

NEW PROJECTS FOR THE FUTURE OF AFRICA / NOUVEAUX PROJETS POUR L'AVENIR DE L'AFRIQUE

AN OPEN SOURCE, OPEN ACCESS JOURNAL DATABASE APPLIANCE: A PROPOSAL

Edward M. Corrado
The College of New Jersey, Ewing, NJ, USA

Abstract

This paper proposes the creation of an Open Access Journal Database Appliance using Open Source Software for use at libraries in developing countries. The paper introduces the reasons for building such a system, including the lack of reliable Internet connections. Some of the issues that will need to be considered during the creation, implementation, and maintenance of such a system are discussed.

Résumé

Un dispositif de bases de données de journaux Open Source, Open Access : Une Proposition

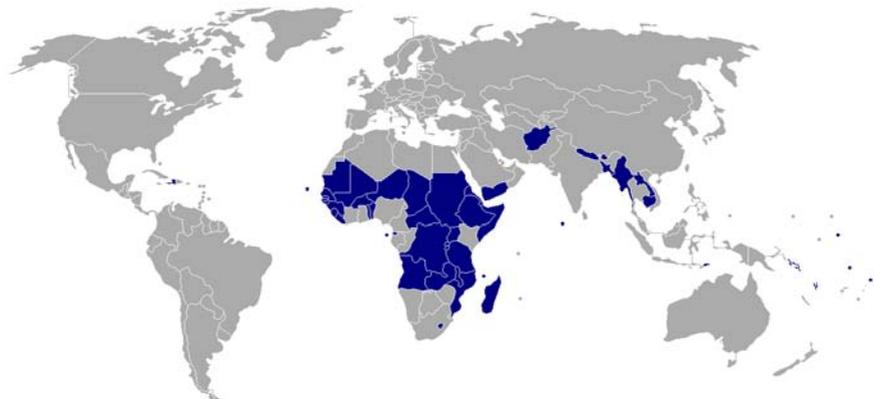
Cet article propose la création d'un dispositif de base de données de journaux en Open Acces utilisant des logiciels Open Source pour les bibliothèques des pays en développement. On y expose les raisons de la construction d'un tel système, y compris l'absence d'une connection Internet fiable. Enfin, il discute certains des problèmes qu'il est nécessaire de considérer pendant la création, l'implantation et la maintenance d'un tel système.

The Open Access Journal Database Appliance (OJDA) is a digital library that includes content from open access journals and other scholarly publications. OJDA will be created using open source software such as Greenstone and distributed (and designed) in such a way that connectivity to the Internet is not a pre-requisite for its use. The purpose of the OJDA is to help libraries in developing countries with insufficient and unreliable Internet infrastructure to have enhanced access to scholarly research. While libraries (and individuals in general) can access open access publications for no cost, and some database providers offer free access to African researchers; in order to access these resources, researchers must have an Internet connection that is fast enough and reliable enough to search for and retrieve articles and other publications.

Slow or limited Internet access in Africa's Least Developed Countries

Faculty and students at colleges and universities located in Africa's Least Developed Countries (LDCs) may have very slow or limited Internet access, if any access at all. LDC is a United Nations designation that is used to describe countries that represent "the poorest and weakest segment of the international community" (United Nations Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and the Small Island Developing States, 2007a, n.p.). These countries are characterized by economic vulnerability and extreme poverty. There are three criteria used for identification as an LDC. The first criterion is low gross national income (under \$745 for inclusion, over \$900 for graduation). The second criterion looks at human capital status and is based on a Human Assets Index (HAI). The HAI takes into account nutrition, mortality rates of children under five years old, education, and adult literacy. The third criterion is based on economic vulnerability and a composite Economic Vulnerability Index (EVI). The EVI is based on indicators involving "(a) population size; (b) remoteness; (c) merchandise export concentration; (d) share of agriculture, forestry and fisheries in gross domestic product; (e) homelessness owing to natural disasters; (f) instability of agricultural production; and (g) instability of exports of goods and services" (United Nations Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and the Small Island Developing States, 2007b, n.p.). As of June 14, 2007, when Cape Verde graduated off the list, there were 49 countries listed as LDCs (United Nations News Service, 2007, n.p.). Thirty-four of the LDCs are located in Africa, twelve in Asia, five in the Pacific, and one is located in Latin America.

