Information Management Resource Kit

Module on Digitization and Digital Libraries

UNIT 5. CREATION AND SHARING OF DIGITAL LIBRARIES

LESSON 2. FACILITIES AND REQUIREMENTS

NOTE

Please note that this PDF version does not have the interactive features offered through the IMARK courseware such as exercises with feedback, pop-ups, animations etc.

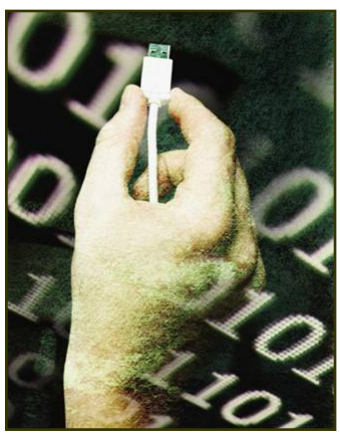
We recommend that you take the lesson using the interactive courseware environment, and use the PDF version for printing the lesson and to use as a reference after you have completed the course.

Objective

At the end of this lesson, you will be able to:

• identify the resources required for creating a digital library collection.

Introduction



Several resources are required for the creation of digital library collections, their maintenance and provision of services.

The two major resources needed are **technology infrastructure** and **personnel**.

Technology infrastructure includes computing and other equipment, software and network connectivity.

Personnel will be required for handling various tasks associated with the creation and maintenance of the collection.

In this lesson we identify these requirements.

Technology Infrastructure: Online and offline access

Access to a digital library collection can be provided **online or offline**.

Online access today typically means that clients use a web browser on a desktop computer and access the collection by connecting to the digital library website over the Internet.

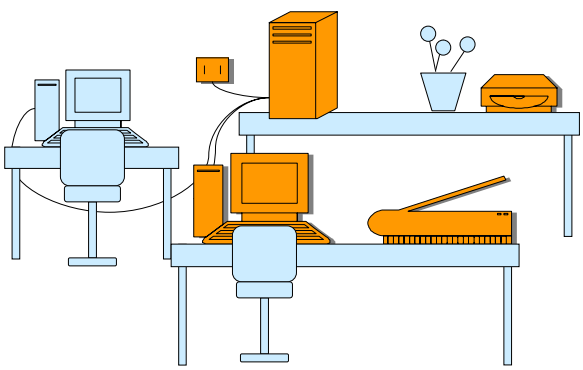
Online access requires a connection to the Internet or to an internal network (Intranet).

In offline access, the digital library is not accessible over a network. One way of providing offline access to a digital library collection is to receive and respond to user queries over e-mail.

Another way is to distribute the digital library collection **on a CD-ROM**.

Hardware

A digital library project would typically require the following equipment:



- Server computer
- Desktop computers
- Digitization equipment
- Network connectivity
- Other equipment

Server computer

The server computer holds the digital library collection and runs the digital library software which provides the searching and processing applications. The server also handles the communication with the user over the network.

The server needs to:

- be reasonably powerful (typically measured in terms of processor speed, for example mega or giga hertz),
- have adequate main memory (RAM),
- have large amounts of hard disk storage, and
- have good communication capabilities (network card and bandwidth).

Desktop computers

Desktop computers will be required for staff handling various tasks related to the digital library collection. These include:

- managing the content in the digital library (conversion, editing, cataloguing, submission and quality checking);
- software development and technical support, and
- digital library server and website maintenance.

Also consider desktop computer requirements for users if public access is provided locally.

You need to assess the actual requirements depending on the nature of your digital library project. For small digital library projects, it may be possible to use existing desktop computers having spare capacity.

Capture devices

These include the digitization equipment needed for converting source material from physical/ analogue format to digital format.

Digitization equipment include scanners for scanning print publications, video and audio capture cards and players for capture/conversion of digital video and audio files, and digital cameras for capture/production of video clips/still images.

Requirements will be minimal if capture and conversion are largely outsourced or if the source material is already available in digital format.

Network connectivity

There are several options for providing **online access** to your digital library collection via the Internet.

