DIGITAL LIBRARY AND ISSUE OF THE STANDARD AND BEST PRACTICES FOR DIGITIZING INFORMATION RESOURCES

M. Solihin Arianto*

Abstrak

A. Introduction

Many libraries of institutions of higher learning are developing their digital resources. The rapid advances in information technology and the exponential increase in digital information enable academic libraries to improve their range of services to their remote clients. In many cases, top priority is given to develop strategies and systems that will maximize access by students and staff to library's information resources and services.\(^1\) Through the digitized resources, the libraries can develop a program that allows authenticated users, regardless of location, to access the libraries' databases. Keeping in view the importance of digital resources, in recent years, some academic libraries are working to increase their pool of electronic information, whether these are electronic databases, full-text journals, or electronic books.

The digital library movement is the most perplexing and promising development for libraries since library automation began almost seven decades ago. Nevertheless, advances in digital library development require that libraries pay close attention to the digital production arena whether or not any local activity is occurring. In this regard, one example that is pointed out by Pace\(^2\) is how the demand for digital content predates any establishment of the best practices for the creation of such content. This indicates that everyday decisions, in today's library world, are based on the availability of the content, rather than on the

---

\(^1\) C. H. Hoo, _Managing the Library: Delivering an Effective Library Service in a Global Environment_ (Dewan Bahasa dan Pustaka Brunei: Bandar Seri Begawan, 2003), p.77

quality or longevity of it. It is true that the main goal of digitization of resources is to improve access to a central part of the scholarly literature and perform well preservation aspects in which the goal enable to be achieved by adherence to prescribed standards and practices. To emphasize the importance of the goal, Pace claims that local digital libraries, in their early years, tended to take off without paying much attention to best practices, standards, or the benefit of good business-model planning.

This paper tries to explore various opinions relating to definition of the digital library and the digitized materials, and the significance of standard and best practices in digitization of information resources.

B. Digital Library

Digital library is not easy to define when some scientists and professionals have different definition based on their own point of view and scholarly base. This is admitted by Haigh that “there is not one single definition of what digital library is.” Similarly, Cool also agrees that “there is no single agreed upon definition of what constitutes a digital library.” To show how difficult we define digital library, below we pick up two examples. Schwartz reports that students in digital library course found 64 different definitions. The difficulty is resulted in a lack of consistency between existing definitions. Another example is what Borgtnan says that the variety of concerns within the digital library research community also reflects the interdisciplinary nature of the topic.

The variety of discussions as for digital library definition will happen to date. Deegan and Tanner also have the same opinion

---


4 A. K. Pace, The Ultimate Digital Library, p.10


observing various digital products created by libraries as they write, "there are many different kinds of digital libraries creating, delivering, and preserving digital objects that derive from many different formats of underlying data, and it is very difficult to formulate a definition that encapsulates all of these".9

The matter of definition as mentioned earlier may confuse. In this regard, Cool10 avoids discussing the definition of digital library. In his research, he prefers to study types of digital libraries and their user communities in order to understand well of what digital library is. The research examines the vast range of projects, initiatives, and services categorized as digital libraries and then present a typology of existing digital libraries, along with their goals and objectives, and their intended user communities. He opines that digital libraries, like traditional libraries as well, have various types. For his study need, he groups five library types that all of them are referred to traditional library's concept. They consist of national libraries, state libraries, public libraries, academic libraries, and special libraries. Four of the first types take two examples each type. These types of digital libraries are a search result of the library and information science literature and of the World Wide Web on the subject of digital libraries. Whereas concerning special libraries' type, he differently treats it. He describes some of the special libraries/collection characterized as digital libraries that refer to information entities sponsored by special groups, such as government agencies, professional organization, non-profit organization, and private companies. The majority of special libraries he studies are affiliated by associations or not-for profit organizations. Within this category, he takes account of numerous special collections that have been developed by academic and public libraries, as well as museum.11

He comes to conclusion that digital libraries are changing information environment in significant ways. Some of digital libraries are valuable for providing remote access to familiar information organizations, such as public and academic libraries.12 At the same time, great strides are being made in the development of more specialized

---

10 C. Cool, "A Typology of Digital Library, p.64
11 ibid, p.64-65.
12 ibid, p.66.
digital collection as shown by special digital libraries. Regarding development of digital libraries, it will rest upon the better understanding of how to communicate about digital libraries and their user.

