

GREENSTONE USER SURVEY

Laura Sheble
School of Library and Information Science
University of North Carolina
June 2006

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. It provides a new way of organizing information and publishing it on the Internet in the form of a fully-searchable, metadata-driven digital library. It has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is open-source, multilingual software, issued under the terms of the GNU General Public License.

The Greenstone User Survey [1] was developed to gather feedback from Greenstone users and developers about the adequacy of current support structures and mechanisms and how support for users might be improved or augmented in the future. The New Zealand Digital Library team wanted to learn more about the organizational and technical environments in which Greenstone is implemented and about the user groups for whom the collections are built. The survey served as a mechanism to gather general feedback from Greenstone users and developers, and gave them an opportunity to indicate how they, or organizations in the countries in which they work, might be interested in participating in the Greenstone support community. Survey respondents were also given an opportunity to contribute information about collections they had built with Greenstone for inclusion in the online directory of collections, available in the *examples* section of the Greenstone website [2]. This paper outlines descriptive survey results for each major objective of the survey.

The survey targeted users and developers of the Greenstone Digital Library Software. These are people who (1) install and build digital collections with Greenstone; (2) develop functionality or language interfaces for the software; (3) use Greenstone as a tool to teach about digital libraries; (4) work to promote or disseminate Greenstone; or (5) use the software in other ways. The survey was not aimed at end users of digital library collections built with Greenstone.

Survey Development and Methods

The User Survey was developed over the course of several months in late 2004. A small team began by composing a core set of questions focusing on support and the organizational and technical environments of users. Many different question types were included: checkboxes, multiple choice, rating scales, and a large number of open ended questions, which were necessary to capture survey participants' ideas about support structures and mechanisms (see Table 1 for a summary of survey section attributes).

A draft of the survey and instructions was distributed to selected members of the broader Greenstone development community for critique, including members in Africa, Europe, India, and New Zealand, and research support staff at Wayne State University. The material was modified according to the feedback received, and both a web-based version and a document-based version of the survey and instructions were developed. In

creating the web version it was necessary to restructure the survey slightly. Though originally we did not want participants to have to answer any specific questions in order to participate, the final structure required participants to respond to 6 of the 40 questions. This change was necessitated by a combination of the technical design of the survey software and the nature of the questions we wanted to ask.

Table 1: Final survey content.

Section	Content	No. Questions	Question Type(s)	Est. time (min.)
I	General questions: the organizations Support mechanisms	24	12 open-ended 7 yes/no 2 checkbox 1 multiple choice 1 rating scale 1 short answer	9-14
II	Collections, technical environments, development, and characteristics of the target populations of end users	8	3 checkbox 2 short answer 2 open-ended	2-3
III*	Contact for follow-up activities Submit collection information for online directory	8	6 short answer 1 checkbox 1 yes/no	

* Optional, supplementary section

The survey was announced on the Greenstone Users' and Developers' lists on December 3rd by a recognized developer of Greenstone Digital Library Software. Two follow-up announcements were posted, one on December 20th and a last call on January 3rd, at which time an extension was announced to the end of that week, January 7th. Although it was not actively promoted or updated, the survey remained available after the initial survey period and further responses were received; responses through January 2006 are considered in this analysis.

The Users' and Developers' lists were convenient means of distribution. Membership indicates some degree of interest in the software, and some support for its use and development—the primary foci of the survey. However, we probably did not reach all potentially interested recipients. In any case the sample population, if taken to represent the larger Greenstone community, would surely be biased by any sampling method. Note that “as with all open source projects, the user base for Greenstone is unknown” [3]. In fact, beyond anecdotal evidence and download statistics from SourceForge, the major distribution point for Greenstone, little is known about the user base. Any insights into the breadth and diversity of the Greenstone user and developer base are likely to be useful, and this initial survey does provide insight into several aspects of this question.

Results

The survey received 62 valid responses from users and developers who work with Greenstone in at least 32 different countries (see Table 2), representing every major

geographic region (see Figure 1). The survey announcement was distributed to approximately 600 email list members in 70 different countries. Though the response rate was fairly low (10%), this could be for several reasons. First, the survey was geared towards current users and developers rather than prospective users or email list members who monitor the list for research or other purposes. Second, anecdotal evidence suggests that in some cases respondents participated in the survey as a representative of their organizations, of whom more than one member might be on the email lists. Third, the survey was administered during December and the beginning of January, a time when staff in many organizations are involved in end-of-period activities and, in some regions, a common vacation period.

Because of the assumed low response rate—and also because one cannot say whether even the list members, let alone the survey participants, are representative of Greenstone users and developers overall—the data presented here is used to indicate the breadth and diversity of the Greenstone community and to serve as a base on which future measures of the community can be built. Quantitative data is used to indicate the nature of the pool of survey respondents. It also introduces a descriptive base on which the characteristics of Greenstone users and developers can be built, particularly regarding the community of users and developers who are members of the Greenstone email lists.