If your organization already has the requisite network infrastructure (intranet and Internet connectivity), your network administrator can help you connect and configure your digital library collection server to be accessible over the Internet.

Other equipment

Other equipment you might need include:

- CD Writers and label printing systems,
- laser printers,
- tapes or CD-ROM backup systems,
- UPS support,
- local and offsite storage backup systems.

Hardware

In regard to the server, options include specialized servers or server-class PCs. Here are some guidelines you might find useful when choosing your server...

- Ensure that the **server is expandable** (RAM, storage, CPU) particularly if the size and usage of your digital library collection is expected to grow rapidly.
- Options for the Server operating system include Windows 2000, Unix and Linux. This is mainly determined by the digital library software you plan to use. **License costs** may be involved if it is a proprietary operating system.
- Ensure that **maintenance support** is available.

Hardware

In addition ...

- In terms of **capacity**, it is useful to plan for a five year life span and estimate the resource requirements at 25-50% higher than projected growth.
- Plan for online and offline **storage backups** (e.g. over a network, CD or DVD) for the digital library collection and procedures for taking regular backups.
- You should also have a **contingency plan** if the server computer fails – for example by moving your collection to another server. This requires that you have the required digital library application programs supported on this server. Another option is to host the collection on one or more mirror servers, i.e. alternate servers containing the same information as the original one, located at different geographical locations.

Hardware

Disc **storage space required** on the server computer is a key consideration for digital library collections. How to estimate it?

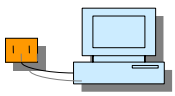
Peter Noerr provides a very useful way of calculating approximate storage space required.

For example, a small collection of 100,000 articles averaging 5 pages, all in English, to be stored in full text and indexed for full text searching will require about 3 GB (Giga bytes):

Characters/Page	2,000	
Characters/Article	10,000	5 pages/article
Characters/Collection	1,000,000,000	100,000 articles/collection
Raw Data Bytes	1,000 MB	1 byte/character
Database Structure Overhead	200 MB	1 page/article = 2 KB/article
Index Overhead (full text index)	1,000 MB	100% of raw data
Bib Records Overhead (Metadata)	150 MB	500 bytes + 200%/article
Subtotal	2,350 MB	
Processing, RAID (redundant disc storage), etc.	780 MB	33%
Total	3,000 MB	= **3 GB**

Peter Noerr provides guidelines on how to calculate approximate storage space required for four main types of electronic documents: text, images, audio and video. For more details see The Digital Library Toolkit. Peter Noerr. Sun Microsystems. Third Edition. January 2003 http://www.sun.com/products-n-solutions/edu/whitepapers/digitaltoolkit.html)

Hardware



Let's now consider some issues concerning **networking** and the location of the digital library server computer…

If your organization has an intranet with Internet connectivity, you can locate the digital library server on the intranet. If your collection is relatively small and does not require a dedicated server, you can host it on your **library web server**, if one exists, or on the institution web server, if this is permitted.

If you don't have network infrastructure and you have the required resources, then you can **obtain dedicated Internet connection** to an ISP (Internet Service Provider). This would require installation of additional equipment and higher maintenance costs.

Another option is to **locate the server computer in a data centre or ISP**, upon payment of a rental fee, or **host your collection** on a server computer owned by the ISP. However, updating the digital library collection may not be easy since such operations will have to be performed over dial-up Internet access. Maintenance and troubleshooting may also be difficult. The ISP may not provide the support you need.

Hardware

Which equipment options are required by the digital library project described on the left?

Access to the collection will be provided online.

The Digital Library software works with any operating system.

The budget is low.

The collection is small.

A library server with an open source operating system is available.

- ☐ Storage backup system
- ☐ CD-Rom writer
- ☐ Open source operating system
- ☐ Windows operating system
- ☐ Hosting on the library web server
- ☐ Server computer dedicated to the digital library collection

Please select the options of your choice (2 or more) and press "Check Answer".

Software aspects

Now let's focus on the choice of the **Digital library application software**.

This needs careful consideration, as it is probably the most important and central aspect of the technology infrastructure for the digital library.

