

PRINCIPLES FOR DESIGNING A DIGITAL LIBRARY

Umesh M., Satish M.V.

JRD Tata Memorial Library,
Indian Institute of Science, Bangalore- 560 012.
E-mail: umeshr@library.iisc.ernet.in
satishmv@library.iisc.ernet.in

Abstract: This article covers some basic principles underlying the design, implementation and maintenance of any digital library. These principles apply not only to conversion projects in which analog objects are converted to digital form, but to digital libraries in which the objects have always been in digital form ("born digitally") and to "mixed" digital libraries in which the objects may be of both types. The key challenges in front of designing any Digital Library are building the Resources, Interoperability, Intellectual Property and Effective Access. These principles acts as a base while building a Digital library.

Key words: Digital library, Digital library Principles,

1. Introduction

Digital libraries aim at unhindered access to content over computer and communication networks, and digitisation may be taken as a visible proposition to enhance the shelf life of non-digital content by preservation apart from the virtue of increased and easy access, thereby furthering usage. As a fresh, lively and dynamic area with a lot of enthusiasm and activity by researchers from different disciplines, institutions and countries, digital libraries are viewed in different perspectives and the single most development that has brought about sweeping changes in the library and information discipline currently in the developed world is that of digital libraries[1].

2. Principles

The purpose of a digital library is to provide coherent organisation and convenient access to typically large amounts of digital information. The following 10 principles are helps to design and continued development of any digital library system [2].

- Expect change
- Know your content
- Involve the right people
- Design usable systems
- Ensure open access
- Be aware of data rights
- Automate whenever possible
- Adopt and adhere to standards
- Ensure quality
- Be concerned about persistence

2.1 Expect change

It may not be apparent why the changing technology landscape is such a thorny problem for digital library projects. Consider, for example, a

conversion project in which documents are converted to some digital format. If the chosen format is part of a proprietary system, viewable only through a proprietary interface, when the company that markets the interface no longer supports the system and format, the digitised documents are all but lost. Consider, too, a scenario in which a document is created in a particular word processing programme and the document is attached to an email message sent to a notable person. Suppose the goal is to preserve all of that person's email messages for future generations. We are all too aware of our dependence on our email technology for reading such attachments. Imagine what today's platform limitations will mean to future generations, when the content of the attachments is likely no longer accessible.

Although the Internet, together with the Web, has made digital libraries possible, this fact may also contribute to unforeseen problems if library designers depend too much on today's paradigms and tools. They might be tempted to create a Web site with Hypertext Markup Language (HTML) pages and Web-accessible digital images of objects and documents—that may all be obsolete when HTML changes or is superseded by something else. Changing technologies can quickly outpace the ability of designers to maintain a particular digital library. An approach that anticipates and plans for change is needed to provide lasting access to its information.

2.2 Know your content

For users, content is the most interesting and valuable aspect of a digital library. Creators of digital libraries need to manage and make decisions about their content, including selecting the objects to be included, digitising items that exist only in analog form, possibly marking-up items using standard languages like the Standard Generalised Markup Language (SGML), and assigning metadata describing the content and other attributes of each object.

It is important for developers to decide on the nature and number of metadata elements early in a project. Although some elements may be added over time, significant costs might be associated with assigning metadata retroactively to already tagged and cataloged items in a collection. Some metadata elements describe the content of an item, including, say, its title, creator, date of publication, and subjects discussed. Other elements might be assigned for managing the collection; examples include scan status, quality-control status, and internal notes, as well as the technical aspects of the digital objects, such as file format and size.

Important, too, is deciding on the basic conceptual units, or objects, the system will include, such as individual documents, photographs, videos, or lab notebooks. This decision affects the level at which metadata is assigned (for example, to an entire book or to each chapter in the book) and how the materials are organised, accessed, and archived. Sometimes overlooked is the practice of assigning each conceptual object its own unique identifier linking it to its metadata record and to other objects in the collection.

2.3 Involve the right people

Ideally, individuals from a variety of backgrounds and offering a variety of expertise contribute to building a digital library. In practice, this may not be the case, but even when it's not, knowing that building the system requires insight from a number of fields yields a better digital library.

The two fields involved most directly are computer science and library science. Computer scientists appreciate the possibilities, as well as the limitations, of technology and are generally the ones who actually build the system. Librarians, including catalogers, indexers, and archivists, have long been the custodians of information resources, understanding not only the information needs of diverse audiences but the issues involved in preserving materials for continued access and use. Digital library research and development have meant that each group has had to understand the other groups' perspectives.

Illustrating the importance of multiple perspectives is the development of the Dublin Core metadata standard. Computer scientists are concerned with the semantic interoperability a digital library metadata standard affords in the very large Internet information space; librarians already have deep experience indexing and cataloging and recognise the importance of these concepts for information retrieval. Moreover, because so much valuable data exists in a variety of metadata systems, including the Machine Readable Catalog (MARC 21) standard, methods to map the data between and among these systems using automated "crosswalks" have been developed.

