99 Questions about Digital Libraries

Dr Peter Noerr
Technical Director
EduLib
peter.noerr@edulib.com

Abstract:
This paper discusses issues and technicalities surrounding the envisioning, planning and creation of a digital library. In general the paper raises issues for discussion, rather than providing prescriptive answers. The range and diversity of individual circumstances is too broad for generalizations and formulas

At the preliminary stage it addresses the issues of organizational motivation, possible content, target audience, long term support and funding.

For planning it asks questions about staff and resources, creation and maintenance, delivery methods, charging policy, and long term support and funding.

During the process of creation aspects to be considered are , available technology, in-house or out-source, project management, design, testing, unveiling, and long term support and funding.

Introduction

In the spirit of over-hyped advertising, I will not be attempting to discuss, or even to raise, 99 questions in the time I have. I will present a slimmed down set of nine questions and some discussion points about them. However, I feel completely free to suggest that anyone who wants more questions should obtain and read a copy of my “Digital Library Toolkit”. Here you will find today’s questions along with over 40 of their colleagues.

A number of copies of this document are available in printed form in this meeting and on the Sun booth in the exhibition. Sun commissioned and paid for the production of the two editions of this document. It is also available at the Sun and Edulib websites for download. URLs are given at the end of this paper.

I do not intend to offer any answers to my questions today, merely to raise the questions as points for discussion.

Structure

I would like to consider some of the aspects of forming a digital library at three stages: Preliminary, Planning, Creation.

A precursor to all of these is the question of what a Digital Library is. This one I have to single out for an attempted answer; even if only to set some sort of playing field for the rest of what I am going to say.

What is...

...a Digital Library?

Conventionally there are two possibilities: a library that contains material in digitised form; a library that contains digital material. The difference is sometimes very subtle.

The really important point is that a digital library has material stored in a computer system in a form that allows it to be manipulated (for instance for improved retrieval) and delivered (for instance as a sound file for computer playing) in ways that the conventional version of the material cannot be.

An automated library is not, per se, a digital library since a library consisting entirely of conventional physical material (such as only printed books) may be very highly automated. This automation does not make it `digital' in the sense we are considering here. However, it is true that a digital library must be automated in some of its essential functions.

Because the material is in digital (or computer readable) form, some new possibilities are opened to the digital library which are not there for a conventional library, even one with the same material.

...Digital Material?

In this computerised day and age information and the medium on which it is recorded can be considered as either digitised or not. There are many other ways of categorising the material, but computer readability is the important criterion here.

`Digital' can be taken as a synonym for `computer readable'. This is a serious generality, but it is this aspect of information which concerns us here.

Today increasing amounts of material are originally produced in digital form. These have, in one sense, no physical presence other than the computer file which is their original form. Thus there are a large number of ways they can be realised for human consumption, all of which are valid, yet all of which vary in the sorts of detail which abound in physical objects.

Preliminary Stage

What are the Motives behind the Digital Library Creation?

Although this is an important initial question, it is also true that the answer is almost irrelevant. The motives may be many, what is important is the organizational commitment to the Digital library. Like any undertaking, the grandeur of the plan is of little consequence if the execution of it is stopped before completion; all is then for naught.

Motives are many: a digital material collection, computer delivery, an expanding conventional library, availability of digital material, etc The important question is: Is the Digital Library going to be important to the organization as a whole. Does it fit squarely in its business plan, its method of operation, or its mission statement?

Obtaining open-eyed commitment to the project with all its possible short-comings and ‘Plan B’ outcomes, as well as the need for long term care and feeding, is more important than accurately determining the motives of the organization.

What will be in the Digital Library?

Having dismissed motives in the previous section, it is now the time to observe them in action again. Perhaps the Digital Library is being created as a “Showcase”,

"Demonstration", or "Prototype" forerunner in some grand plan. Many are. Too many are. The temptation here is to utilize material which looks good in a demonstration either to potential customers or partners, or in presentations at conferences such as this.

Thus we have many Digital Libraries which have their content defined for them with little thought to its utility to the intended audience (see next section). That is if an intended audience is considered and defined. We have many Digital Libraries which contain beautiful high resolution images of works of art. Who uses them on a regular basis? What do the clever zoom and pan tools that allow minute examination of the images do for their users? How does this relate to the average man on the web and a desire to read a good thriller?

Amount of content is another problem. Too many Digital Libraries are small, very small in terms of numbers of objects, and limited in scope. This compounds the finding problem. Just when the Internet and Z39.50 are showing promise of allowing more conventional, but automated, libraries to be searched high and wide, the 'next generation' Digital Libraries become more insular and specialized.

Who are we aiming the Digital Library at?

This leads directly to a consideration of who the organisation's Digital Library is aimed at. In the case of the "Showcases", "Proofs of Concept", "Technology Testbeds" there really is no audience, just as there really is little intention of longevity.

For any endeavour to survive it must have an interested audience which is willing to use and evolve with it. The audience is paramount for the growth of a particular Digital Library and for Digital Libraries in general.

Factors in the design of the Digital Library such as usefulness, ease of use, cost, etc will determine the popularity and possibly the long term future of a Digital Library, but if it does not have a targeted audience to start with it will have no future at all. Bad design or operational decisions will lose audience and use. Policy decisions ignoring the intended users will ensure there is no one to lose in the first place.

Planning Stage

What sort of Resources will be Needed?