Digital library software works with the web server (and the operating system) in providing various **digital library functionalities** including creation, organization and maintenance, indexing, search and retrieval.

Other software you may need
Other software you may need include: Digitization software, such as Acrobat, OCR software (e.g. FineReader), image editors, web server and browsers, word processing software including HTML/XML editors, and DBMS package.
Many of these are freely available, both for Windows and Linux operating systems.

Software aspects

A clear understanding of the **requirements and features of the collection** you plan to build is very important in order to assess and select among software options that are available today.

A good understanding of what features to look out for in digital library software will also help you make an informed decision.

In the next screens, we will discuss a few software options you may want to consider when creating and providing access to your digital library collection.

Software aspects

What features should you look out for in off-the-shelf digital library software?

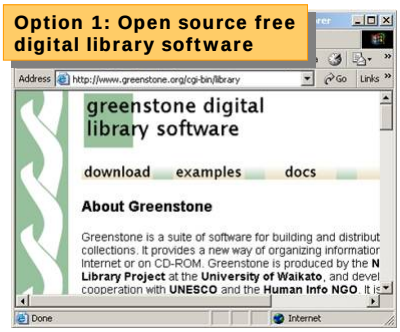
Support for different document types (e.g. journal articles, technical reports) and formats (e.g. PDF, Word Doc)
Support for customized metadata for the identified document types,
Online/batch document and metadata submission
Indexing and storage (metadata and/or full text)
Search and retrieval (metadata, full-text)
Multi-lingual support
Interoperability support for sharing your digital library collection
Access and usage management for managing and monitoring usage
Collection administration
Support for standards like Dublin Core (metadata standard), Unicode and XML (for document mark-up).

Software aspects

Several **free digital library software packages** are now available, enabling the easy creation and sharing of information through digital library collections.

All of these are available for the Linux operating system platform - a popular, freely available operating system.

Option 1: Open source free digital library software



Specific features offered by these packages vary, but since they were all developed to meet the specific requirement of establishing and providing access to digital collections, you should consider them seriously. Most of these packages also support the Open Archives Initiative interoperability protocol (http://www.openarchives.org/), an attractive proposition if you wish to share your digital library collection with other collections on the Internet.

On the downside, you may not have ready access to technical help when using these packages, though most have mailing lists where you can seek e-mail based help.

Software aspects

Examples of Open source free digital library software

Example packages include:

Greenstone Digital Library Software: New Zealand Digital Library
- http://www.greenstone.org/

ARNO: Academic Research in the Netherlands Online, Tilburg University, The Netherlands
- http://www.uba.uva.nl/arno

CDSware: CERN Document Server Software (CDSware), CERN, Geneva, Switzerland
- http://cdsware.cern.ch/

Dspace: MIT Libraries, Cambridge, MA USA
- http://www.dspace.org/

Eprints: University of Southampton, U.K.
- http://software.eprints.org/

Fedora digital object repository management system: University of Virginia, USA
- http://www.fedora.info/

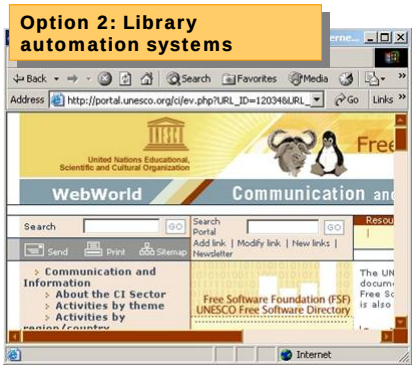
i-Tor: Tools and technologies for Open Repositories; Netherlands Institute for Scientific Information Services
- http://www.i-tor.org/en/toon

MyCoRe: Essen University Library, University of Duisburg-Essen,Germany
- http://www.mycore.de/engl/index.html

See also tools on the open archives website: http://www.openarchives.org/tools/tools.html

Software aspects

Option 2: Library automation systems



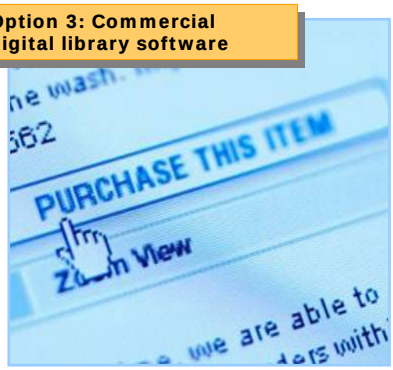
Since most library automation packages today are web enabled (i.e.**accessible over the Web**), it should be possible to extend access to digital material in a simple way by including relevant **hypertext links** in catalogue records.

