Digitizing Era of Library

*Neeta Samagh, **Rachna Dewan
*Sr.Librarian, G.C.W, Bhodia Khera
**Sr.Librarian, District Library, Sirsa
nitisamagh87@gmail.com, rachnadewan@gmail.com

Abstract: A digital library is "a world of literature, history, photographs, movies and maps open, free of charge, to any curious mind that wants to meander through the electronic equivalent of library stacks. a focused collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization, and maintenance of the collection.
Keywords: digital library, texts, digital objects, era.

[I] Digital library

In the full definition a digital library would be one in which all the texts and the spoken books would be held as digital files. But that will take a long time to achieve. Let us start by considering text files. If you have a text as a word processor file on a floppy disc or the hard disk you have the basis of the digital library. What are the advantages of the digital file? It takes up very little room. One can make safety copies in seconds. From the digital files on can drive a printer - either for a standard printout or for large letter. One can even define the needs of particular clients and produce a copy that is of the size and typeface that the partially sighted can best read.

As you all know there are, in many languages, software programmes that can automatically translate from the original letters into braille - many of these also contracting the braille to fit the particular rules of a country or language. Once the text has been passed through this software one has yet another digital file, but now in a code which generates braille characters rather than letters. It is important to keep this in mind since we will be coming back to these digital braille files.

The digital braille file can be used to drive a braille printer. Suddenly we can break away from the former restriction of whether to print a whole book. If you are hand-making a book the tendency has been to make the whole book. Most libraries have rules about producing the integral text of a book. But if we regard the creation of the digital braille file as the act of production, so that we have the integral text should we need it, then we can save time and money by printing off in braille only those parts of a book that are needed. This does not apply to novels but may well apply to a study book or a reference work. Over time the savings can be considerable. Remember that the text and the braille files can both be used in this way.

[II] Introduction

If you could guarantee that all your clients had computers adapted to their particular limitations and that they had printers at home (and supplies of paper) to braille out materials that they wanted to read directly - then you might set up a totally electronic service. Not even the richest and most advanced country in the world can guarantee this level. A less wealthy country may have few, if any, clients with computers. But that doesn't matter. What the argument presented here suggests is that the libraries should go digital in order to produce traditional reading materials at lower costs and faster. Leaving aside audio books for the moment (though they will be dealt with in depth during the conference) let us stay with texts. How much time does it take to prepare a page of braille by hand? Any mechanical method is slower than using a computer (assuming a moderate level of operator skills). If libraries or the braille production centres start to use computerised methods then dramatic improvements can arise in the production of braille text. To change the infrastructure within a country is a very different question. However desirable it might be to equip all visually impaired people with audio players and adapted computers this is just not feasible. In countries without a technological infrastructure it makes no sense in the short term to assume that one can change the main forms of materials and the methods of distribution.

[III] Digital Production of texts

There are three main methods of producing digital texts. The simplest is to use a scanner and a good Optical Character Recognition program that turns the scanned text from an image into letters that can be handled by a word processing program. The accuracy of these OCR programs is now very high. The clearer the printing in the original the fewer mistakes there will be in the transfer. Of course one then has to run a spelling checker and proof read the resulting file. However if one uses a sheet feeder, whole books can be turned into digital files in tens of minutes rather than tens of hours. The availability of good software varies depending on the sort of alphabet that is used, of course.

The second and most common method is to copy-type the text. In libraries for the blind such a method is reasonably efficient. The text still needs to be proof read but a sighted typist can key in text using a conventional keyboard at speeds far in excess of traditional brailling. The third method is to get the text directly from the publishers as a digital file. In theory this should be possible. After all, the printer has produced the book from a digital file. In reality these files are difficult to find, are often full of setting codes and in any case the publishers are frightened that somehow the files will escape onto the open market. Experience shows

that it is often quicker to scan or re-key the work. As indicated above, once one has a digital file then passing this through a braille translator to arrive at a braille file which can then be used to drive a braille printer.

