Date submitted: 28/05/2009

**General guidelines for designing bilingual low cost digital library services suitable for special library users in developing countries and the Arabic speaking world**

**Ramadan Elaiess**
Department of Computer and Information Sciences
University of Strathclyde & University of Garyounis
Benghazi, Libya

**Meeting:**           **175. Government Information and Official Publications**

## Abstract

*The World is witnessing a considerable transformation from print based-formats to electronic-based formats thanks to advanced computing technology, which has a profound impact on the dissemination of nearly all previous formats of publications into digital formats on computer networks. Text, still and moving images, sound tracks, music, and almost all known formats can be stored and retrieved on computer magnetic disk. Over the last two decades, a number of special libraries and information centres in the Arab world have introduced electronic resources into their library services. Very few have implemented automated and integrated systems. Despite the importance of designing digital libraries not merely for accessing to or retrieval of information but rather for the provision of electronic services, hardly any special library has started the design of digital library services.*

*Managers of special libraries and information centres in developing countries in general and in the Arab world in particular should start building their local digital libraries, as the benefit of establishing such electronic services is considerably massive and well known for expansion of research activities and for delivering services that satisfy the needs of targeted end-users. The aim of this paper is to provide general guideline for design of special low cost digital library providing services that are most frequently required by various categories of special library users in developing countries. This paper also aims at illustrating strategies and method approaches that can be adopted for building such projects. Seeing the importance of designing an inexpensive digital library as basic principle for the design accordingly, the utilisation of today's ICTs and freely available open sources software is the right path for accomplishing such goal. The paper intends to describe the phases and stages required for building such projects from scratch. It also aims at highlighting the barriers and obstacles facing Arabic content and how could such problems overcome.*

1

## Introduction

The world is witnessing a considerable transformation from print based-formats to electronic-based formats thanks to advanced computing technology, which has had a profound impact on the dissemination of nearly all previous formats of publications into digital formats on computer networks. Text, still and moving images, sound tracks, music, and almost all known formats can be stored and retrieved on computer magnetic disk (Ashoor, 2000). Over the last two decades, a number of special libraries and information centres in the Arab world have introduced electronic resources into their library services. Very few have implemented automated and integrated systems. Despite the importance of designing digital libraries, not only for accessing or retrieving information, but also for the provision of electronic resources and services, a handful of special libraries in the Arab region have begun to design digital library services. Managers of special libraries and information centres in developing countries in general and in the Arab world in particular should begin building their local digital libraries, as the benefit of establishing such electronic services is considerable and known to expand research activities and deliver services that are most wanted and most frequently required by target end-users. Building low cost digital libraries is becoming uncomplicated thanks to the availability of open sources software, which is required for building such services. Nowadays, various software, such as Greenstone, Dspace, and Eprint, is freely available. The Greenstone software was produced to build New Zealand's digital library, and afterwards became available to assist particular libraries in developing countries to build their local digital libraries in order to enhance the services presented to target end-users and to provide adequate and appropriate information resources. The software provides a new way of organising information and making it available over the Internet or on CD-ROM. The following section will discuss briefly how to build a low cost digital library using freely available open sources software, and what sort of services could be provided through the proposed digital library.

## Digital libraries

There is no doubt that there are many different views in the literature as to the actual nature of digital libraries. This paper does not intend to provide a comprehensive collection of definitions of the digital library, but rather a number of representative definitions. A variety of terms are still used interchangeably worldwide, such as electronic library, hybrid library, library without walls, cyber library, virtual library etc. Arms (2000) views digital libraries as "managed collection of information with associated services, where the information is stored in digital formats and accessible over a network". Witten (2003) defines the digital library as a focused collection of digital objects, including text, video, and audio along with methods for access and retrieval, and for selection, organisation and maintenance of the collection. The Digital Library Federation (DLF) defines digital libraries as "organisations that provide the resources, including the specialised staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities." According to Jeng (2006) digital library in the 21st century has the following characteristics:

- It is an organised and managed collection of digital objects
- It is accessible over the internet or a server
- It is a global information infrastructure and;
- It should offer services.