2.4 Design usable systems

Most digital libraries are made available over the Internet through Web technology, though, strictly speaking, this is not a necessary attribute of a digital library. However, as the advantages of the Web are so great, most library systems today are designed to be Web-accessible. The most successful Web site designs account for a number of factors, including the technical differences among computers and browsers, including speed of access, and differences among users, including Web navigation preferences. Browsers differ in the way they display information, even though they use the same basic communication protocols (such as the Hypertext Transfer Protocol, or HTTP, and File Transfer Protocol, or FTP) and standard markup languages (such as HTML and perhaps the Extensible Markup Language, or XML)[3]. Since users may change default settings, including font size and other parameters, it is always preferable to create as simple an interface as possible and avoid server-side control of the exact display of the data. Providing multiple access points not only makes a digital library more interesting, it also acknowledges the differences among its potential users.

Accessibility for users with a range of physical disabilities should also be a concern when developing the interface to a digital library. This includes user access to all content; documents that are clear and simple; user control of

styles; the availability of context and orientation information; inclusion of clear navigation mechanisms; and standard markup.

2.5 Ensure open access

Ensuring open access is closely related to usability concerns, including access to the information in the digital library, as well as to the digital library itself. Christine Borgman defines access to information as "connectivity to a computer network and to available content, such that the technology is usable, the user has the requisite skills and knowledge, and the content itself is in a usable and useful form". Michael Lesk writes that open access to information raises a number of public policy issues, including whether or not all segments of society are given equal access to information.

One way to ensure open access to content is to avoid proprietary hardware and software solutions whenever possible. That is, while it may be reasonable to create content using commercially available systems and tools, avoid requiring special software or hardware to access that content. A number of companies, while charging for the tools used to create digital images, make their interfaces available for free. As long as the software is easy to download and install, and the developers of the digital library make it clear which software is needed to view the images, the content will be accessible, as long as the interfaces are available. In all cases, however, for continued accessibility and use, open, nonproprietary systems are preferred.

2.6 Be aware of data rights

A possible threat to open access to information arises because of intellectual property concerns. Existing intellectual property and copyright law provides economic and legal protection to publishers of physical artifacts. "Fair use" (allowing libraries to make, say, single copies of portions of books or journals) and "first-sale" rights (allowing individuals to, say, lend or resell copies of books they have purchased) have promoted greater access to physical artifacts than might be possible otherwise, but these notions are only indirectly applicable to networked information.

There are no straightforward answers to the enforcement of intellectual property rights for information available in digital form. The Internet and Web have emerged from communities that believe in sharing information, rather than restricting access to it. This has led to the perception, and perhaps even expectation, that anything available on the Web is freely available and may be redistributed at will. Some argue there will be a realignment in the way intellectual content is published and distributed worldwide (for example, so authors are the primary "publishers," largely replacing existing commercial publishers). Others argue for technical solutions to restrict illegal copying. Still others argue that copyright law needs to be strengthened to prevent unauthorised use of digital information. And also the legal protections should come from contract, rather than copyright, law, so data is licensed, rather than sold.

Some resolution to these problems will undoubtedly arise in the coming decades. In the meantime, however, those creating digital libraries need to be aware of the issues, participate in the public debates about their resolution, and establish procedures to manage them in order to protect their collections to the extent possible. For example, in conversion projects, every attempt should be made to seek permission from the copyright owner for the materials that are to be digitised. Privacy issues should also be considered when, for example, the full papers of a prominent individual are being digitised. Ideally, the donor will have marked the items that are sensitive in some way and left instructions about how they should be handled. In all cases, however, careful tracking of permissions and privacy information ensures the collection will not be at risk at some future time.

2.7 Automate whenever possible

Because building a digital library requires significant intellectual effort on the part of the system's creators, the more automated tools that can be built and used, the better will be the use of precious human resources. These tools need to be easy to use and incorporate real-time aids, including data validation, pull-down lists, report generation, and other time-saving devices, thereby allowing the content expert to concentrate on the intellectual tasks at hand. Content experts use the metadata entry system to add metadata to a master database, entering the information only once. Subsequently, the information is extracted and combined as needed from the master database to generate HTML pages, search indexes, and reports. Entering the data only once saves human time and effort, reduces the error rate, and allows maximum flexibility. Nearly the entire Web interface is generated from the database, allowing regeneration whenever necessary, while adhering to the latest Web standards. The system is designed to be modular, allowing existing modules to be modified easily and new modules added for additional functionality.

