

Digital Libraries for the Developing World

Ian H. Witten

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

Digital libraries (DLs) are the killer app for information technology in developing countries. Priorities here include health, agriculture, nutrition, hygiene, sanitation, and safe drinking water. Computers are not a priority, but simple, reliable access to targeted information meeting these basic needs certainly is. DLs can assist human development by providing a non-commercial mechanism for distributing humanitarian information on topics such as health, agriculture, nutrition, hygiene, sanitation, and water supply. Many other areas, ranging from disaster relief to medical education, from the preservation and propagation of indigenous culture to educational material that addresses specific community problems, also benefit from new methods of information distribution.

Traditional publishing and distribution mechanisms have tragically failed the developing world. By decoupling production and distribution costs from intellectual property charges, DLs offer a sorely needed lifeline. A wealth of humanitarian material is produced by inter-government agencies, national bodies, and non-governmental organizations. Much of this—at least in principle—resides in the public domain and be made freely available if suitable distribution mechanisms exist. More broadly, increasing the level of education through knowledge and technology transfer, particularly for doctors, teachers, and other professionals, is an important social and economic benefit of any kind of library.

In developed countries we think of the Internet as the infrastructure for information distribution. But network access varies widely across the globe. Schools and hospitals in developing countries are poorly connected; even universities have appalling access by western standards. While global satellite communication networks may eventually bring relief, this takes time and money. However, the structure and organization of DLs can be separated from their distribution media. Physical distribution of information on recordable devices can provide an attractive alternative to networks. CD-ROM is a very practical format for areas with little Internet access. Disks can hold a small library of a few thousand illustrated and fully-indexed books; DVDs increase this manifold. Information remains accessible even when the network goes down. It can be carried around and distributed physically even in countries with unreliable infrastructure or repressive regimes.

Computers are not so hard to come by in developing countries as one might think. Their extraordinary rate of obsolescence, coupled with the developed world's voracious appetite for the latest and greatest, makes low-end machines essentially free. Instead of clogging landfill sites many (though certainly not enough) find their way to developing countries. Of course, they tend to be low-end. A worse problem in our experience is that they come with poorly-installed software that is inappropriate to their network environment and intended use. They may lack network access, browser software, adequate disk storage and main memory, and CD-ROM drives. Developing countries have an endemic lack of skilled people to reconfigure the software and resolve technical problems.

DL software must be open source to be relevant in this context. And no matter how good it is, software that cannot be installed satisfactorily is useless. Developers must ensure that their systems operate on any computer system and can be easily installed. Manual configuration is beyond the skills of most end-users. Installation takes place in an environment that lacks any software support infrastructure. The variability of the network configuration raises particular problems for DL software that can use both the Internet and removable storage. The technical challenges are daunting. They demand arduous work and user testing, and the results are mundane and decidedly unsexy. As just one example, we resorted to software auctions to find compilers that generate code for obsolete systems (e.g. Windows 3.1). The rewards are nebulous at best: the work is neither commercial nor computer science research in any conventional sense.

We have extolled the virtues of DLs for disseminating information originating in the developed world. There are many excellent examples of this useful work. However, effective human development blossoms from empowerment rather than gifting. A greater and far more ambitious aspiration is to foster the ability for people in developing countries to build information collections of their own. Then, those with access to information could create and distribute their own libraries. For example, someone in a university or technology center might select information from the Web and combined it with local documents into a unified collection that is written to CD-ROM and distributed to people without network access.

This presents an array of user interface challenges. Complex library-building software, including metadata assignment and full-text indexing, must be targeted to end users with limited technical skills and no technical

support. Documents come in a variety of formats. Many users want to incorporate multimedia such as audio and video. To distribute the collection it should be possible to transfer it to CD-ROM, along with the programs needed to access it. Probably it should be distributed on the Internet too, raising issues of compatibility between networked and standalone versions.

Imagine a librarian user creating a DL for readers. In order to gain institutional support, they must be able to brand the collection appropriately. The reader's interface should be available in local languages, and librarians or their colleagues should be able to add new ones. The librarian's interface is more extensive than the reader's, and need not be available in local languages. For example, in Africa the librarian's interface may be in English, French and Arabic, whereas the reader's interface may be needed in local languages such as kiSwahili, Kinyarwanda, Shona.

Open source DL software has the potential to enable non-specialist people to conceive, assemble, build, and disseminate new information collections. In practice, training people how to use this software to build their own collections is a serious bottleneck in the application of this technology. Internationally-run commercial courses are infeasible in poorer parts of the world for economic reasons. Inter-governmental organizations such as UNESCO sponsor regional training workshops that train trainers who can return to their local organizations and run their own workshops there. Attendees must be supplied with a full set of teaching material that they can use back home, and training material should be published in the Internet with no restrictions on redistribution.

To make a serious impact on the developing world, large-scale training is necessary. For example, UNESCO's Communication and Information program has mounted workshops on the use of the Greenstone DL software for building collections in India (Bangalore), Eastern Europe (Almaty), Africa (Dakar, Cape Town) and the Pacific region (Suva). In India several institutions have run follow-on workshops and a user group is forming. A feasibility study has been produced for an African support group, with input from several major language communities. There are Internet sites for French- and Spanish-speaking users. A CD-ROM self-study course, including a software resource kit, is produced by the FAO. There are encouraging signs that the news is beginning to spread.

Ian Witten directs the New Zealand Digital Library Project, which produces the open source Greenstone software that is widely used in developing countries (see www.greenstone.org)