Figure 1: map of LDCs (Source: http://en.wikipedia.org/wiki/Least_developed_country/)



A 2004 study that looked at selected LDCs in Africa showed, partially because of slow connections and limited availability of computers connected to the Internet, that “most research is still done using traditional print periodicals” and “in Nigeria, only 19.2% [of researchers] use the Internet for collecting teaching materials and 18.5% for publishing purpose. 36.2% use it for networking exchanging, 57.7% for academic research, and 6.2% for e-commerce” (Oyelaran-Oyeyinka and Adeya, 2004, p. 75). Even in African universities that have Internet connections, many students and professors are unable to get online because of costs and limited bandwidth (Miner and Missen, 2005, p. 27). Connected universities in Africa may lose access to the Internet for days or weeks at the time and research suggests that the average connected African university is only connected for an average of about five hours per day (Miner and Missen, 2005, p. 28).

Despite these issues, the future may not be so bleak. There are a number of positive programs that are underway to improve network infrastructure in Africa’s LDCs. The African University Network, a joint United Nations Agency project involving the United Nations University and the International Telecommunication Union, has the goal of having every African college and university connected by 2015 (Global Virtual University, 2007, n.p.). Another program, Connectivity Africa, aims to improve access to information and communication technologies (ICTs) in Africa by providing support in four areas: 1) innovation and use of ICT, 2) African regional ICT, 3) research and development in African ICTs, and 4) partnerships and networks (Connectivity Africa, 2007, n.p.). These and other programs offer the promise of improved Internet connectivity in the future. However, addressing these issues will take time, so new approaches to providing access to access to online scholarly need to be explored that can be implemented in the mean time.

The OJDA project proposal

In order to provide improved access to scholarly materials in academic libraries located in LDCs, any plan needs to consider the sub-standard network infrastructure, limited numbers of computers and computer professionals, and the low income of these areas. OJDA does this in multiple ways. First it addresses the issues associated with sub-standard network infrastructure by storing data locally. While local data storage helps alleviate problems associated with poor Internet connectivity, it does create other challenges. Because the content is stored locally, someone at the local site will need to have the knowledge and ability to operate the equipment and load new content into the system. This also increases the amount (and cost) of computer hardware needed on site. A computer system on site with enough hard drive space, processor speed, and memory will need to be acquired.

Free/Open Source Software

This project relies on open source software so there will be no charge for acquiring the software. However, as with any software project, there will be a need to continually develop and maintain the software. Open source software is called Free Software by some. While there are some differences in philosophy between Open Source and Free Software, the characteristics of the software are very similar and often identical. The characteristics of Free/Open Source Software (FOSS) include that the source, or programming, code and that the software is typically available at no cost. Although access to the source code is one of the conditions necessary in order for software to be considered FOSS, it is not the only one. Users of the software must also be able to make changes to the software and redistribute the software. Additionally, the ability to use of the software cannot discriminate against any person or field of endeavor. The open source software model allows for distributed development of the software by volunteers from around the world. Besides the low costs and the ability to freely re-distribute open source software, an added benefit of using open source software is that by doing so, developing countries will not be dependent on Western corporations whose primary aim is accumulating capital and not on solving the digital divide (Fuchs and Horvak, 2006, p. 15).

Free Software provides the user with four core freedoms: the freedom to 1) run the program, for any purpose, 2) study how the program works, and adapt it to her needs, 3) redistribute copies so he can help his neighbor, and 4) improve the program, and release the improvements to the public, so that the whole community benefits (Free Software Foundation, 2007, n.p.). With these freedoms also comes the responsibility to FOSS users to share their improvements with others. This combination of freedoms and responsibility often leads to vibrant communities of users of FOSS projects that help one another learn how to use and improve the software.

