

Information Management Resource Kit

Module on Digitization and Digital Libraries

UNIT 4. CREATION AND MANAGEMENT OF DIGITAL DOCUMENTS

LESSON 2. WHAT TO DIGITIZE?

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.



© FAO and UNESCO, 2005

Learning Objectives

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

- understand the main **purposes of digitization**; and
- apply **criteria for selecting and prioritizing** documents to specific situations.



Why digitize?

Mr. Touré, a library manager, is evaluating the advantages of digitizing his library's hardcopy documents.

Hmm... Digitizing documents could be the best answer to our needs!



Why would he want to digitize documents?

He may want to:

- **preserve** his documents, that is to allow people to read them without **damaging** the originals;
- make the documents more **accessible** to the users; or
- **reuse** the documents for different purposes.

Let's analyze these three needs...

Why digitize?

These are the three main needs for digitization; two (or all three) of them may apply to your digital library project.

NEED	DESCRIPTION
To preserve the documents	<ul style="list-style-type: none">• To allow people to read older or unique documents without damaging the originals
To make the documents more accessible	<ul style="list-style-type: none">• To serve existing users better (e.g. to allow users to search the full text of the documents)• To serve more users than is now possible (e.g. people at remote locations, or more than one person at a time)• To bring together scattered materials on a specific topic• To respond to a particular request for a digital library
To reuse the documents	<ul style="list-style-type: none">• To convert documents into different formats (e.g. to use images in a slideshow)• To adapt the content for a different purpose (e.g. to convert the text of a report into training materials)

Why digitize?

For example, imagine you run a library or a documentation centre. You want to make sure that content and look and feel of documents will be displayed as originally intended over time and regardless of the software utilized for their creation.

What is your main goal?

- Reuse
- Preservation
- Providing access

Please click on the answer of your choice

What to digitize?

Let's come back to Mr. Toure's case. A meeting with his colleagues raises some important questions...

The idea is interesting, but, before starting, we should clarify some points. Are we sure it is worth it? And, if yes, which of our documents should we digitize?



- Is it **worth** digitizing the documents?
- Is there enough **demand** from users to justify the costs?
- Do the documents form a **collection**?
- How **easy** will it be to digitize them?
- How **many pages** have to be digitized?
- What about **copyright** and **security**?

We will go through these questions in this lesson. They provide criteria for selecting and prioritizing documents to digitize.

What to digitize?

Is there a particular request to digitize documents?

You may be asked to digitize a particular set of documents – say, a research journal, or the institute's personnel records. This request may come with an offer of funding. If the request comes from the boss, it can be hard to refuse!

Requests to digitize documents have advantages and disadvantages:

ADVANTAGES:

The requester will support the digitization effort; this can be very useful if **funding** is tight, or if your digitization efforts are just **beginning**.

DISADVANTAGES:

The particular documents to be digitized are not always the most appropriate ones; they may be low priority from **other** points of view.

Is it worth digitizing?

First question is: is it **worth** digitizing the documents?



Digitizing documents can take a lot of time, effort and money. Make sure it is worthwhile.

Do the documents contain **information** that is **valuable enough** to warrant the costs? Is the information of lasting value? There is little point in digitizing documents (such as bulky seminar reports) that are already out of date.

Or perhaps the **documents themselves** are valuable? Digitizing old, unique documents that are easily damaged (such as illuminated manuscripts) allows people to use the documents without handling the originals. These unique documents are sometimes called "heritage documents".

Are there **other electronic copies** of the documents available? Perhaps they have already been digitized by someone else.

Who is your audience?

Is there **enough demand from users** to justify the costs?



If there are only a **few** users (such as scientists in a highly specialized subject), perhaps they can be served by sending them photocopies.

Maybe there is a **large** number of potential users (such as farmers or extension workers), but they do not have computers to **access** the digital library.

It may be difficult to judge the demand for documents. You may wish to get other people's opinions. Ask potential users of the documents what **they** see are the priorities.

Remember, the users may not be familiar with electronic documents, so you may have to interpret what they say carefully.



Do the documents form a collection?

Could some of the documents we want to digitize be grouped into **collections**?



It's important to verify if the documents form a **collection**.

In fact, the documents in a digital library should have something in common, like a common **subject focus** or **the same author** (or group of authors).

For example:

- agricultural extension manuals for farmers in the drylands of Southeast Asia; or
- publications produced by the Indonesian Research Institute for Rice.

It may be possible to enrich the library with documents from other organizations or collections (either by including them in the collection directly, or by adding links to them if they are stored on other servers). Of course you need to consider copyright issues.

How easy is it to digitize documents?

Another important factor to take into account is **how easy** it will be to digitize the documents.

Not all hardcopy documents are easily converted to electronic format.

