Digital library futuristics:
Developing countries, universal access, and information for all

Ian H. Witten
Department of Computer Science
University of Waikato, New Zealand
ihw@cs.waikato.ac.nz

Key Words

Digital libraries, developing countries, internationalization, universal access, social issues in computing.

Abstract

Digital libraries are large, organized collections of information objects. Well-designed digital library software has the potential to enable non-specialist people to conceive, assemble, build, and disseminate new information collections. This has great social import because, by democratizing information dissemination, it provides a counterbalance to disturbing commercialization initiatives in the information and entertainment industries. This talk reviews trends in today's information environment, introduces digital library technology, and explores applications of digital libraries—including their use for disseminating humanitarian information in developing countries. We illustrate how currently available technology empowers users to build and publish information collections. Making digital libraries open to all, as conventional public libraries are, presents interesting challenges of universal access.

1 Introduction

Futuristic: We think it will run properly on the next generation of supercomputers.

Software Marketing Glossary, Anon (1996)

Digital libraries are large, organized collections of information objects. Whereas standard library automation systems provide a computerized version of the catalog—a gateway into the treasure-house of information stored in the library—digital libraries incorporate the treasure itself, namely the information objects that constitute the library's collection. Whereas standard libraries are, of necessity, ponderous and substantial institutions, with large buildings and significant funding requirements, even large digital libraries can be lightweight. Whereas standard libraries, whose mandate includes preservation as well as access, are "conservative" by definition, with institutional infrastructure to match, digital libraries are nimble: they emphasize access and evolve rapidly.

What will the future hold for digital libraries? The dizzying rate of change in the core technologies clouds even the brightest crystal ball. Perhaps the most striking feature of the digital library field, at least from an academic point of view, is the inherent tension between two extremes: the very fast pace of technological change and the very long-term view that libraries must take. We must reconcile our aspiration to surf the leading edge of technology with the literally static ideal of archiving material "for ever and a day." Any future we create must certainly not be "futuristic" in the jokey sense of the quotation above: it must run on everyone's computer today—for libraries are universally accessible, and should remain so—and it must also preserve the treasures of the past, including past digital libraries.

This paper is particularly concerned with the future for developing countries. It sometimes happens that technological advances in developing countries leapfrog those in developed ones. This occurs because established infrastructure, a strong and necessarily conservative force, is absent. Alternative sources such as solar energy are widely used in place of traditional power generation and distribution, while many developing countries have experienced far higher levels of mobile phone growth than developed ones. Digital libraries provide another example, compensating for the failure of traditional distribution mechanisms to address local requirements and deliver information where it is needed. Indeed, developing countries already have a competitive edge, for the labor-intensive process of optical character recognition (OCR) is often outsourced from the Western world to countries such as India, the

Philippines, and Romania. More intellectually demanding tasks such as metadata assignment and collection building will not be far behind. Under the leadership of Ranganathan, an influential librarian and educator who is considered the father of library science in India, India already has a distinguished history in metadata work.

In the next section we examine the social need for digital libraries, particularly in developing countries, by briefly sketching some trends in commercial publishing and contrasting them with a growing international perspective of information as a public good. Then we review a project that is applying digital library technology to the distribution of humanitarian information in the developing world, a context that is both innovative and socially motivated. Next we discuss issues of universal access and illustrate them with reference to the Greenstone digital library software (Witten and Bainbridge, 2003). We include a brief demonstration of a system that is intended to allow anyone to build and disseminate information collections, and illustrates some human interface challenges that arise when providing necessarily complex functionality to a non-computer-oriented user base. We close with the hope that future digital libraries will find a new role to play in helping to reduce the social inequity that haunts today's world, both within our own countries and between nations.

2   Books, libraries, and the socially disadvantaged

Today, the long-standing three-way tension between the commercial interests of publishers, the needs of society and information users, and the social mandate of public libraries, is being pulled and stretched as never before.