FOSS projects have been getting a lot of attention in the library world of late. The FOSS Evergreen integrated library system developed for, and by, the State of Georgia Public Library system in the United States has caused many librarians to take notice of the feasibility of FOSS in mission critical services. Libraries and information centers in developing African nations have also taken notice. Koha, a FOSS integrated library system, has been implemented at the University of Kinshasa. One of the college libraries in the University of Malawi system is also migrating to Koha. Slums Information Development and Resource Centers (SIDAREC) community library located in the slums of Nairobi, serves twenty secondary and primary schools and is using Koha and other FOSS projects to provide information and resources to the youth of Nairobi. Some libraries in Africa are using Greenstone (a FOSS digital library program) for digital repositories and others have

attended Greenstone workshops organized by Electronic Information for Libraries (eIFL). In short, libraries in Africa and elsewhere in the developing world have shown they can successfully use FOSS. This is especially so when there is a user community or outside organizations that can help provide support and training.

Open Access

Open access is a term that has varying definitions and may mean different things to different people, but basically open access “calls for scholarly publications to be made freely available to libraries and end users” (Corrado, 2005, n.p.). By focusing on open access content in this project, there will be less legal hurdles and copyright negotiations needed for acquiring and repackaging the scholarly materials. This will allow project participants and volunteers to focus on identifying and collecting content rather than negotiating with content providers.

Open access not only offers researchers in developing countries (and elsewhere) access to information they need to do their research, but it also can provide “a means by which they can more efficiently contribute their work to the global research community” (Electronic Information for Libraries, 2007, n.p.). As Matsika (2007, p. 16) observes, “open access tends to be very appealing in developing countries [...] where demand outstrips supply and poverty has wrecked havoc and made information a commodity priced beyond the reach of most of the community.”

How it can be done

Simple solutions work the best. The system needs a simple, straightforward, user friendly interface that is easy to maintain and operate. OJDA and similar solutions need to be considerate of the sub-standard network infrastructure, limited number of computers, limited number of computer professionals, and inadequate funding. Luckily for this proposal, there are already a few well-developed open source article repository systems that are easy to install and configure. These systems require little day-to-day maintenance and can be adapted for use in this project. Two such projects are E-prints and Greenstone. The first step in the process will be to identify appropriate open access content. Gathering articles can be done from many open access repositories using the Open Archive Initiative Harvesting Protocol for Metadata Harvesting (OAI-PMH), the Open Archive Initiative Object Reuse and Exchange (OAI-ORE) (when it becomes available), and other data exchange protocols. While relying on these protocols may limit the potential repositories that can be accessed, it also should allow for better automation and lower costs. Additional data can be acquired using web archiving software such as the open source Heritrix crawler that is used by the Internet Archive.

One of the biggest challenges for this project will be how to get the data to libraries located in rural areas of LDCs. While delivering the data over the Internet in

blocks might be feasible in some cases, in others this might not work due to network infrastructure issues. One solution is to send the data can be delivered during low use times to minimize the effects on the network. Where the network infrastructure is completely lacking, the content will need to be physically shipped. One possible method to achieve this while minimizing costs is to ship pre-configured hard drives. While the costs of hard drives should not be discounted, the old ones can be shipped back and reloaded with new data. With this method, the local site will only need to take the old hard drive out and replace it with the new one. An alternative option that would be more cost effective but may require more local expertise is to ship the updates on one or more DVDs, CD-ROMs, or USB memory storage devices. If the program is designed properly, there should not be a significant amount of computer knowledge necessary to load the updates into the system from DVDs.

Besides the initial programming design, it will be necessary to maintain the system and keep it up to date. Who is going to do this? Who will cover the costs? One possible partner, obviously, is IFLA. Besides IFLA developing the OJDA as an IFLA core program, there are a number of non-governmental organizations (NGOs) that can be approached to help offset costs, and volunteers can be recruited to help with the computer programming of this worthwhile cause. One NGO that might be able to work with this project is eIFL. eIFL already has a core program eIFL-FOSS that is dedicated to helping eIFL libraries take advantage of the benefits of FOSS. This program, combined with eIFL's open access program seems to be an appropriate place for such a project (assuming eIFL is interested). A possibility for dealing with on-going updates is the creation of a sister library program. This sister library program can create partners between libraries in the developed world and those in LDCs. Libraries in the developed world can create copies of the updates on DVDs or hard drives and can be responsible for shipping them to their sister library located in the LDC. This OJDA proposal fits in with Fuchs and Horvak's (2006, p. 15) call for "solutions to the material and social causes of the global digital divide [...] based on open standards and copy-left licenses."