The last point indicates that there is a difference between a digital collection and a digital library in that a digital library should offer service to end users. Therefore, a digital library is considered a collection of information objects and a collection of services that should be provided by the digital library. "The definition of a digital library that came up in March 1994 in the Digital Library Workshop emphasised that a full service digital library must accomplish all the essential services of traditional libraries and also exploit the well-known advantages of digital storage, searching, and communication" (Chowdhury , 2002).  Leiner (2009) reports that "There are a large and varied set of services, including services to support the management of collections, services to provide replicated and reliable storage, services to aid in query formulation and execution, services to assist in name resolution and location, etc."  At the initial meeting of the WG (held January 7-8, 1998 at Stanford University), the following definition was proposed by Leiner (2009) "A digital library is

- The collection of services
- And the collection of information objects
- That support users in dealing with information objects
- And the organisation and presentation of those objects
- Available directly or indirectly
- Via electronic/digital means."

## Phases of building low cost digital library services

The first step in designing a low cost digital library for any special library or information centre is the planning of the new system. However, before the planning stage, the objectives and goals of the project should be considered.

## Objectives

In the beginning there is a need to identify the project's goals before taking any further steps because the whole project would be worthless without clear objectives. The objectives of the design could be, for example:

- Developing the standard of services presented to targeted end users to enhance the level of research activities within the organisation.
- Producing the design of a low cost digital library providing not only access to and retrieval of information, but rather services that are most wanted and frequently required by end-users.
- Using and putting into practise modern ICTs to improve the standard of services provided to end-users and to deliver services that are most required and aspired by special library' users.

## Methodology

An experimental and system design approach should be used for the design of the proposed digital library. The BNET dictionary defines this approach as "the planning of the procedures to be used in an experimental study" (BNET dictionary). This approach should be applied due to its suitability for such types of projects, as it has been proven by various studies that this approach is the most appropriate method for creating digital libraries. The design of a low cost digital library comprises the following stages:

## Planning

This is the fundamental process of understanding why a digital library should be built and determining how the developer will go about building it. Questions examining economic feasibility and organisational feasibility (e.g., if we build digital library, will it be used?) should be answered in this phase to determine the expected value of the proposed digital library. The planning phase consists of the following steps:

## Survey and fieldwork

A survey has to be conducted to gather general information and basic data on the current state of the targeted organisation or institution to which the library belongs. The survey has to focus on the real status of the following:

❖ ICT infrastructure within the organisation. To focus on the number of computer, e.g. how computers are distributed in the organisation, how many for library use and how many for general use.

❖ Type of internet connection

❖ Policy related to:

- Computers

- Networks

- Internet

- Intranet

❖ Policy related to content : To focus on

- Electronic resources

- Budget

- Access control

❖ Policy related to management:

- Familiarity

- Willingness

- Vision:  e.g. visions with regard to developing ICT policy.

- Plans:  e.g. any plan for moving towards an electronic library services.

❖ Policy related to training

- Librarians training

- Users training

- Cost of training courses

- Type of training courses

## Requirements analysis

Developing services that meet the expectations of users and customers is critical to success. "Requirement analysis is the foundation of a user-centred approach, creating projects and services that appeal and meet user needs" (D'Hertefelt, ). User requirement analysis is not about asking users what type of services they want, but rather it is about understanding users' current practices and the problems they encounter. This stage comprises the analysis of fieldwork which has been done to ensure the availability of required components in favour of the design of the digital library. This stage should provide answers to questions such as:

- For whom is the digital library going to be designed?

- How many users are going to benefit from the proposed digital library?

- What is the real status of the ICT infrastructure in the organisation?

- What are the current practices and problems users encounter?

- What type of material is already in digital form?

- What sort of material is to be digitised?

- What Items are distributed within the organisation or the institution?

- What Items or resources are outside the organisation or the institution?

In order to answer the questions above, fieldwork is required for this phase. Observations and on-site visits are significant tools for this stage. Interviews should also be conducted during this phase to gather information pertaining to information resources and services in the targeted organisation. The interviews have to be conducted to identify most required and frequently used resources, as well as to gather information on tools and techniques currently used, problems and so forth. The output will be a list of most frequently used resources and services in addition to recognise how users would prefer to use these resources in order to decide the metadata and the requirements for designing the digital library. Identifying frequently used resources and their format, i.e. hard copy or electronic form, should be the focus of the interviews. Internal resources, which consist of reports and publications created by employees within targeted organisations should also be specified to assist in building the digital library in the most effective way. In brief, the requirements analysis should also indicate the following:

- ❖ Internal documents which include:
  - Reports
  - Statistical data
  - Publications by individuals within the organisation
- ❖ External documents:
  - Purchased materials, whether in
    - Hard copy
    - Electronic format

❖ Documents obtained from other resources

## User studies

User studies are considered to be one of the most important pillars required for designing digital libraries. The aforementioned survey should highlight the extent of demand for electronic resources. The survey should focus on what users really want and what is actually provided for them. In other words, the gap between what users really need and aspire to and what they actually obtain. The results will show whether there is a gap between supply and demand. If the finding of the survey clearly indicates a gap between supply and demand, then designing a low-cost digital library is anticipated to bridge this gap and to improve the services presented to end-users to a greater degree, as well as to meet users' diverse needs and requirements.