For example, which of the following documents do you think are easy to convert to digital format?

- Documents printed on coloured paper.
- Journal articles, consisting mainly of text.
- Scientific papers with equations and tables.
- Extension leaflets with one or two line drawings per page.

Click on the answers of your choice

How easy is it to digitize documents?

Check the **physical characteristics** of the documents to understand **how easy** it will be to digitize them.

Easy to convert	Difficult to convert
White, clean opaque paper	Coloured, damaged or thin paper
Simple layout, single columns	Complex layout, multiple columns
Single sheets	Fragile, heavy bindings
Straight text with headings, few pictures	Many pictures, equations and tables
Standard computer typefaces	Unusual typefaces, poor quality printing, typewriting, handwriting
Unaccented Roman scripts	Accented and non-Roman scripts

If you have a lot of documents that are hard to digitize, you might choose:

- not to include them in the digital library;
- include them as image files rather than as searchable text; or
- have them retyped rather than scanned.

How many pages have to be digitized?

How many pages are there in the documents?



The number of pages is more important than the number of documents.

More pages means a higher cost of digitizing, and more time will be needed.

If you have only a few thousand pages, or a few dozen documents, you might consider using a simple menu-based system to allow users to retrieve documents.

If you have more pages or documents, you should consider using a database or digital library software so users can find information easily.

Remember, though, that a large digital library with a lot of documents is more likely to be useful than a small one.

You can always start off with a small number of documents, then add more as you go along.

One-off or ongoing project?

To correctly evaluate the amount of material to be handled, you should also understand if your digitization project is one-off or ongoing.



One-off digital library project

A **one-off** digital library project has a limited set of documents to digitize. For example: the publications produced by an institute up to a **certain date** (newer publications may already be in digital format and directly put into the digital library).



Ongoing digital library project

An **ongoing** project continually adds new documents. For example: a set of **newspaper clippings** on a particular topic (the number of new clippings will be hard to predict, as it will depend on whether the topic is in the news).

If you have an ongoing digital library project, you will have to make sure funding and staff will be available to carry on the work in the future.

One-off or ongoing project?

Which of the following is an ongoing digital library project?

- A collection of an ancient author's writings.
- A collection of student theses at a university.
- All the documents collected during a workshop.

Click on the answer of your choice

What about copyright and security?



Check the **copyright** to make sure there are no restrictions on digitizing the documents and making them available in a digital library. If there are restrictions, you should obtain the necessary permissions before scanning the documents. Remember that:

- **government** documents are increasingly being copyrighted;
- **commercially published** documents are almost always copyrighted; and
- even if there is no © **symbol** in the document, it may still be copyrighted.



Digitizing documents makes them easier to access. But some documents are **confidential**, or access to them must be **restricted**. Some examples:

- minutes of policy meetings;
- personnel, financial and salary records; and
- commercial secrets.

It is possible to restrict access to all or part of a digital library, but this involves extra steps. You will have to password-protect the library (or parts of it), and take extra care to ensure that the computer system is secure.

Selecting and prioritizing documents

Once you have considered all the previous criteria, you should be able to decide **which documents to include** in your digital library.

It seems that not all our documents are easy to digitize...

True. And we have many more documents than we can digitize immediately.



Well, let's use the discussed criteria to prioritize them!

Selecting and prioritizing documents

Try developing a scoresheet to help you **prioritize** them. A table like the one below can be used to give scores for each of the documents (or set of documents, such as a journal series).

Criteria		Doc. Set A	Doc. Set B	Doc. Set C
Document value	3 = high 1 = low			
User demand	3 = high 1 = low			
Fit into collection	3 = yes 1 = no			
Ease of digitization	3 = easy 1 = hard			
Number of pages	3 = few 1 = many			
Copyright	3 = no restriction 1 = restricted			
Security	3 = no problem 1 = problems			
Particular request	3 = yes 1 = no			
Total	Max = 24			

Selecting and prioritizing documents

So let's evaluate our first set of documents. They are research reports. I have taken notes of their features using a checklist. Now, we can give scores!

The following **checklist** is based on the criteria we have analyzed. It describes the features of the set of research reports.

- 1) Click on the following icon to read it.

Checklist for research reports

- 2) Type the numbers in the relevant fields in the table.

- 3) Click on **View Answer** to see the suggested answer.