Related Projects

The OJDA proposal is not the first initiative to attempt to provide improved access to scholarly materials for African researchers. Programs like The Access to Global Online Research in Agriculture (AGORA) provides access to 747 journals from major scientific publishers and eIFL is an independent foundation that's main focus is to negotiate affordable subscriptions to electronic resources on a multi-country con-

sortia basis (Smart, 2005, p. 262). African Journals OnLine (AJOL) provides online access to scholarly research published inside of Africa using the Open Journal System, “an open source software developed by the Public Knowledge Project in Canada” (Smart, 2005, p. 263).

These above projects and others like them are not specifically designed to deal with issues related to Internet connectivity. Even with free access to online resources, the “dearth and expense of bandwidth [...] severely restrict African researchers’ ability to take full advantage of free or subsidized resources” (Miner and Missen, 2005, p. 23). The eGranary Digital Library is designed to be a practical solution to the problem of not being able to access scholarly resources because of poor, or non-existent, Internet connectivity. The eGranary Digital Library is a “collaborative effort to garner and deliver a wealth of digital information to the world’s information poor” (Miner and Missen, 2005, p. 21). eGranary Digital Library is open source and is designed to be inexpensive and easy to use. It provides instant access to Web pages, audio, video, and other multimedia content from within an institution’s local area network. It does not require Internet connectivity. By relying on an institution’s LAN the eGranary Digital Library replaces bandwidth with local data storage. It has been estimated that Web documents open 5,000 times faster from an eGranary than it does from a typical African Internet connection (Fritz, 2005, p. 11). The eGranary includes over 3 million digital documents on a 250 Gb disk drive. As of 2005, there have been over seventy-five installations in over fifty campuses in Africa and elsewhere.

The eGranary Digital Library contains collections in multiple subject categories. The content for these categories are selected by subject experts, called content editors. Besides identifying content, these content editors may also attempt to secure permission to use copyrighted material. Unlike the proposed Open Access Journal Database Appliance proposed in this paper, eGranary actively seeks out all kinds of information. About 90% of the content is from the public Internet while the other 10% of materials are donated by authors or publishers (Missen, 2005, p. 196). Once collections are identified the collections are harvested using HTTrack, an open source program that is a free and easy-to-use offline browser utility that can be used to “mirror” Web sites.

The eGranary Digital Library may be updated using a variety of transport mechanisms. When a full Internet connection is available, that can be used. Otherwise the content can be shipped via DVD, CD-ROM, USB memory stick or on other digital media. The eGranary Digital Library includes software that makes it simple to apply the updates, once they are acquired.

While the WiderNet Project that is responsible for the eGranary Digital Project has not published a formal evaluation, they report that the early feedback has been generally positive (Missen, 2005, p. 197). Besides providing access to digital con-

tent, the eGranary Digital Library also provides subscribers with a publishing platform that can be used to help African scholars share their work with their African colleagues.

There are a few differences between OJDA and eGranary. OJDA only includes open access content. Because WiderNet's agreements with copyright holders for commercial content may not allow for the eGranary to be installed (at least with complete access to the eGranary content) in the developed world, it limits who can use the product. Since OJDA consist only of open access materials, it will be available to more institutions, including those in the developed world without having to worry about copyright concerns. Making OJDA available to institutions in the developed world may have the benefit of increasing awareness of librarians in the developed world, and leading to more interest and additional volunteers to work on the project. Also OJDA has one simple interface for all included content while eGranary uses the multiple native interfaces of the included content.

What can be learned?