2.1   Books

What future has the book in the digital world? The question is a complex one that is being widely aired (see Lynch, 2001, for a particularly thoughtful and comprehensive discussion). Authors and publishers ask how many copies of a work will be sold if networked digital libraries enable worldwide access to an electronic copy of it. To counter the perceived threat, the entertainment industry is promoting "digital rights management" (DRM) schemes that permit a degree of control over what users can do that goes far beyond the traditional legal bounds of copyright. Indeed, they seem to be concerned solely with content owners rights and not at all with user's rights. Anti-circumvention rules are sanctioned by the Digital Millennium Copyright Act (DMCA) in the US (similar legislation is being enacted elsewhere).

Can DRM be applied to books? The motion picture industry can compel manufacturers to incorporate encryption into their products because it holds key patents on DVD players. Commercial book publishers are promoting e-book readers that, if widely adopted, would allow the same kind of control to be exerted over reading material. Basic rights that we take for granted (and are legally enshrined in the concept of copyright) are in jeopardy. DRM allows them to be controlled, monitored, and withdrawn instantly, and DMCA legislation makes it illegal for users to seek redress by taking matters into their own hands. Fortunately, perhaps, standardization and compatibility issues are delaying consumer adoption of e-books.

In scholarly publishing, digital rights management is more advanced. Academic libraries license access to content in electronic form, often in tandem with purchase of print versions too. They have been able to negotiate reasonable conditions with publishers. However, the extent of libraries' power in the consumer book market is moot. One can envisage a scenario where publishers establish a system of commercial, pay-per-view, libraries for e-books and refuse public libraries access to books in a form that can be circulated (Roehl and Varian, 2001, describe a parallel between historical circulating libraries and video rental stores).

These new directions present our society with puzzling challenges, and it would be rash to predict what society's response will be. But one thing is certain: they will surely increase the degree of disenfranchisement of those who do not have access to the technology.

2.2   Public information

In parallel with publishers' moves to reposition books as technological artifacts with refined and flexible control over how they can be used, an opposing trend has emerged: the ready availability of free information on the Internet. Of course, the world-wide web is an unreliable source of enlightenment, and undiscriminating use is dangerous—and widespread. But search engines and other

portals have enormously increased our ability to locate information that is at least ostensibly relevant to any given question.

Teachers complain bitterly that students view the Web as a replacement for the library, harvesting information indiscriminately to provide answers to assignments that are at best shallow and at worst incoherent and incorrect. Nevertheless, the Web abounds with accessible, high-quality information. Many social groups, non-profit societies and charities make it their business to create sites and collect and organize information there. Widespread use is strongly encouraged, and arrangements could surely be made for re-distribution of the material, particularly as a not-for-profit service, with appropriate acknowledgement.

A key problem with information distribution via the Web is that it disenfranchises developing countries. Although the Web does not extend into the homes of the socially disadvantaged in developed countries either, programs are working to provide access. But network access varies enormously across the world, and it is still true that, as Arunachalam wrote in 1998, the Internet "is failing the developing world". Prompted by this inequity, the importance of public information is today being highlighted by prominent international bodies. For example, UNESCO's "Information for all" programme was established in 2001 to foster debate on the political, ethical and societal challenges of the emerging global knowledge society and to carry out projects promoting equitable access to information. Information literacy is described as "a new frontier" by the Director of UNESCO's Information Society Division (Quéau, 2001). The International Telecommunications Union's World Summit on the Information Society (Geneva in 2003 and Tunis in 2005) will promote a global discussion of the fundamental changes that are being brought about by the transformation from an industrial to an information society, and confront the disparities of access to information between the industrialized countries and the developing world.

2.3   Libraries and their role

What is the librarian to make of all this? The mandate of public libraries is to facilitate the open distribution of knowledge. Librarians strive to enable the free flow of information. Their traditions are liberal, founded on the belief that libraries should serve democracy. A recent promotional video from the American Librarian's Association exults that "the library is democracy's place of worship" (ALA, 2002).