## User requirements

The aforementioned survey should shed light on the current status of the special library and the level of users' satisfactions. As is well-known, reliance on homogeneous resources, which comprise only one sort of information material, i.e. printed material, has proved by information specialists to be inefficient in satisfying users' requirements. Consequently, continuing of provision of homogeneous resources will affect the services provided in a negative way. Findings of the survey should also reveal the volume of electronic materials *vis-à-vis* traditional resources. Electronic services constitute a new horizon for special library users that should be rediscovered in order to find out what is new in a world that expands every day with new innovations. Providing heterogeneous information resources as well as materials suitable for teaching, learning and research should enrich research activities in many ways and should improve the status of the current services. Moreover, access to a multi-content digital library will contribute to the development of ongoing research and supply targeted end-users with new experiences.

## Design

Having concluded the requirements analysis, the developer of the digital library can then start the design. The design phase decides how the system will operate in terms of hardware, software and available infrastructure. The first step in the design phase is to develop the design strategy and examine the suitability of the chosen software. It is also important at this stage to decide types of content and the collections, which will comprise a significant part of the digital library. Special collections – printed and/or digital – that are of importance to users in the organisation should be designated at this stage. Knowing the content of the proposed digital library is essential in order to be able to decide on the metadata, as well as on the language of the content. The content of the proposed digital library may be in a single language or it could be in several languages. However, this depends on the purpose of the proposed digital library and the policy of the institution or the organisation to which the library belongs.

### Software

Various software is freely available for creating the proposed digital library. However, as the basic principle of the design depends mainly on the creation of an inexpensive digital library, selection of appropriate software from various freely available open sources software is therefore a complex question. This step is very important in order to reduce the cost of the design and to keep it to a minimum. What is more, due to the availability of a variety of open source software, such as Dspace, Greenstone and many others, that can serve the purpose of designing the proposed library,  selection of appropriate software for building the digital library appears to be complicated issue and must be considered carefully before choosing the software. If Arabic is the language of the content of the proposed digital library or if Arabic is its second language, then Greenstone software would seem to be most appropriate. This software is one of the leading open source software due to its good reputation and support of different languages, which is an essential aspect required for the design of bilingual content. Producing a design for a bilingual digital library that supports both the Arabic and English languages is a crucial aspect. However, it should be taken into consideration that:

- English is the second language in many Arabic countries.

- Almost all literature written in any branch of knowledge is published in English.

- A substantial number of scientists and research workers in almost all Arabic countries are acquainted with the English language.

Greenstone is regarded as one of the leading software in this domain, as mentioned previously, and can deal with documents in different languages. In addition, the software is capable of displaying the user interface in multiple languages and handling collections of text, pictures, audio, and video. The software also offers flexible browsing facilities and can run on Windows and UNIX. Dspace software would also be a very good choice. However, fairly good knowledge of dealing with open source software is required to modify Dspace to support the Arabic language. Despite the availability of various open source software required for building digital libraries such as Dspace;  Bepress;  Eprints; Open repository ; Open publication system; Fedora, and so forth,  many organisations and institutions worldwide have built their digital libraries using Greenstone software. Among these are the Indian Institute of Science Publications database; the Indian institute for management; the Sudanese Association of Libraries and Information;-Sudan Open Archive.
In fact, the most important reason for choosing Greenstone software is its ability in building the content in two languages. This feature is not available in other open source software. No other software mentioned above is capable of building Arabic content or displaying a user interface in the Arabic language. Greenstone software is capable of building and displaying Arabic content because the software is supported by the UNESCO, which encourages developing countries to participate in the current information revolution by adopting modern IT applications.