In the other part, Fecko\textsuperscript{13} views that digital library offers great potential benefits and its use is regarded as a challenging and valuable test domain. Some of the benefits give on the new horizon in developing of library. Firstly, it creates a new dimension of scholarship and education because users can search large amounts of information quickly. Second, based on the collection facet, the digital library can help to remote the divisions between the information rich and poor by providing universal access to information. She further mentions that digital collections encourage users to take a more active role in learning since they may directly seek out the information needed rather than having to access a broad set of information that may be potentially useful. In doing so, borrowing privileges and physical location are not matter in digital libraries. Next, it provides the core essential functionality since it serves as structured repositories of multimedia documents. The documents can be added to and retrieved from the library.

In the different way, William Arms gives definition of the digital library as "a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network."\textsuperscript{14} He also states that there are some beneficial things in terms of digital technology. Firstly, scientific works can be written by an easier way. Furthermore, the libraries may protect and preserve their collections easily and cheaply. In addition, most people may communicate in an easy, cheap, fast, reliable way. This becomes true through e-mail and Internet. These entire indicate that digital information can be easier required from time to time.\textsuperscript{15}

Additionally, Michael Lesk defines the digital library as "a collection of information that is both digitized and organized, gives us

\textsuperscript{15} ibid, p.1.
powers we never had with traditional libraries.” He further points out that building digital libraries is not just a question of piling up disk drives. It involves creating an entire organization of machines and people, perhaps a culture, in which people are able to find information and use it. In this regard, he claims that the digital library will be useless if user will not use and utilize it.

Based on various definitions mentioned above, we can conclude that, in nature, it is not easy to find the definition of digital library in general. However, the author prefers to use the definition prescribed by the Digital Library Definition (DLF) because various works, both articles and books, discussing on digital libraries refer to the DLF when defining on digital library. For example, Liu writes up an article entitled “Best Practices, Standards, and Techniques for Digitizing Library”, Shiri authors “Digital Library Research: Current Developments and Trends”, and Cleveland also writes on “Digital Libraries: Definitions, Issues and Challenges”. In the same way, three books written by Pace The Ultimate Digital Library: Where the New Information Player Meet, Tennant in Managing the Digital Library, and Hughes Digitizing Collection: Strategic Issues for the Information Manager also refer to the DLF’s definition.

The DLF defines the digital library as follows:

“....organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of

17 M. Lesk, Practical Digital Libraries, p.2
collections of digital works so that they are readily and economically available for use by a defined community or set of communities."

The definition mentioned by the DLF above involves three key components that pose the theoretical framework underlying digital libraries namely (1) people, (2) information resources, and (3) technology. Therefore, the digital library can be meant as a managed collection of information and associated services. The information had been collected will be stored in digital formation and accessible over network. At the same time, various types of users who come from different background are able to access into the information needed via internet network.

C. The Digitized Materials

Regarding the digitized materials, there is a rapidly increasing volume of information, which exists in digital form. Whether created as a result of digitizing non-digital collections, created as a digital publication, or created as part of the day-to-day business of an organization, more and more information is being created digitally and the pace at which it is being created is accelerating.

The digitized materials may consist of selected articles (conference papers), newspaper articles, pictures/images, multimedia resources, photograph of artifacts (scanned and download from Internet, etc.), manuscripts, personal collection, artworks, examination questions, important information resources for classes not available in the local library, out of print materials, special letter of well-known personality identity, and background sounds and music.