Clearly, the impending redefinition of the book as a digital artifact that is licensed rather than sold, tied to a particular replay device, with restrictions that are mechanically enforced, goes right to the heart of libraries. The changing nature of the book may make it hard, or even impossible, for libraries to fulfill their mandate by providing quality information to readers. And the emergence of a vast storehouse of information on the Internet poses a different kind of conundrum. Librarians, the traditional gatekeepers of knowledge, are in danger of being bypassed, their skills ignored, their advice unsought. Search engines send users straight to the information they require—or so users may think—without any need for an intermediary to classify, catalogue, cross-reference, advise on sources.

The ready availability of information on the Internet, and its widespread use, really presents librarians with an opportunity, not a threat. Savvy users realize they need help, which librarians can provide. A good example is Infomine, a cooperative project of the University of California and California State University (amongst others) (Mason *et al*., 2000). Infomine contains descriptions and links to a wealth of scholarly and educational Internet resources, each of which has been selected and described by a professional academic librarian who is a specialist in the subject and in resource description generally. Participating librarians see this as an important expenditure of effort for their users, a natural evolution of their traditional task of collecting and organizing information in print.

New trends in information access present librarians in developed countries with difficult and conflicting challenges. Meanwhile, however, the situation in the developing world is dire. Here, traditional publishing and distribution mechanisms have failed tragically. For example, according to the 1999 UN Human Development Report (UNDP, 1999), whereas a US medical library subscribes to about 5,000 journals, the Nairobi University Medical School Library, long regarded as a flagship center in East Africa, last year received just 20 journals (compared with 300 a decade before). In Brazzaville, Congo, the university has only 40 medical books and a dozen journals, all from before 1993, and the library in a large district hospital consisted of a single bookshelf filled mostly with novels.

Figure 1 A selection of recent humanitarian digital library collections on CD-ROM

2.4 Open-source software

Open source software is a powerful ally for librarians who wish to extend liberal traditions of information access. Open source projects make source code freely available for others to view, modify, and adapt; and the very nature of the licensing agreement prevents the software from being appropriated by proprietary vendors. But the open-source movement is more than just a vehicle for librarians to use: its link with library traditions goes much deeper. Public libraries and open source software both enshrine the same philosophy: to promote learning and understanding through the dissemination of knowledge. Both enjoy a sense of community, on the one hand the kind of inter-institutional cooperation exemplified by inter-library loan and on the other teams of designers and programmers that frequently cross national boundaries.

3 Disseminating humanitarian information with DLs

Digital libraries provide perhaps the first really compelling *raison d'être* for computing technology in the developing world. Priorities in these countries include health, agriculture, nutrition, hygiene, sanitation, and safe drinking water. Computers *per se* are not a priority, but simple, reliable access to practical information relevant to these basic needs certainly is. Witten *et al*. (2002) mention ten information collections in which Greenstone is being used to deliver humanitarian and related information in developing countries. For example, the *Humanity Development Library* is a compendium of practical information aimed at helping reduce poverty, increasing human potential, and giving a useful education. Rather than recapitulating parts of the above-cited paper, we describe four new ones (Figure 1).

The *Researching Education Development* library is a project of the Department for International Development (DFID), a British government department whose central focus is a commitment to a target of halving the proportion of people living in extreme poverty by 2015. Associated targets include ensuring universal primary education, gender equality in schooling, and skills development. It has created a CD-ROM library containing many education research papers and other documents. Each one represents a study or piece of commissioned research on some aspect of education and training in developing countries.

The *Energy for Sustainable Development* library, initiated jointly by the United Nations Development Programme (UNDP), the United Nations Department of Economic and Social Affairs (UNDESA), and the World Energy Council (WEC), contains a collection of 350 documents (26,000 pages). It includes titles that all these organizations have published on the subjects of energy for sustainable development—technical guidelines, journals and newsletters, case studies, manuals, reports, and other training material. The documents are in English, Spanish and French, and one document has Arabic, Russian and Chinese translations as well.

The *UNAIDS Library* contains publications that form a unique resource for those working in planning and practice. It is produced by the Joint United Nations Programme on HIV/ AIDS, whose mission is to lead, strengthen and support a response to the AIDS epidemic that will prevent the spread of HIV, provide care and support for those infected by the disease, reduce the vulnerability of individuals and communities to HIV/AIDS, and alleviate the socioeconomic and human impact of the epidemic.