Criteria		Research reports
Document value	3 = high 1 = low	
User demand	3 = high 1 = low	
Fit into collection	3 = yes 1 = no	
Ease of digitization	3 = easy 1 = hard	
Number of pages	3 = few 1 = many	
Copyright	3 = no restriction 1 = restricted	
Security	3 = no problem 1 = problems	
Particular request	3 = yes 1 = no	
Total	Max = 24	

View Answer

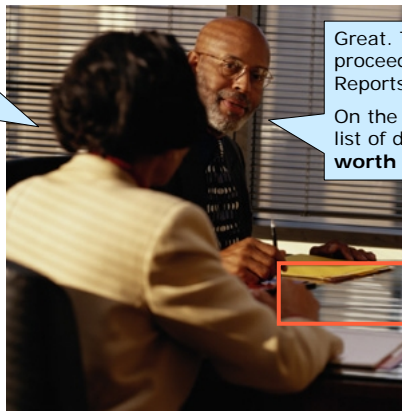
Selecting and prioritizing documents

Mr. Touré and his colleagues have carried out the evaluation for another two sets of documents. As you can see, **Extension brochures** have the highest score.

Criteria		Extension brochures	Research reports	Seminar papers
Document value	3 = high 1 = low	2	3	1
User demand	3 = high 1 = low	3	2	1
Fit into collection	3 = yes 1 = no	3	3	2
Ease of digitization	3 = easy 1 = hard	3	2	2
Number of pages	3 = few 1 = many	2	1	2
Copyright	3 = no restriction 1 = restricted	3	3	3
Security	3 = no problem 1 = problems	3	3	3
Particular request	3 = yes 1 = no	1	1	3
Total	Max = 24	20	18	17

Selecting and prioritizing documents

So, **Extension Brochures** should be given the top priority for digitizing!



Great. Then, later on, we could proceed with Research Reports.

On the other hand, here is the list of documents that are **not worth digitizing**.

Documents that are not worth digitizing

- Documents that are **out of date**
- Documents with **poor quality** printing or that are otherwise hard to digitize
- Content that is **not relevant** to the intended collection
- Materials where you do not have **copyright** permission
- **Sensitive** material (unless you have a suitable security system in place)

Guidelines and procedures

Here you can download and print the documents provided in this lesson.

You may use them as tools for your job.



Checklist - to evaluate your documents following the defined criteria



Table - to give scores to your documents for prioritizing them



Summary

When choosing documents to digitize, you should evaluate if:

- it is **worthwhile** to digitize them;
- they are **easy** to digitize;
- they will form a **viable** collection; and
- they are free from **copyright** or **security** restrictions.

Also make sure there is enough demand from **users** to justify the effort of digitizing documents and putting them into a digital library.

Once you have considered all the previous criteria, you should be able to decide **which documents to include** in your digital library.



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

Which is the **main** purpose for digitizing in each of these situations?

- a
- You have a collection of old documents and books that are of historical interest. They are becoming damaged as they get older.
 - You have a collection of publications your organization has produced over the last 25 years. You want to make them available to users.
 - You have a set of several thousand photographs of insects. They will make an excellent basis for an illustrated guide for students and farmers.
 - You have a collection of students' theses that are currently available only in your library building.

Preservation

1

Reuse

Accessibility

Accessibility

Click on each option and drag it to the correct box.
Then, click on Check Answer.

Exercise 2

You are considering digitizing various documents. What types of documents are you more likely to include in a digital library?

- A series of popular commercially published books
- A series of books published by your own institution

Please click on the answer of your choice

Exercise 3

You are considering digitizing various documents. What types of documents are you more likely to include in a digital library?

- A set of rare, historical photographs
- A set of photographs of prominent modern politicians

Please click on the answer of your choice

Exercise 4

You are considering digitizing various documents. What types of documents are you more likely to include in a digital library?

- A collection of manuscripts about a highly specialized field of agricultural research
- A collection of easy-to-understand manuals on common farming problems

Please click on the answer of your choice

Exercise 5

You are considering digitizing various documents. What types of documents are you more likely to include in a digital library?

- A large pile of unsorted documents, the contents of which you are not sure of.
- A small collection of documents on a specific subject.

Please click on the answer of your choice

If you want to know more...

Online Resources:

Cornell University Library/ Research Department. 2000. Moving theory into practice: Digital imaging for libraries and archives. Research Libraries Group. (<http://www.library.cornell.edu/preservation/tutorial/>)

Institute of Museum and Library Services (IMLS). 2001. A framework for building good digital collections. (<http://www.ims.gov/scripts/text.cgi?/pubs/forumframework.htm>)

Noerr, Peter. 2003. The Digital Library Toolkit. Sun Microsystems. (http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital_library_toolkit.pdf)

Sitts, Maxine K. (ed.) 2000. Handbook for digital projects: A management tool for preservation and access. Northeast Document Conservation Center, Andover, Massachusetts. (<http://www.nedcc.org/digital/dman.pdf>)

Additional Reading:

Witten, Ian H, and David Bainbridge. How to build a digital library. Morgan Kaufmann, San Francisco. 2003. Chapter 2: Preliminaries: Sorting out the ingredients.