From a catalogue (or **OPAC**) search output, a user can simply click on a link in the catalogue record display to bring up the complete electronic document. If your library is already using a web enabled library automation package, you may want to consider this option for providing access to your digital library collection. Of course, this would require you to store electronic documents in appropriate folders on the server computer disc and insert links to these document files in the library OPAC records.

Examples of Free Automation and Bibliographic Database Packages
You might also want to explore freely available library automation and bibliographic database packages. A prime example is UNESCO's CDS/ISIS package for bibliographic/textual database management software. It is now possible to enable web access to CDS/ISIS. You can embed links to electronic documents in CDS/ISIS records and allow users to access these documents by selecting these links in the retrieved records. Another free package you may like to consider is KOHA. (See resources section for details).

Software aspects

Option 3: Commercial digital library software



If you have the resources, you have the option of selecting from several commercial digital library software packages that are available today.

Onsite **technical support** is an advantage of commercial packages (this may not be available in all locations: you need to check).

On the downside, these packages can be quite **expensive**. Also, you may not be able to extend their functionality as these packages are proprietary and you usually do not have access to **source code**.

Examples of Commercial Digital Library Software Packages

ENCompass of Endeavor Information Systems (http://www.endinfosys.com/)
DigiTool of Ex Libris (http://www.exlibris.co.il/, or http://www.exlibris-usa.com/)
Visual MIS (Multimedia and Imaging Solutions) of VTLS (http://www.vtls.com/)
TEAMS of Artesia Technologies (http://www.artesia.com/).
Insight of Luna Imaging (http://www.luna-imaging.com/)
MuseSearch of MuseGlobal, Inc. (http://www.museglobal.com/)

Software aspects

Option 4: In-house software development

If you have in-house software development and maintenance support, another interesting possibility is **to develop the software in-house**. Simple, database-driven web applications to set up and provide access to your digital library collection can be developed relatively easily.

Open source database packages (e.g. MySQL and PostGress) and programming tools (e.g. Perl, PHP, Python and Java) available for the Linux operating system platform can be used for developing the digital library application.

An advantage of this approach is that you can **customize** the application to suit your requirements. A disadvantage is the **cost of development** as software projects are known to run over schedule. **Maintenance** may also be difficult if you do not have a dedicated software team.

Software aspects

Let's try to summarize some features of different types of software options.

Can you match each option with its related advantage?

Option 1: Open source free digital library software

Option 2: Library automation systems

Option 3: Commercial digital library software

Option 4: In-house software development

Readily customizable to suit your requirements.

Also available for the Linux OS (for free).

It provides onsite technical support.

Simple implementation.

Click each option, drag it and drop it in the corresponding box.
When you have finished, click on the "check answer" button.

Personnel

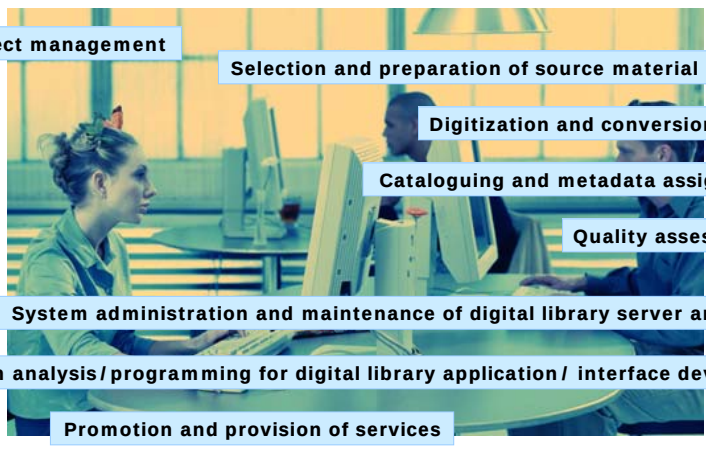
Personnel are a digital library's most important resource, not only during its initial creation and set up, but also for its operation, maintenance and provision of services.