OJDA can learn a number of things from eGranary. The eGranary project has shown that replacing bandwidth with storewidth can be a viable way to begin to minimize the information divide and we can learn a number of things from their success. By recruiting subject specialist to identify open access content, we can be sure that we are including quality content from the Internet. eGranary includes content other than scholarly articles. For example, it includes educational Web sites and books. This type of content may also fit in well with OJDA. Also, OJDA can borrow ideas on how to distribute the software and content. Like eGranary, OJDA can distribute data via hard drives as time and travel permit and via one-way satellite communications.

Conclusions

There is a definite need for "offline" access to "online" scholarly publications in the developing world. Many universities and colleges in LDCs and other developing countries do not have adequate bandwidth for scholars to connect to Internet-based research, even when access is offered without cost. The "come and get it" approach to accessing scholarly information on the Internet does not work for scholars at colleges and universities in some parts of the world because it is a huge hurdle for them to get on the Internet. As other papers presented at this conference show, FOSS can be indeed be used successfully in libraries located in the developing

world. By promoting communities of FOSS in the developing world, libraries can work together by teaching and helping each other with the use and support of FOSS. This proven model can be used with the OJDA. Likewise, by focusing on open access literature, librarians will be promoting the sharing of knowledge among libraries. An additional benefit of open access is that because the license does not restrict who can view the content, the OJDA can be used by any library around the world. Librarians that use OJDA in the developed world may find the project useful based on hands on experience and will be more willing to volunteer to help improve the project by participating in a sister library program, providing software code, or working to identify additional content.

Bibliography

Connectivity Africa (2007). About us. 16 August 2007 <<http://www.connectivityafrica.ca/>>.

Corrado, Edward (2005). The importance of open access, open source, and open standards for libraries, *Issues in Science and Technology Librarianship*, 42. Available: <http://www.istl.org/05-spring/article2.html> [Accessed May 27, 2007]

Electronic Information for Libraries (2007). Open access. 4 September 2007 <<http://www.eifl.net:9080/cps/sections/services/eifl-oa>>.

Free Software Foundation (2007). The Free Software definition. 15 August 2007 <<http://www.gnu.org/philosophy/free-sw.html>>.

Fritz, Mark (2005). Food for thought: localized content aids disconnected Africa. *Econtent*, 28(7/8): 10-11.

Fuchs, Christian and Horak, Eva (2006) Africa and the digital divide, *Telematics and Informatics*, In Press, Corrected Proof, Available online 31 July 2006: <http://www.sciencedirect.com/science/article/B6V1H-4KJ0FDW-1/2/08eed7043af4fe673e375596a732531a>

Global Virtual University (2007). African University. 16 August 2007. <<http://gvu.unu.edu/docs/African%20University%20Network.doc>>.

Matsika, Kathy (2007). Intellectual property, libraries and access to information in Zimbabwe. *IFLA Journal*, 33(2): 160-167.

Miner, Edward A. and Missen, Cliff (2005). "Internet in a box": augmenting bandwidth with the eGranary Digital Library. *Africa Today*, 52(2): 20-37.

Missen, Cliff (2005). Internet in a Box: The eGranary digital library serves scholars lacking Internet bandwidth. *New Review of Information Networking*, 11(2): 193-199.

Oyelaran-Oyeyinka, Banji and Adeya, Catherine Nyaki (2004). Internet access in Africa: empirical evidence from Kenya and Nigeria, *Telematics and Informatics*, 21: 67-81.

Smart, Pippa (2005). African Journals OnLine (AJOL). *Serials Review*, 31: 261-265.

United Nations News Service (2007). UN advocate salutes Cape Verde's graduation from category of poorest States. 14 June 2007 <<http://www.un.org/apps/news/story.asp?NewsID=22918&Cr=cape&Cr1=verde>>.

United Nations Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and the Small Island Developing States (2007a). Criteria for identification of LDCs. 2 September 2007 <<http://www.unohrrls.org/en/ldc/related/59/>>.

United Nations Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and the Small Island Developing States (2007b). About least developed countries. 2 September 2007 <<http://www.unohrrls.org/en/ldc/related/58/>>.