## Content

As content is considered to be the key to success for any digital library, further attention should be given to this issue in order to guarantee the achievement of design goals and the success of the created digital library. If the aim of the constructed library was to serve research workers with different background and languages, then creation of bilingual content would possibly be the right decision.  In this case, the display of users' interface in multiple languages is essential so that users

can browse and search the constructed library in both languages with ease. By so doing, the constructed library would be effective and efficient in providing a variety of electronic services to targeted end-users. The content of the digital library may vary according to the specialisation of the organisation and the area of interest of targeted end users. However, the following collections could be part of any special library or information centre: statistics; standards and specifications; publications of the organisation; reports; patents; technical reports; dissertations and theses; e-journals; videos and galleries; and bibliographical data.

### Library catalogue

The standalone CDS/ISIS database could be converted into a web interface enabling users to browse and search the library catalogue from anywhere, with open access resources and links to free portals in the field of interest.

## Content creation

This step covers the digitisation processes and prerequisite selections and decisions, in addition to the subsequent manipulation and management processes. (Dawson, 2003). A variety of decisions must be made on various aspects before starting the actual design. For example, whether OCR should be carried out on text files held as image or text. If OCR as image is found to be feasible it must thereafter be carefully checked before dissemination. Choice of formats is another challenge - which file format should be used to display the content of the library. Decisions also have to be made on other aspects, such as whether to use XHTML or rich text to display plain text. Another issue is related to PDF files, that is, whether to use PDF or another plug-in such as Ms Word plug-in for displaying documents. PDF plug-in is ideal for retaining complex layouts. "There are also various formats for image and even more choices. Whilst JPEG is common format, PNG is spreading widely, and TIFF is possibly recommended for printing or preservation" (Dawson, 2003). These issues will be discussed in the digitisation stage.

## Digitisation

Before starting the digitisation process, it is crucial to find out what types of collection are to be digitised, and available resources that will be part of the collections and comprise a great deal of their content. Suggested collections have to be divided into two groups. The first set comprises those already in electronic form (digital-born) or converted to digital form. The second set comprises traditional documents (printed). Suggested materials should also be separated into different sets according to their language.

## English collections

Materials in English may already be in digital form, but in different formats, e.g. PDF, rich text, and Ms Word. It is worth mentioning in this context that a decision has to be made before starting to design the digital library whether to display the content of the library in the original format of the document or in HTML format. It is desirable to display the library's content in more than one format to give users the choice of viewing the content in various formats. PDF format is popular for downloading purposes. Traditional print documents, scanned using Optical Character Recognition (OCR), would be easily displayed in HTML without any problem. However, the Greenstone system

is not capable of retrieving such materials when a search is performed. To overcome this problem all materials converted into PDF using OCR would have to be transformed into searchable PDF using a transformation program such as ABBYY PDF Transformer 2.0, Adobe PDF professional or Nitro PDF professional. After transformation, Greenstone software would be able to retrieve target content when a search was performed. Born-digital materials in PDF format should also be converted into searchable PDF format in order to be retrieved. Regarding documents in rich text and Ms Word, no problems would be faced in displaying such documents as HTML.

## Arabic collections

Considerable problems could be faced in the process of building Arabic collections. In fact, the problem with the Arabic language in this respect is due to the weakness of the Optical Character Recognition system in this domain. Born-digital documents in the Arabic language are easily displayed and retrieved as HTML by the Greenstone system when a search is performed. Documents scanned and converted either to PDF or Ms Word can be neither displayed nor retrieved by Greenstone. Moreover, in this case the Greenstone system is not capable of performing a full text searching. This means that the Greenstone system is unable to extract words and terms occurring in the text if digitised materials are entered as scanned images. Despite the availability of many specialised OCR programmes that claim the capability of converting Arabic from digital image to digital text format such as Readris 11, none of these programmes is capable of dealing with Arabic characters and of converting documents 100 percent. Experience has shown that no programme in this domain is able to give reasonable results. Even the retrieval of digital-born documents in PDF format by the system is difficult when a search is performed because they are simply not searchable. The Greenstone system cannot recognise and search characters that are not written in ASCII code, which is a standard code used so that data can be moved between computers that use different programmes. It is worth mentioning in this regard that born- digital documents in Arabic can be displayed and retrieved by the Greenstone system provided they are in Ms Word, rich text, or searchable PDF. To overcome problems associated with the retrieval of objects in PDF format, documents of such types should be especially reviewed and more attention should be given to their associated metadata so that users could exploit enabled browsing facilities instead of search facilities for retrieval purposes. Developers should therefore avoid scanning Arabic content using the Optical Character Recognition system. Nevertheless, conversion of Arabic content already in digital form from rich text or Ms Word to PDF has proven to be practical. As regards the format for image, there are also various formats for image and even more choices, as mentioned previously.