Since access to the digital library is easy compared to a physical library, **more users are likely to access it**. If the digital library does not meet the expectations of the users in terms of currency and quality of content, they will lose confidence and are less likely to visit the digital library again.

It is therefore important that you assign personnel with the right skills and attitude to handle various tasks associated with the digital library project.

Personnel

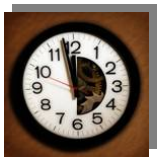
Broadly, **personnel will be required for the following tasks**:



Personnel

Personnel required for a digital library project depends on the type and volume of tasks to be carried out. Let's consider the following suggestions.

For small and medium size digital library projects, a part-time project manager working with just one or two well-trained volunteers may be sufficient to successfully complete the project. For some digital library projects (e.g. dissertations and theses), there may already be staff in the library handling the source material in paper-based form. In converting this material for a digital library, much of this staff may be trained to handle processes in the digital environment.



For temporary and part-time tasks, you may also consider seeking support from enthusiastic volunteers from within the library or other departments. You can also take advantage of student's assistance/intern programmes available in most educational institutions for employing students on hourly payment basis. It may also be possible to contract services, particularly for grant-based projects.

While hiring new personnel or re-deploying existing personnel, it is important that their education and skills match the requirements of the job.

It is also important that all staff employed in the digital library project have their positions described in clear terms, including responsibilities and expected performance standards.

Personnel

Personnel skills include both **technical and soft skills**. Technical skills will be required for staff handling various technical tasks related to the digital library collection. Soft skills include:

- communication skills,
- ability to work in a team,
- flexibility,
- people skills,
- capacity to learn constantly and quickly,
- ability to work under pressure,
- attention to details,
- goal oriented,
- time-consciousness,
- capacity to work independently,
- public-service perspective,
- presentation skills,
- negotiation skills, and
- teaching/ training skills.

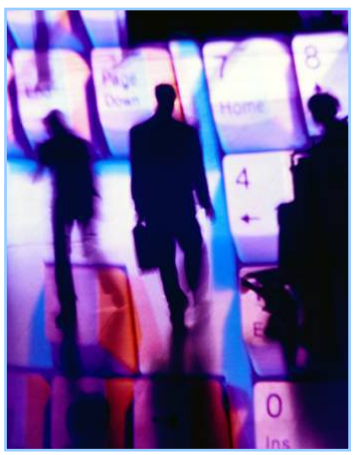
Personnel

It may not always be possible to find personnel with the right background: **Staff training** becomes critical.

Training should be appropriate to the skill requirements of the concerned operations. In-house training options include special **training programmes** and **on-the-job training**.

Staff may also be deputed to relevant training programmes. It will be useful to prepare a **training manual** which will be very useful for new staff. There are also a large number of online resources, including tutorials, available from authoritative sources related to digitization.

Personnel



The rapid changes in technology and practices in digital library technologies require constant **re-training and re-positioning of staff**.

This usually means that the project manager spends a good bit of time keeping up-to-date on developments in the field and learning about emerging standards and best practices. He or she then takes what is valuable, incorporates it into the project plan, and trains other staff in its implementation.

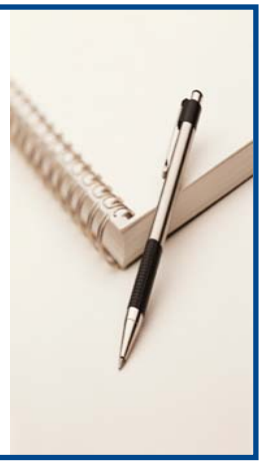
Summary

Information technology infrastructure and personnel comprise two critical resources required for digital library projects.

Hardware requirements include a server computer for hosting the collection, desktop computers, digitization equipment, network connectivity, and other equipment.