[IV] General Principles

1. The technical framework exists within a legal and social framework

Early networked information systems were developed by technical and professional communities, concentrating on their own needs. The emphasis was on making information available to colleagues and the public, without charge. The digital library of the future will exist within a much larger economic, social and legal framework.

For example, musical works represent the livelihood of composers and musicians. Their artistic reputations depend on their work not being changed in storage or transmission. They require payment, as do recording studios and concert halls. Such work will only be part of the digital library, if the library supports their interests.

The legal system's task is to codify this rapidly changing economic and social framework. The relevant areas of law include copyright, performance, and other intellectual property, libel and obscenity, communications law, privacy, and international law.

The Kahn/Wilensky architecture can not write the law, but it provides a technical design that matches the legal structure that is expected to emerge. The architecture respects the creators and owners of intellectual property. It allows the preservation of rights that can last for more than one hundred years, and recognizes that digital works may include material from many sources, with separate property rights.

Society expects the creators of works to be responsible for their content, and for those who make decisions about content to behave responsibility. However, the digital library will not thrive if legal liability for content is placed upon parties whose only function is storage and transmission. Therefore, the architecture establishes clear boundaries between the areas of responsibility of the various parties.

2. Understanding of digital library concepts is hampered by terminology

Terminology proves to be a barrier in describing a digital library. Some words have such strong social, professional, legal, or technical connotations that they obstruct discussion between people of varying backgrounds. Simple words mean different things to different people. For example, the words "copy" and "publish" have different meanings to computing professionals, publishers, and lawyers. Common English usage is not the same as

professional usage, and the versions of English around the world have subtle variations of meaning.

Certain words cause such misunderstandings that they are best expunged from any precise discussion of the digital library. The list includes "copy", "publish", "document", and "work". Other words have to be used very carefully and their exact meaning made clear whenever they are used. An example is "content".

In the Kahn/Wilensky architecture, items in the digital library are called "digital objects". They are stored in "repositories" and identified by "handles". Information stored in a digital object is called "content", which is divided into "data" and information about the data, known as "properties" or "metadata".

3. The underlying architecture should be separate from the content stored in the library.

A conventional research library stores more than books, and the digital library is the same. Almost every type of information can be represented in digital form, including text, pictures, musical works, computer programs, databases, models and designs, video programs, and compound works combining many types of information.

The underlying architecture of the digital library, as described by Kahn and Wilensky, specifies those characteristics that apply to all types of material. For example, every object needs to have a name or identifier; the actions of adding objects to the library or deleting them apply to all material; general purpose methods of security can be provided.

This underlying architecture is a base for extensions that can be tailored for various types of information. The extensions typically include specific formats, protocols, and rights management that are appropriate for the type of material. For example, the extensions for digitized movies will be very different from those for video games; texts are usually described by bibliographic terms, such as author and title, which are of little relevance to a computer program; a protocol designed for interaction with a database is unlikely to be useful in manipulating graphic designs.

Separating general functions from those specific to the type of content has other benefits. It encourages different markets to emerge, and allows a legal framework in which storage, transmission and delivery of digital objects is separate from activities to create and manage the intellectual content.

4. Names and identifiers are the basic building block for the digital library

Names are a vital building block for the digital library. Names are needed to identify digital objects, to register intellectual property in digital objects, and to record changes of ownership.

They are required for citations, for information retrieval, and are used for links between objects.

These names must be unique. This requires a administrative system to decide who can assign them and change the objects that they identify. They must last for very long time periods, which exclude the use of an identifier tied to a specific location, such as the name of a computer. Names must persist even if the organization that named an object no longer exists when the object is used. There need to be computer systems to resolve the name rapidly, by providing the location where an object with a given name is stored.

The Corporation for National Research Initiatives has implemented a handle system which satisfies these requirements. A "handle" is a unique string used to identify digital objects. The handle is independent of the location where the digital object is stored and can remain valid over very long periods of time. A global handle server provides a definitive resource for legal and archival purposes, with a caching server for fast resolution. The computer system checks that new names are indeed unique, and supports standard user interfaces, such as Mosaic. A local handle servers is being added for increased local control.