## Information organisation: classification and indexing

In order to make searching more reliable and allow users to browse across collections there is a need for a controlled method of information classification. Therefore, Library of Congress subject headings could be used as means of linking diverse collections into a coherent information structure, in addition to Arabic subject headings as a primary means for linking collections in the Arabic language. The terms of both subject headings could be included in the metadata and used in contexts where international compatibility is required. In order to keep consistency among collections, controlled vocabulary for place names and organisation names, as well as for technical

terms, should be used where appropriate as authority files e.g. names of persons, names of concessions, and names of areas.

## Metadata

Regarding metadata, the question of which metadata standard to use and implement is a key issue. Because of the availability of diverse metadata standards, there is a need to find out which standard could best serve the purpose of building the collections of the proposed library. The Dublin Core metadata standard is expected to meet the requirements of building the required collections. However, the standard should be tested and some field elements may be added to correspond with the methods of information retrieval by target end-users and to make the retrieval process more efficient and effective.

## Authentication or access control

This issue is concerned with policies for controlling access to different types of content and therefore it raises various questions. For example, should all digital content be accessible for the public over the internet? Is it necessary to control who can access specific collection and make use of it? Should the content of the proposed digital library contain sensitive documents that would harm a company's business if such documents were accessible over the internet? In fact, such issues have to be discussed and studied before starting the actual design of the digital library. Greenstone software has a built-in access control mechanism that allows collections and even individual documents to be restricted to authorised users using a password protection scheme. This mechanism can be applied if the organisation prefers to apply restrictions on access to some specific content. A policy for controlling access to different digital content should be written by managers of information units in companies or organisations seeking to build digital libraries and should take into consideration the negative and positive impacts of open access.

## Standards

The core standards that should be applied to the proposed digital library are Dublin Core and MARC 21 for metadata, Library of Congress subject headings and Arabic Subject headings for subject vocabulary, and AACR2 for resource descriptions. These standards are important for the consistency of resource description across collections and for assisting the process of information retrieval.

## Users interface

As user interface is considered an important issue and comprises an essential part of the proposed digital library the design of the prototype has to take into consideration simplicity, consistency and flexibility. The main issue before starting the design of the user interface is to understand the main priorities from the users' perspective. Therefore, how the interface should be designed would depend on users' opinions and could be decided after conducting interviews with people in charge and with a number of potential users. Because user interface can be designed in different ways, it is crucial to understand users' primary preferences in order to produce a practical design. The Greenstone system has the ability to display multiple interfaces without difficulty. However, producing a practical design that enables users to browse and search collections on dif-

ferent interfaces is not an easy task, as both interfaces should have the same features to display collections effectively and efficiently.

## Prototype

After designing the proposed digital library, the library should be evaluated in order to discover any problems within the design. The prototype is useful in many ways. In fact, it is usually built to test the function of the constructed library and to solve unexpected problems. The objective of the prototype is to assist in building the full design.

## Evaluation

This is the last stage in designing the digital library. This phase is considered to be an important stage as the evaluation will assist in developing the constructed digital library and lead to the full design.

## References

Arms, W. (2000) Digital Libraries. Cambridge, MA:MIT press.

Ashoor, S. (2000) Planning the elctronic library- suggested guidlines for the Arabian Gulf region. The Electronic Library. Vol. 18, no. 1

BNET dictionary. http://dictionary.bnet.com/ [Online] [Cited: 20 January.2009]

Chowdhury,G (2000) Access to Information in Digital Libraries: Users and Digital divide . Journal of Information Science. 26, 2000, Vol. 5.

Dawson, A. Building a digital library in 80 days: the Glasgow experience. http://cdlr.strath.ac.uk/pubs/dawsona/ad200301.htm. [Online] [Retrieved: 24 January 2009]

Digital library federation(DLF).http://www.diglib.org/about/dldefinition.htm. [Online] [Cited: 21 January 2009]

Jeng, J.(2006) Usability of the digital library: An evaluation module. PhD dissertation.New Jersey:The State University of New Jersey.

Leiner, M. A working definition of digital library http://www.diglib.org/about/dldefinition.htm. [Online] [Retrieved: 16 January, 2009.]

UNESCO. World Summit on the Information Society (WSIS). http://www.itu.int/wsis/index.html. [Online] [Cited: 21 January 2008]

Witten, I and Bainbridge, D. (2003) How to build a digital library. San Francisco: Morgan Kaufmann, 2003.