The relationship between the underlying master version of the system's data and its derivatives. High-quality lossless TIFF images, for example, serve as archival copies and as master copies for creating a variety of derivatives, such as PDF or JPEG files, for optimal use on Web sites. Then, as better compression algorithms are developed, it will be easy to discard the derivatives (precisely because they are derivatives) and generate a newer file format from the original master.

2.8 Adopt and adhere to standards

The use of standards in system building has many benefits. Applications are more readily scalable, interoperable, and portable; these characteristics are all important for the design, implementation, and maintenance of digital libraries. Using standards is especially important for the aspects of digital libraries that are most labor-intensive. Scanning, metadata entry, and document markup, all involving the evaluation and handling of individual items in a collection, are resource-intensive and best done carefully and only once. Data might still have to be migrated to other forms and formats in the

future, but migration will be easier, because standards have been used consistently. Images scanned using standard file formats, such as TIFF, or texts saved in the open ASCII or Unicode formats will be more easily accessible in the future than images or texts encoded in proprietary formats[4].

The benefits of using standards for interoperability should be clear to any developer. For example, if all finding aids for historical collections are marked-up in SGML using the Encoded Archival Description document type definition, one can easily imagine distributed access to all the finding aids held by the members of a library consortium. Delivering the contents of the digital library on the Web, using standard, valid, and current HTML, including metadata tags, and other standard Web technology, increases the chances that other Web search engines will be able to find the library, as well as the specific items in it.

Because technology is changing so rapidly, it is important to question whether it will be possible to port an existing system to another platform, user interface, or software. In principle, while porting should always be possible, in practice it is much easier when the system's design and implementation adhere to certain standards. It's also possible that sometime in the future the entire digital library will be part of another more-encompassing system, either within or outside the original institution. The transition will be smoother and more successful if the system is designed with open standards in mind.

2.9 Ensure quality

Quality can be applied to all the processes and outcomes involved in creating a digital library. They are relevant to selection, metadata entry, image capture, and the overall usability of the system. Complete and correct metadata yields many benefits; incomplete or incorrect metadata affects the quality of the entire digital library. Metadata plays a vital role not only in resource discovery but in managing the collection. If, for example, subject codes are applied haphazardly or incorrectly, access could be more difficult, and attempts to generate browse hierarchies based on these codes could be foiled.

If scanning procedures and guidelines in conversion projects involve immediate review and evaluation of the scanned images for appearance, including orientation, resolution, color, and tone, there will be fewer future problems. Images that are skewed, dark, or cut off are not welcome in a digital library. Digitised video and audio need to be reviewed periodically for adherence to evolving standards and for their ability to be viewed and heard using current tools[5].

Some the quality-control metrics can be automated; others require careful human review. Digital library projects should define and then carry out quality-control methods as part of their normal procedures. Adhering to such

methods ensures that quality assessment becomes an integral part of building and maintaining the digital library.

2.10 Be concerned about persistence

Digital Preservation is gaining importance today. At present there is no way to guarantee the preservation of digital information. While preservation has long been a concern of archives and libraries, it has only recently been of interest to a much larger community.

Anyone creating a digital library has a stake in the outcome of preservation approaches, yet only some address the problem directly. Several suggestions have been made by researchers to address preservation of digital objects. Perhaps the one discussed most often is the "migration strategy,"[6] which entails the transformation of data from one file format to another, converting it from one software environment to another, or moving it from one physical medium to another. Migration implies a strong and enduring commitment on the part of an institution to continually refresh its collections to keep up with the technology. Another proposed method is the "emulation strategy," which entails the emulation of an entire software system so it runs on future unknown systems.

3. Conclusion

When creating digital library systems containing valuable content, the principles and the practices that evolve from them places us in a much better position to keep these promises. Valuable content should be handled with care and rendered in the highest quality possible. Valuable content should not disappear. We need to understand how to preserve and safeguard digital material, so it doesn't become obsolete simply because we didn't pay attention.

4. References

- [1] Elisabetj,J.show.(2000). Building a Digital Library: Atechnology managers point of view. *The Journal of Academic Librarian Ship*, Vol.26 No 6, pp 394-398
- [2] R.K. and Vishwanthan K.R.(2001). Digital Libraries: development and principles, *Library Review* , Vol 50 No 1, pp 10-16
- [3] Rob kling and Margaret Elliott(1994). Digital library Design for organisational usability. *SIGOIS Bulletin*, Vol.15 No 2
- [4] National Information Standards Organisation. The Dublin Core Metadata Elemen Set: Draft Standard S39.85; see www.niso.org/S3985.html.Sharma
- [5] Lesk, M.(1997). Practical Digital Libraries, *Books, Bytes, and Bucks*. San Francisco: Morgan Kaufman Publishers.

- [6] Marilyn Deegan and Simon Tanner.(2002). *Digital futures: Strategies for the information age*. London: Library Association Publishing.