The *Health Library for Disasters*, a collaboration between the emergency and disaster programs of the World Health Organization (WHO) and the Pan American Health Organization (PAHO), with the participation of many other organizations, contains over 300 technical and scientific documents on disaster reduction and public health issues related to emergencies and humanitarian assistance. A follow-up to the Spanish-language *Biblioteca Virtual de Desastres* discussed by Witten *et al.* (2002), it includes technical guidelines, field guidelines, case studies, emergency kits, manuals, disaster reports, and training materials.

4   Universal access

Universal access to digital libraries presents huge challenges to software engineers. The Greenstone digital library software (Witten and Bainbridge, 2003) allows us to glimpse some of the issues. We summarize some technical details in the next subsection, before turning to more interesting questions of access for readers, collection builders, and international users.

4.1   Platforms and distribution

Most digital libraries are accessed over the web. However, in many environments in developing countries, web access is insufficient and the system must run locally. And if people are to build and control their own libraries, a centralized solution is inadequate: the software must run on their own computers. Thus digital library systems intended for broad access should run on a wide variety of computer systems, particularly low-end ones.

The Greenstone server runs on any Windows, Unix, or MacOS/X system. All versions of Windows are supported, from 3.1 up. Supporting primitive platforms poses substantial challenges of a rather mundane nature: for example, Microsoft compilers no longer support Windows 3.1 and it is necessary to acquire obsolete versions (e.g. at software auctions).

In an international cooperative effort established in August 2000 with UNESCO and the Belgium-based Human Info NGO, Greenstone is being distributed widely in developing countries with the aim of empowering users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. UNESCO recognizes that digital libraries are radically reforming how information is acquired and disseminated in its partner communities and institutions in the fields of education, science and culture around the world, particularly in developing countries, and hopes that this software will encourage the effective deployment of digital libraries to share information and place it in the public domain.

4.2   Access for readers

Greenstone collections can be published as standalone libraries on removable media such as CD-ROM, or presented on the Web. Any Greenstone collection can be converted into a self-contained Windows CD-ROM that includes the Greenstone server software itself (in a version that runs right down to Windows 3.1) and an integrated installation package. The installation procedure has been thoroughly honed to ensure that only the most basic of computer skills are needed to install and run a collection under Windows.