Digital library software is another critical technology component. Options include: Open source free digital library software, library automation software, commercial digital library software and in-house software development. Each of these have advantages and disadvantages.

Personnel comprise the most important resource for the digital library apart from technology. Actual personnel required for a digital library project depend on the type and volume of tasks to be carried out.



Exercises

The following five exercises will help you test your understanding of the concepts covered in the lesson and provide you with feedback.

Good luck!

Exercise 1

What is the computer which hosts the digital library collection and runs the digital library?

Write the correct word in the box. Then, click on the "Check Answer" button.

Exercise 2

If your digital library collection is to be accessed online, the server computer which hosts the collection should have dedicated Internet connectivity.

- True
- False

Please click on the answer of your choice

Exercise 3

Which of the following determine the capacity of a server computer for hosting a digital library collection?

- ☐ Monitor size
- ☐ RAM
- ☐ CD-ROM drive speed
- ☐ Processor Speed
- ☐ Disk storage

Please select the options of your choice (2 or more) and press "Check Answer".

Exercise 4

Given a choice, it is better to host the digital library collection on a computer located in the ISP's data center.

- ○ True
- ○ False

Please click on the answer of your choice

Exercise 5

Which of the following are soft skills?

- ☐ Teaching/ training skills
- ☐ Scanning skills
- ☐ Cataloguing skills
- ☐ Negotiation skills
- ☐ Software development skills

Please select the options of your choice (2 or more) and press "Check Answer".

If you want to know more...

Online resources

Peter Noerr. Digital Library Toolkit. 3rd edition. Sun Microsystems, January 2003. http://www.sun.com/products-n-solutions/edu/whitepapers/digitaltoolkit.html

North Carolina ECHO (Exploring Cultural Heritage Online), State Library of North Carolina. Digitization guidelines (best practice guide). http://www.ncecho.org/guidelines.asp

Kentuckiana Digital Library. Digital Library Production Guide version 1.0. http://www.kyvl.org/kentuckiana/bpguide/about.shtml.

Building Digital Collections: Technical Information and Background Papers. National Digital Library Program (NDLP) at the Library of Congress. http://memory.loc.gov/ammem/techdocs/index.html

UNESCO Free Software Portal. http://www.unesco.org/webworld/portal_freesoft/index.shtml

Oss4lib - Opensource systems for libraries. http://www.oss4lib.org/

Koha – Open source library system. http://www.koha.org/

Tools on the open archives website. http://www.openarchives.org/tools/tools.html

A guide to Institutional Repository Software. 2nd edition. Open Society Institute. January 2004. http://www.soros.org/openaccess/software

If you want to know more...2

Online resources

Example Open Source Packages

Greenstone Digital Library Software - New Zealand Digital Library
http://www.greenstone.org/
ARNO - Academic Research in the Netherlands Online, Tilburg University, The Netherlands (http://www.uba.uva.nl/arno)
CDSware - CERN Document Server Software (CDSware), CERN, Geneva, Switzerland (http://cdsware.cern.ch/)
DSpace - MIT Libraries, Cambridge, MA USA (http://www.dspace.org/)
Eprints
University of Southampton, U.K. (http://software.eprints.org/)
Fedora digital object repository management system - University of Virginia, USA (http://www.fedora.info/)
i-Tor - Tools and technologies for Open Repositories - Netherlands Institute for Scientific Information Services (http://www.i-tor.org/en/toon)
MyCoRe - Essen University Library, University of Duisburg-Essen,Germany (http://www.mycore.de/engl/index.html)

Examples of Commercial Digital Library Software Packages

ENCompass of Endeavor Information Systems (http://www.endinfosys.com/)
DigiTool of Ex Libris (http://www.exlibris.co.il/, or http://www.exlibris-usa.com/)
Visual MIS (Multimedia and Imaging Solutions) of VTLS (http://www.vtls.com/)
TEAMS of Artesia Technologies (http://www.artesia.com/).
Insight of Luna Imaging (http://www.luna-imaging.com/)
MuseSearch of MuseGlobal, Inc. (http://www.museglobal.com/)