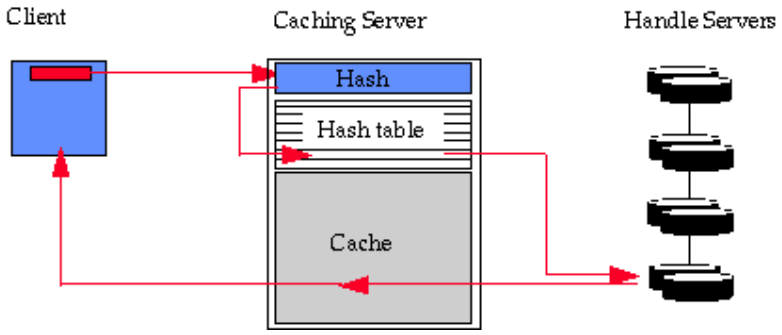

Figure 1. The CNRI handle system

5. Digital library objects are more than collections of bits

In the digital library, information is stored as "digital objects". A primitive idea of a digital object is that it is just a set of bits, but this idea is too simple. The content of even the most basic digital object has some structure, and information, such as intellectual property rights, must be associated with the digital object. Figure 2 shows that a digital object in a repository has two parts, content and associated data, sometimes called "metadata". To enable the content to represent useful information, its type must be known. Thus part of the content may be of type text (perhaps encoded in a mark-up language),

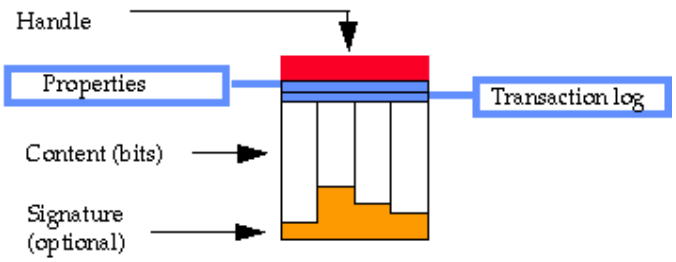

Figure2. Parts of a digital object while another part may be of type audio. A single digital object may contain many types of content. It turns out that arbitrarily complex data types can be constructed from a few basic types, notably bit-sequences, handles and other digital objects. By combining these in various combinations, any digital content can be represented.

To manage valuable intellectual property, certain metadata is required. This is shown in the figure. It always includes a unique identifier (the handle). It may also include properties such as rights and access methods. One property states whether a digital object is mutable, in that it may be altered after being placed in a repository. Another is a digital signature or other method of validating that an object has not been changed. Frequently, it is useful to keep a log of all transactions associated with each digital object.

6. The digital library object that is used is different from the stored object

In the digital library, what you store is not what you get. The architecture must distinguish carefully between digital objects as they are created by an originator, digital objects stored in a repository, and digital objects as disseminated to a user.

The user receives the result of executing a program on the stored object. This may be a simple program, such as a file transfer program, or something very complex. For example, an image is stored in a library as a set of wavelets. To use it, the stored wavelets are used to generate an image with the characteristics requested. This is transmitted over the network to a user's computer, where it can be further processed or displayed.

Some classes of digital objects can be provided it to a user in more than one way. For example, the score of a musical work is held in the library. One form of use is to transmit a representation of the score to the user's computer. Alternatively, the user could request the repository to execute a synthesizer program, which would perform the score, and transmit the digitally encoded audio over the network. For some types of object, such as a data base or a video game, the use consists of an interaction between the user and the execution of the program.

Legal scholars see an interesting parallel between the computer viewpoint of executing a program to supply a digital object to a user and the legal concept of performance. This may prove to be the correct framework for managing rights in a digital library.

7. Repositories must look after the information they hold

A repository stores digital objects, both the content and the metadata.