4.3   Access for librarians: building new collections

Effective human development blossoms from empowerment rather than gifting. As the Chinese proverb says, "Give a man a fish and he will eat for a day; teach him to fish and he will eat for the rest of his days." Disseminating information originating in the developed world is no doubt useful for developing countries. But a more effective strategy for sustained long-term human development is to disseminate the capability of creating information collections, rather than the collections themselves. This will allow developing countries to participate actively in our information society, rather than observing it from outside. It will stimulate the creation of new industry. And it will help ensure that intellectual property remains where it belongs, in the hands of those who produce it.
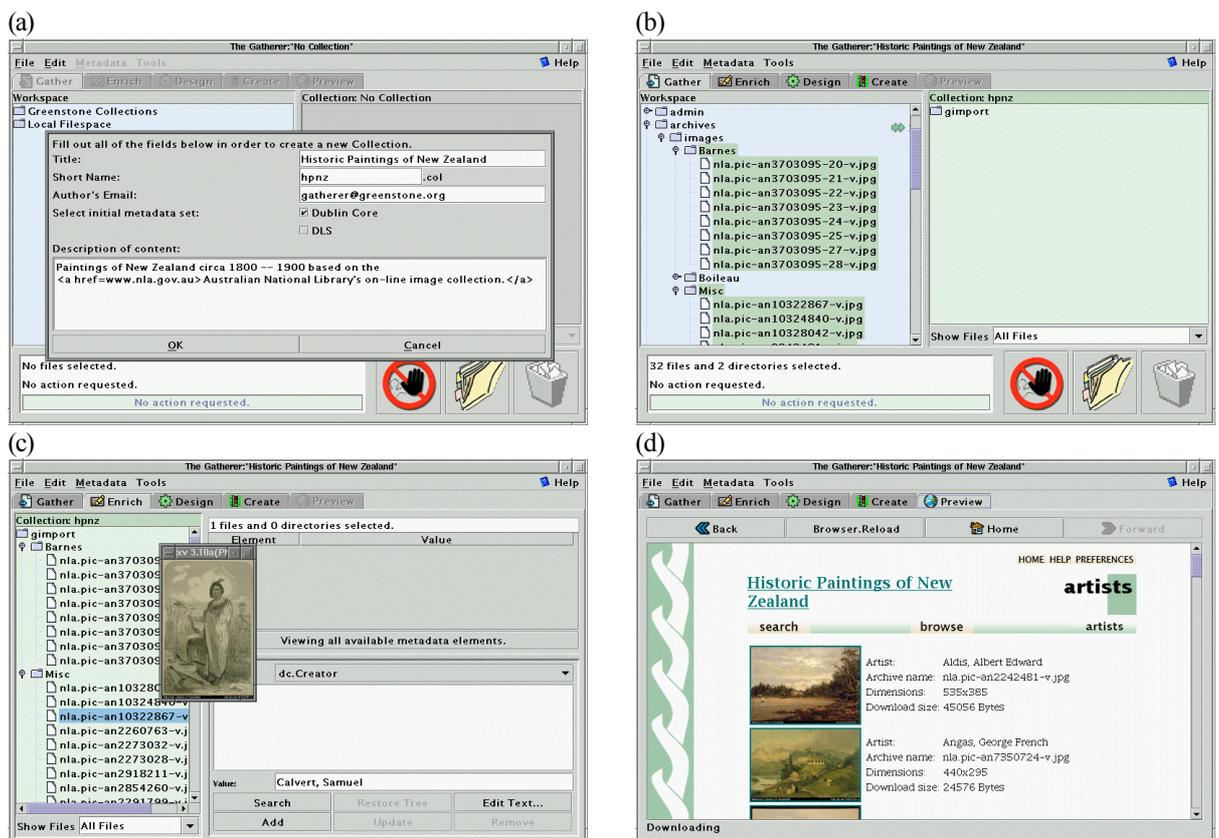
Figure 2 Building a collection with the Greenstone Librarian Interface.

Users whose skills are those of librarians rather than computer specialists can use Greenstone to build and distribute their own digital library collections. Figure 2 shows a user collating a selection of images for a digital library collection, augmenting these source documents with textual metadata and then building and viewing the collection. From here, a couple of clicks can produce a self-installing CD-ROM version of the collection.

This user is developing a digital library collection of historic paintings of New Zealand. They create a new collection using the file menu (Figure 2a), and fill out general information about the collection. Then a series of panels guides the user through the processes required to build the collection. The left-hand pane of the *Gather* panel (Figure 2b) shows the file system and the right-hand one represents the contents of the collection, initially empty, which the user populates by dragging and dropping files. Then (Figure 2c) the user moves to the *Enrich* panel and adds textual metadata (the name of the artists) to the selected documents. The *Design* and *Create* panels help structure the collection, control its appearance, and build it. Figure 2d shows a page from the newly built collection, in which source documents are listed by artist. Alongside each thumbnail are the artist's name, its catalog number, image dimensions, and its download size. The full-size image is shown by clicking on the thumbnail.

4.4 Internationalization

The Unicode character set is used throughout Greenstone, and documents in any language and character encoding can be imported. Collections in Arabic, Chinese, Cyrillic, English, French, Spanish, German, Hindi, and Maori are publicly available (see nzdl.org). The Greenstone web site (greenstone.org) links to sites that contain further examples.

It makes little sense to have a collection whose content is in Chinese or Hindi, but whose supporting text—instructions, navigation buttons, labels, images, help text, and so on—are in English. Consequently, the entire Greenstone interface has been translated into a range of languages,

and the interface language can be changed by the user as they browse from the *Preferences* page. Currently, interfaces are available in 30 languages (including Hindi and Kannada).