The most important resource for the whole exercise is staff time and expertise. Although there is a lot of hi-tech and computers involved in creating and running a digital library, most of it is hard work.

There are a number of specialist tasks during creation which are candidates for contracting. You will need expertise now and not again. Such things as data and record conversion, and even the biggest skilled job of all to do; the cataloguing and indexing.

Resources for emergencies need to be considered and contingency plans (stand-by machine(s), access to temporary staff, etc.) need to be made.

Whatever other resources you have available there will never be enough time. Do not publicly set too aggressive a timetable. Allow time to produce things, test them, correct them, and then do it all again. Problems will arise and will set the timetable back. All that can be emphasised here is to plan as thoroughly as possible and to be conservative.

How will the Digital Material be Created?

This rather depends on where you start from. If you already have digital forms of material then its all done. Except of course for ensuring that the computer format is correct, that no data is lost during conversion, that resolution, fidelity, colour balance and the like remain true, that the data is compatible with the delivery system, and that it is properly catalogued and indexed.

Maybe you can by the records you need. This may avoid some of the above pitfalls, but not all of them, and, of course, you may end up with half your record in Spanish.

Create them your self and control the quality throughout. An excellent choice,; just remember the expensive equipment, the trained operators, and the fact that your beautiful images may end up as 1Gigabyte files and be essentially unreadable, at least within the lifetime the users are prepared to wait.

A combination of the above is most probable, but be very aware of the pitfalls (like file size and format) that flow from seemingly obvious decisions at the early stages.

What form of Delivery will be Offered?

The Web's the Thing! It may well be, but beware of the siren call of high technology in the form of fast computers and, specially, high bandwidth links.

This is really a plea to again be aware of your audience and to plan to be inclusive rather than exclusive. Try not to demand so high a price of admission that a large potential audience is left out.

It is even possible to run a very successful Digital Library with plain old fashioned paper copies delivered by snail mail. Aim for multiple access methods, trading off performance and functionality against simplicity and expense.

Creation Stage

Which of the Available Technologies to Use?

"Only the very best!" Yes, but best for who? One of the inherent problems in a Digital Library is that the material is digital. By the definition earlier that means 'computer readable'. And computers change. For those who haven't noticed, they change very fast. In particular things such as standards become obsolete and are superceded. This means all your data may have to be 'updated' every so often.

Digital Libraries are not permanent storage, they are transient and today's data will be unreadable without continual updating over any long period. Even the programs and the delivery systems will fade away on the sort of timescale that libraries normally think in. Believe it or not Microsoft's Windows will not be the dominant user interface in 5 or 10 or 15 years, it will be as interesting an antiquity as DOS is now (that is DOS for PC's not IBM's DOS, which is a different form of antique).

You have to decide if long term stability is important, or can you keep the Library constantly at the forefront? Will your audience (there they are again) demand it? Will your organization stand for it?

All the above apply to all the elements inherent in the digitizing (or sampling) process rather than the characteristics of the material. Take colour fidelity. 5 Years ago 8 or 16 bits for describing a colour were all you could have – the technology of scanners couldn't provide any more. Then came 24-bit colour and that was it. The human eye couldn't discriminate any more. Now we have 30-bit and 32-bit representations – there will be more. Of course you can't move that 16-bit image to a 30-bit one, you have to re-scan with the latest equipment….

Handle things In-house or Out-source them?

It may be obvious by now that a lot of special considerations have to be taken into account and technical decisions made, well outside the normal scope of librarian training. Many of them lead to work which is outside the competence and expertise of library staff.

Here is a fundamental human resources issue. Do you train people to do a task which they will probably only perform once, and thus as a learner; or do you hire in experts and have no technology transfer to permanent staff? Both are right and both are wrong. Special circumstances prevail.

As an aside to calm growing fears (possibly), the situation is really not much different to installing an automated library system. Most librarians go through the trauma a small number of times, and learn relatively little of the process and the problems; they are not

relevant for 95% of their work. Some libraries hire a systems librarian and try to have a little expertise on staff, if only to keep check on what is being done. Some just pass the whole thing off to outside agencies. The options are the same and many of the arguments are the same, both during creation of the Digital Library and running it.

Is this a Project?

Tied up with the last is the degree of seriousness associated with the creation of the Digital Library. For some organizations it is a vital step in their continued existence (there must be some, but not a large number), for most it is a 'nice thing to have', for some it is a pet project of a staff member.

however it starts, it is a complex process and must be treated seriously if it is to have any chance of success. Even as a part time personal activity, treat it seriously and plan it as a project. Take the time to lay out what has to be done and who will do it. Set times and deadlines. They may scare you, but better at the beginning than after a year of work has been wasted.

Creating and running a Digital Library is not a simple task. It is a very close marriage of librarianship and computing along with a fair amount of marketing and administration. This is a learning ground for every one treat it sensibly and you will produce a worthwhile result and learn an immense amount.

Bonus Question

What about Long Term Support and Funding?

This is not a question it is a necessity. The one important, vital sine qua non is that of a long term commitment to the Digital Library. The library is different to all other libraries, it cannot be left alone or it will become useless.

Download address

Copies of the Digital Library Toolkit in PDF or RTF formats can be freely downloaded from:

http://www.sun..com

http://www.edulib.com

The exact location on each site is subject to change, but should be easily found. Within the EduLib site an updated and maintained database of the resources from Section 6 of the Toolkit is available.