In general, Tedd and Large classify digital information into four major groups as follows:

1. Full-text material that cover e-journal, open access and open archive collections, e-books, and e-newspapers.

a. *E-journals.* The majority of e-journals still have linked printed counterparts although the content may vary between the printed and the digital version.

b. *Open access, e-print collections, and open archives.* Open access is a system of providing users access to the full text of quality, peer reviewed research articles, which uses a funding model that does not charge users or their institutions for access. In addition, the term e-print is used to cover the electronic version of the reprint of articles as well as the electronic postprints or version that have been published in peer-reviewed scholarly journals. Besides that, the phrase 'open archive' is used to describe reprints of articles published in the area of high-energy physics.

c. *E-books.* E-books are considered important sources in many digital libraries. These collections are now predominantly available from companies, which license their use via the web.

d. *Electronic newspapers.* With the advent of the web, it was natural for newspaper publishers to make their product also available on it. As with e-journals, web editions of newspapers do not always contain the same information as their printed counterparts and often provide more features such as links to related news items and so on. Some digital libraries include news clipping services for users to be kept up-to-date with current affair.

e. *Theses and dissertations.* Concept of the theses and dissertation development in electronic version was first aired at a meeting in 1987 run by UMI (formerly known as University Microfilms). In order to support the development Electronic Theses and Dissertations (ETD), a Networked Digital Library of Theses and Dissertations (NDLTD) is being developed at Virginia Tech in the US (http://www.ndltd.org).

f. *Archives.* In the archives area many local and national archives are beginning to move from providing computer-based catalogues and finding aids to their collections, linking to digitized versions of archives.

2. Metadata sources that covers catalogues, indexes and abstracts, and sources that provide 'information about information'.
3. Multimedia material
4. Websites

Taking into account Tedd and Large’s elucidation as mentioned above, we can identify that types of digital library collections consist of electronic sources including e-book, e-journal, e-magazine, e-newspaper, e-theses and dissertation, e-archives, e-manuscript, e-print collection, open access, metadata, website, software, game, movie or video, and image that are saved on electronic storage media such as CD, DVD, hard disk, or database server. These electronic sources may contain still images, moving images, or audio sequences, or some combination of these different media.

Based on economic perspective, digital materials provided by digital library will give economic advantages compared to the traditional library. Chapman and Kenney\textsuperscript{27} mention four assumptions below proving that the digital collection will become more cost-effective.

1. Institutions can share digital collections
2. Digital collections can alleviate the need to support full traditional libraries at the local level
3. Use will increase with electronic access
4. The long-term value of digital collections will exceed the costs associated with their maintenance and delivery.

Both further assert that those four assumptions will be true if electronic files have relevant content and meet baseline measures of functionality over time.

To know how a digital library can organize library’s electronic documents into an organized collection, or convert a print library into an electronic library, Catenazi and Sommaruga\textsuperscript{28} describe that there are five interface designs for digital libraries as follows:

1. \textit{Plain text interface}. This is the simplest and immediate method of presenting information. ASCII text is used, and no images or multi-media images are included.


2. **Image-based interface.** Most information in digital collections is presented as image, and only some portions, such as a table of content, may be converted to text.
3. **Hypertext-based interface.** Documents using this interface are presented and organized using hypertext links.
4. **Book-like interface.** Documents are presented in book format.
5. **Mixed strategy.** This combines two or more of the previously mentioned interface.

In the other part, Pettengill and Arago\(^{29}\) conducted research regarding some of digital collection problems that are encountered by public user of digital collection and internet accessible archives. Firstly, it is difficult to find information due to poor organization and lack of search tool. Second, it is lack of consistency in the presentation of similar information, outdated information, and obvious errors in grammar and spelling. Next, there are too many links empty or useless information. In addition, reorganization is frequent so that users are forced to guess where to find previously located references. Finally, the documents are not available in formats suitable for both online use and printing.

Deegan and Tanner also have the same opinion observing various digital products created by libraries as they write, “there are many different kinds of digital libraries creating, delivering, and preserving digital objects that derive from many different formats of underlying data, and it is very difficult to formulate a definition that encapsulates all of these”\(^{30}\).

To classify kinds of digital resources is more or less as difficult as defining of digital library. However, in brief we can say that digital materials range from relatively simple, text-based files (e.g. word processing files), to highly sophisticated web-based resources, which fully exploit the benefits of the technology (e.g. combining sound with images, the ability to link to other resources, the ability to interrogate the data). All of them come with digital signal or code that must be read electronically.