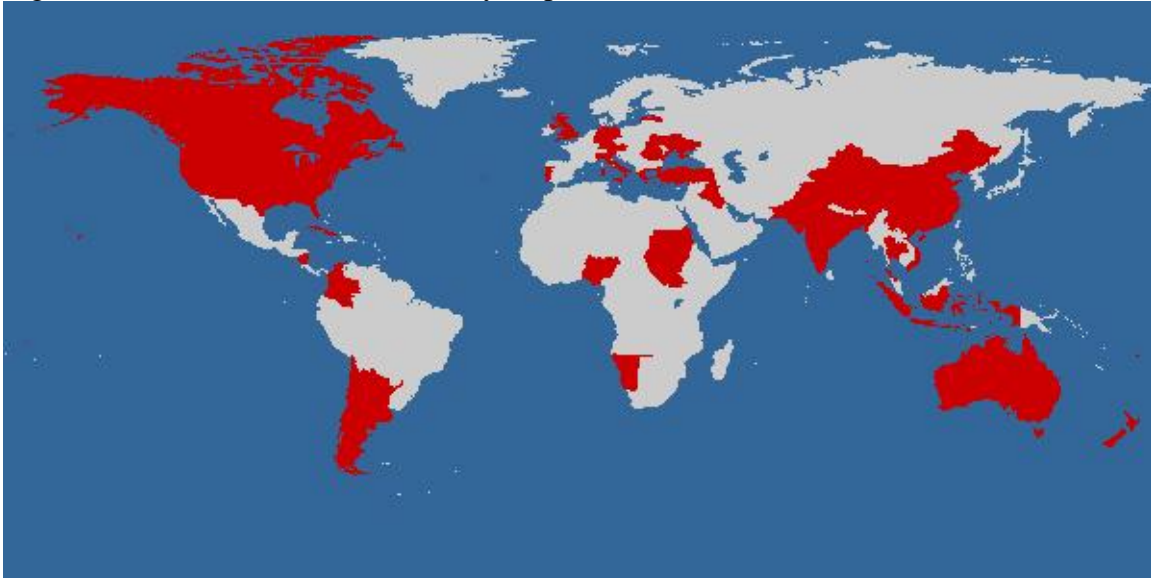
Of the 62 respondents, 90% indicated they work with Greenstone in a single country (including those who distribute or target collections to audiences in several countries), and 10% in multiple countries. Additionally, 3 respondents indicated that though they work with Greenstone in one country their collections are developed for users elsewhere as well. 87% of the responses were submitted via the web survey, and 13% in document format, which the respondent acquired either through email or by downloading the document from the survey website.

Table 2: Countries in which survey respondents work with Greenstone.

<u>Africa</u>	<u>Europe</u>	<u>Latin America</u>	<u>North America</u>
Namibia	Austria	Argentina	Canada (2)
Nigeria	<i>France*</i>	Chile	United States (21)
Sudan	Germany (2)	Colombia (2)	
	Greece	Cuba	<u>Oceania</u>
<u>Asia</u>	Italy	Latin America - all	Australia (3)
Bangladesh	Latvia	Nicaragua	Fiji (2)
China (2)	Portugal (2)		New Zealand
India (6)	Romania	<u>Middle East</u>	Samoa
Indonesia (2)	<i>Spain*</i>	Iraq	Tonga (2)
Pakistan (2)	United Kingdom	Turkey	
Thailand			<i>Global</i>
Vietnam			<i>distribution*</i>

** Countries/Regions in which respondents do not work with Greenstone but are home to primary target audience(s)*

Figure 1: Global distribution of survey respondents.