5   Conclusions

By allowing people to easily create and disseminate large information collections, digital libraries extend the applications of modern technology in socially responsible directions, and counter a possible threat towards the commercialization of information in line with practices developed by the entertainment industry. As far as the developing world is concerned, digital libraries may prove to be a "killer app" for computer technology—that is, an application that makes a sustained market for a promising but under-utilized technology. The World-Wide Web is often described as the Internet's killer app. But the Internet does not really extend to developing countries, and the developing world is missing out on the prodigious amount of basic, everyday human information that the Web provides, and its enormous influence on promoting and internationalizing business opportunities. There is little incentive to make copies of the entire Web available locally because of its vast size, rapid change, and questionable information value per gigabyte. However, it is easy to provide focused information collections on both the Web and, in exactly the same form, on removable media such as CD-ROM, DVD, or bulk disk storage devices—indeed, the Greenstone software described above allows one to create a complete, runnable, self-installing CD-ROM image from a Web collection in just a few mouse clicks.

Public libraries are founded on the principle of universal access, and digital libraries should be too. Universal access means running on low-end devices, but one does not want to provide a lowest-common-denominator solution that sacrifices high-end capability where it is available. Universal access means that interfaces should be available in the world's languages, but one does not want the burden of translation to stifle the development of new functionality and features. Universal access means educating users: UNESCO is mounting training courses on building collections with Greenstone in Almaty, Bangalore, Senegal, and Suva, and discussions are underway for Latin America; the Tulane Institute has run courses that use Greenstone collections as a resource in many locations in Africa (e.g. Burkina Faso, Cameroon, Cote d'Ivoire, Democratic Republic of Congo, Ghana, Rwanda, Senegal, Sierra Leone, Togo) and Latin America (e.g. Argentina, Bolivia, Colombia, Ecuador, Guatemala).

Universal access also means that non-textual material should enjoy first-class status in a digital library—perhaps first-class status in "the literature." This has important cultural ramifications. It should be possible to create digital library collections intended for people in oral cultures, who may be illiterate or semi-literate. Or people who, though literate in their own language, cannot speak or read the language of the documents. Opening digital libraries to the illiterate is a radical and potentially revolutionary benefit of new interface technology.

Digital libraries give software engineers a golden opportunity to help reverse the negative impact of information technology on developing countries and reduce the various "digital divides" that cleave our world (Norris, 2001)—the "social divide" between the information rich and the information poor in our own nations, the "democratic divide" between those who do and do not use the panoply of digital resources to engage, mobilize and participate in public life, as well as the "global divide" that reflects the huge disparity in access to information between people in industrialized and developing societies.

Acknowledgements

References

Anon (1996) *Software marketing glossary*. http://www.flinders.edu.au/isd/Newsletter/may96.html

ALA (2002) "Rediscover America @ your library." Video produced by the American Library Association, Chicago, IL. Available from www.ala.org/@yourlibrary/rediscoveramerica.

Arunachalam, S. (1998) "How the Internet is failing the developing world." Presented at *Science Communication in the Next Millennium*, Egypt; June.

Lynch, C. (2001) "The battle to define the future of the book in the digital world." *First Monday* 6(5)..

Mason, J., Mitchell, S., Mooney, M., Reasoner, L. and Rodriguez, C. (2000) "INFOMINE: Promising directions in virtual library development." *First Monday*, Vol. 5, No. 6; June.

Norris, P. (2001) *Digital divide? Civic engagement, information poverty and the Internet worldwide*. Cambridge University Press, New York

Quéau, P. (2001) "Information literacy: a new frontier." *UNISIST Newsletter*, Vol. 29, No. 2, pp. 3-4.

UNDP (1999) *Human development report 1999*. UNDP/Oxford University Press, New York.

Witten, I.H., Loots, M., Trujillo, M.F. and Bainbridge, D. (2002) "The promise of digital libraries in developing countries." *The Electronic Library*, Vol. 20, No. 1, pp. 7–13.

Witten, I.H. and Bainbridge, D. (2002) *How to build a digital library*. Morgan Kaufmann.