D. Standards and Best Practices

Comprehensive standards and best practices are currently starting to emerge, and ongoing work has deepened the understanding of the needs and requirements that must be met to carry out for digitization of library collections. However, various researches conducted in the area of standards and best practices for digitization of library materials are responded with different concepts. Some experts argue that standards and best practices are highly required to make sure the consistency, reliability, and longevity of digital sources. In this regard, they contend that thoughts of best practices as accepted and documented ways have been long established in the library, archival, and museum world. On the contrary, some remind that implementing standards and best practices may be counter productive because each library that digitizes the library materials has different resource abilities.

There are a number of reasons stated by some experts that the standards and best practices are very essential to consider. Hughes believes that developers of digital collection will discover many difficult decisions because they have to decide every aspect of the project, starting from the selection of material, cataloguing, displaying, archiving and preserving digital materials, technical specification about hardware, software, and file formats, to metadata. She further concludes that a single set of prescribed standards and rules will be very helpful to the project manager.

In addition, Tedd and Large declare that librarianship world has been very familiar to standards. Use of the second edition of the Anglo-American Cataloguing Rules (AACR2) and the Machine-Readable Cataloguing (MARC) record format is two good examples to prove that standards are very important for libraries. They agree if the standards should be emphasized in the digital library environment as important as they have been implemented traditionally for libraries. This is in line with what they state, “It is impossible to consider digital libraries without at least dipping into the critical waters of standards.”

The standards can espouse cooperative relationships and provide the

---

31 L. M. Hughes, Digitizing Collection, p.199.
32 ibid, p.200.
33 L. A. Tedd & A. Large, Digital Libraries, p.86.
34 ibid, p.85.
common bases from which individual developments may emerge. They can also bestow opportunities to share across institutional and political boundaries. In line with Tedd and Large's views about the importance of standard, a number of important institutions or organizations were actively involved in the development and promotion of standards relevant to digital libraries. Some of them were the Digital Library Federation (DLF), the Library of Congress (LC), the Consortium for Computer Interchange of Museum (CIMI), Museum Documentation Association (DMA), and the International Federation of Library Association and Institutions (IFLA).

Other scholars who view that standard as fundamental need to digitize library materials are Anderson and Maxwell. They point out that when the projects are created, there are two tasks performed at once i.e. digitizing the actual material for online presentation and preserving the material for long-term archiving. These tasks can be attained by referring to something known as standard. The standards provide the option of how the materials have to be displayed online and what formats are available so that they are not only fundamental for file formats but also for long-standing preservation of digital object.

The standards' application is not only necessary in digitization process, but it also makes sure that users can access the process result. By this way, the standards enable for the dissemination of knowledge that conform to specific guidelines for usage. It is true as they note that “standards were developed to ensure the seamless operation of any given product so that it can be interpreted, manipulated, built upon, and recognized by certain protocols that allow users to view, interpret, and use the information that is available to them.” Like standards, best practices also play an important role to perform practices and procedures that can provide a safety net for practitioners as they do all aspects of digitization workflow.

In tune with the experts' views mentioned above, McDonough asserts that standards are very substantial to apply in digital libraries. He believes that with the spread of the Internet, numerous communities suddenly confronted both the opportunity to exchange information on

---

36 ibid, p. 156.
37 ibid, p. 67.
a scale never before realized, and the necessity for establishing standards for encoding content in order to exchange information more sophisticated than a HTML page.\textsuperscript{38} As central theme on his article, he elaborates that Metadata Encoding and Transmission Standards (METS) standard can be considered as one of many efforts to try to determine how complex sets of data and metadata might be best encoded to support both information exchange and information longevity.\textsuperscript{39} Furthermore, he concludes that the METS provides digital libraries with a practical and flexible packaging mechanism for digital objects to support their long-term preservation and promote the interoperability of digital library objects between different repository systems.\textsuperscript{40} However, METS is not certain to provide a guarantee of interoperability and there are some obvious practical difficulties in using METS for the long-term preservation of digital objects. Actually, those issues are not unique to METS; many of those working on similar standards for complex digital content used in research and higher education are encountering the same issues that digital library developers employing METS are experiencing.