A digital object as stored in a repository may be very different from the digital object that is made available to users' computers. Different repositories will have very different internal organizations, but for each digital object every repository will have a properties record, which holds attributes of the object, and a transaction log.

Since digital objects contain valuable intellectual property, the stored form of a digital object within the repository includes information that allows for it to be managed within economic and social frameworks. The repository maintains this information, provides basic reference information, and provides security to ensure that only valid operations are carried out on the digital objects.

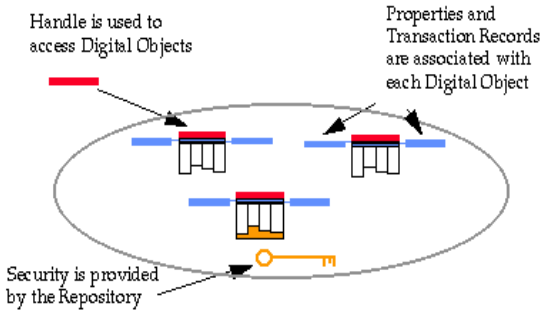

Figure 3. A repository

The internal organization of a repository and the way that digital objects are stored are hidden from the user. A simple protocol is provided for interactions with the repository. This protocol is called the "repository access protocol." The basic commands in this protocol are those to access a digital object and its metadata, and the service request to disseminate a digital object. In addition there are commands to add and delete digital objects.

8. Users want intellectual works, not digital objects

Digital objects are the basic building blocks of the digital library, but users of the library usually want to refer to items at a higher level of abstraction. Common English terms, such as "report", "computer program", or "musical work", often refer to many digital objects that can be grouped together. The individual objects may have different formats, minor differences of content, different usage restrictions, and so on, but certain users are willing to consider them as equivalent.

Which digital objects should be grouped together can not be specified in a few dogmatic rules. The decision depends upon the context, the specific objects, their type of content and sometimes the actual content. The underlying architecture has to support two main needs. It must provide methods for grouping digital library objects and must provide means for retrieval.

The Kahn/Wilensky architecture supports these higher level ideas in several ways. One is to have a digital object containing several digital objects. Thus several formats of a text might be assemble into a single digital object. Another approach is to have these variants stored as separate digital objects, each with its own handle. These handles are contained in a digital object, known as a "meta-object", which acts like a catalog record. It contains a list of the variants with their handles and information about the differences amongst them.

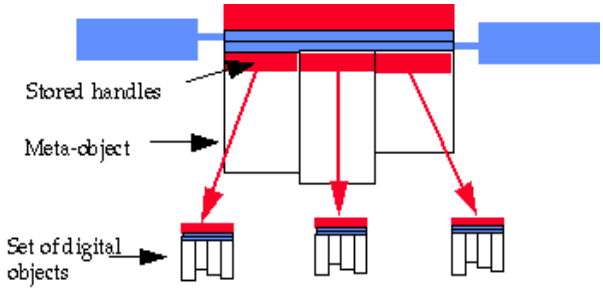

Figure 4. A digital object used as a catalog record

[V] CONCLUSION

A digital library collection, which, as noted above, may be logically defined as a set of criteria for selecting resources from a broader information space, may be less formally understood as a set of digital objects

REFERENCES

1. hdl:cnri.dlib/tn95-01 Kahn, Robert and Wilensky, Robert. "A framework for distributed digital object services". May, 1995. (http://www.cnri.reston.va.us/home/cstr/arch/k-w.html)
2. Key Concepts in the Architecture of the Digital Library William Y. Arms Corporation for National Research Initiatives Reston, Virginia warms@cnri.reston.va.us **D-Lib Magazine**, July 1995
3. Introduction to Digital Libraries hussein suleman uct cs honours 2009

4. Digital Libraries: Assumptions and Concepts DAVID BAWDEN AND IAN ROWLANDS Department of Information Science, City University London, UK.
5. www.google.com
6. www.wikipedia.in