Unlike the scholars mentioned earlier, some of other experts claim that the standards and best practices for digitization of library materials have shortcomings to apply. For example, Ron Zweig underlines and admits that best practices may mean delaying a project until technical standards are agreed upon, until large sums of money are available for digitization, or even until new technologies are invented. He further affirms that implementing “best practice, or even good practice, may be counter productive, and it may even constrain the creativity or innovation sometimes required to make thing work on a budget.”\textsuperscript{41} In addition, Haigh\textsuperscript{42} declares that constant and unpredictable technological change, coupled with a staggering a number of new and proposed standards emerging from a variety of communities,

\textsuperscript{39} \textit{ibid}, p.149.
\textsuperscript{40} \textit{ibid}, p.155.
\textsuperscript{41} L. M. Hughes, \textit{Digitizing Collection}, p.202
have created a confusing and complicated standards environment. Picking up two examples for descriptive metadata such as Dublin Core and MARC, she considers that given the plethora of standards to choose from, the wide range of applications and interpretation many support and the sometimes prohibitive cost of standards-based approaches, libraries can be uncertain which standards to embrace.

Similarly, William Arm also acknowledges there are some inconveniences in implementing standards for digitization of library materials. He opines that adoption of shared methods will supply digital library with extra functionality. However, it is also expensive because there are expenditures related to directly financial such as, purchasing equipment and software, hiring and training staff. Additionally, introducing a new standard requires interrelated changes in existing system, changes in the flow of work, changes in relationships with suppliers, and other changes. To take the opposite Arm’s view but still pertaining to the standards’ shortcoming, Ted and Large comment that the standards may retard innovation by maintaining outworn technology and safeguarding established practice.

However, examining existing guidelines will allow the extrapolation of the best and most recent standards currently available that can be modified to fit the intended purpose, institution, and budget. Using community-accepted standards also allows the possibility of collaboration with other institutions in the future. Moreover, digitization of library materials needs to follow a number of standards in order to achieve the objectives of interoperability among the various resources and information systems as mentioned earlier. This is in line with Chowdhury and Chowdhury’s view in which they reveal that digital libraries are necessary to adhere to a number of standards for various activities. The standards are very essential because they are useful at different levels of digital library design and development. In the other part, both add that standards are much needed when digital library design links with differences in computer systems, file structures, formats, information organization and retrieval features of the various

---

43 W. Y. Arms, *Digital Libraries*, p.208
44 L. A. Tedd & A. Large, *Digital Libraries*, p.86
information systems or collections that are accessible through a digital library.\textsuperscript{46}

Both also remind that the web has emerged as the prime means of delivery information in digital library, behind the web interface the integration of systems for seamless access depends on the use of agreed standards, implemented in agreed ways.\textsuperscript{47} Application several standards are also highly taken by digitization project into account in order to meet the need of distributed digital library for text and multimedia file formats, as well as indexing, and storage. Regarding the best practices, Hughes\textsuperscript{48} considers we may use any best practices as long as they are intended to create the best result by avoiding redundancy, adding value to the resources created and ensuring that digitization creates content of the broadest use and appeal to multiple and diverse audiences. A simple, local solution can be applied in the way that the local standards and best practices carry out processes to check that work is being done to agreed standards, by regularly examining work that is undertaken, overcoming potential barriers, and evaluating outcomes.\textsuperscript{49}

E. Conclusion

Many institutions around the globe such as libraries, museums, achieves, information centres are developing their digital resources. The rapid advances in information technology and the exponential increase in digital information enable the institutions to improve their range of services to their remote clients.

The digital library is meant as a managed collection of information and associated services. The information had been collected will be stored in digital formation and accessible over network. At the same time, various types of users who come from different background are able to access into the information needed via internet network. Whilst for the digitized materials is simply signified as digital materials range from relatively simple, text-based files to highly sophisticated web-based resources, which fully exploit the benefits of the technology.

\textsuperscript{46} G. G. Chowdhury \& S. Chowdhury, Introduction to Digital, p.70
\textsuperscript{47} ibid, p.104
\textsuperscript{48} L. M. Hughes, Digitizing Collection, pp.200-201
\textsuperscript{49} ibid, p.203
It is essential to apply the standard and best practices for digitizing library materials. The guidelines are highly required to make sure the consistency, reliability, and longevity of digital sources that allow the possibility of collaboration with other institutions in the future and achieve the objectives of interoperability among the various resources and information systems.
BIBLIOGRAPHY