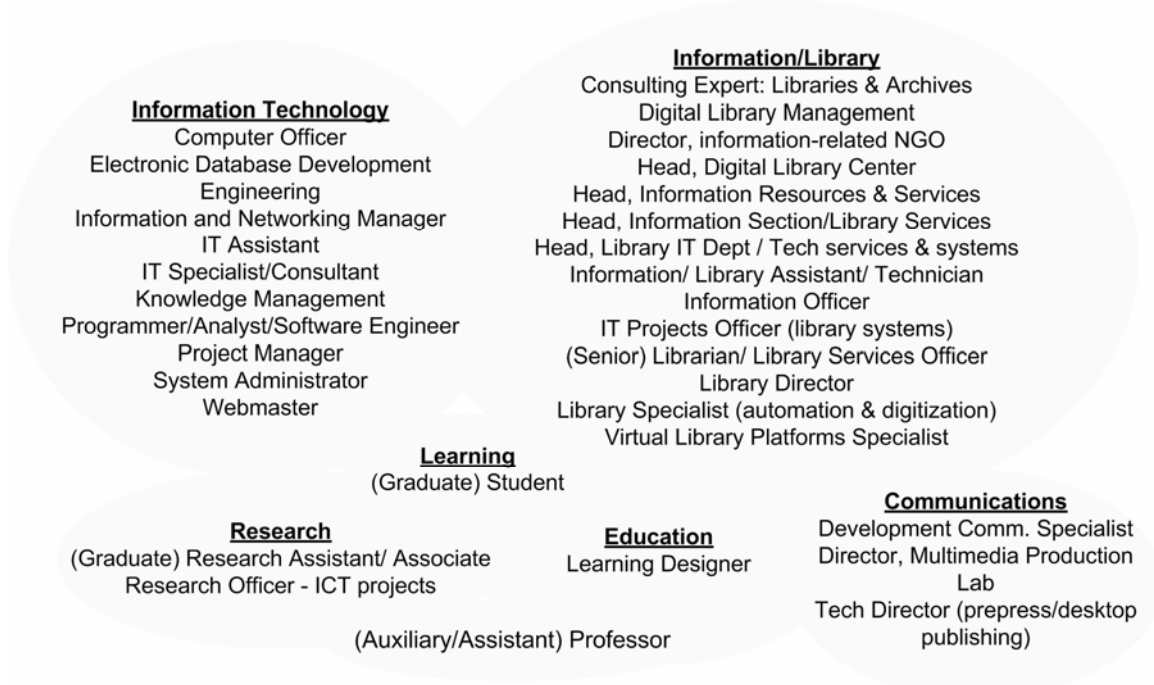
Respondents, their environments, and collections

Most survey respondents (57%) considered themselves to be basic or occasional Greenstone users or developers rather than intensive (knowledgeable and regular) ones. It is evident from the 46 respondents who indicated their professional function or title that the pool includes individuals who perform many functions and hold a wide variety of titles related to Information Technology, Libraries and Information Services, Communications, Research and Education and Learning (see Figure 2). When asked to describe the organizational context in which they used Greenstone, half the respondents indicated that they worked with it in a university or academic institution, 26% at a public service (not for profit) organization at the national level, 20% at a regional or international organization, 20% in an individual capacity, and 7% (3 respondents) in a commercial enterprise. Additionally, 3 people indicated they used Greenstone in other organizational contexts; 2 at public libraries and one as a consultant for clients in international development.

Respondents

When asked to describe the functional modes in which they use Greenstone, half of those who answered this question (and 48% of all respondents) said they worked with it in multiple functional modes. Nearly all (93%) develop digital library collections with Greenstone. 33% said they use Greenstone to teach about digital libraries, 26% promote or disseminate Greenstone, 11% develop new Greenstone functions, and 7% develop language interfaces for Greenstone. Of those who indicated how many Greenstone collections they had developed, over half had developed between 1 and 5 collections (56%), about 28% 6 or more collections, 11% were developing their first collection, and 6% had not developed any collections. 7 respondents did not answer this question.

Figure 2. Representative functions performed and titles held by survey respondents.



Access to collections

Given a list of terms and asked to choose which best described access to their Greenstone collections, 50% said they were open to the public; 39% that they were for staff (23%) and/or organizational affiliate (20%) use only; 2 people developed personal or individual collections; and 1 has demo collections available for customer use and develops collections for others, which are available to various user groups. 14% did not indicate a user group able to access their collections but did indicate the mode of access. Nearly a third (32%) said their collections were available via CD-ROM, half (50%) responded that their collections were available via the internet, 27% via a local network, and 21% who answered other access questions did not indicate access via the internet, local network, or CD-ROM. It is likely that at least some of these collections are available on a single computer, an option that was not an available selection. Of the 18 respondents who indicated their collections are available via CD-ROM, 8 said access was via CD-ROM only, 7 internet and CD, 2 a local network and CD, and 1 via all three modes of access. Respondents with collections on CD-ROM are not limited to any single geographical location. They represent a mixed group of organizational types, including academic institutions, NGOs, government agencies, libraries, and private enterprises.

Operating systems

Almost 40% of respondents who indicated the type(s) of operating systems indicated they had installed Greenstone on multiple systems. Windows XP was the most popular, at 52% of respondents, closely followed by Linux at 41% and Windows NT/2000 at 35%. In terms of Microsoft versus non-Microsoft, 72% of organizations

install Greenstone on some version of Windows and 52% installed Greenstone on a non-Microsoft system. As a percentage of the total installs reported, the picture looks slightly different. Microsoft installs account for approximately 63%, Linux 26%, and others (Macintosh and Solaris) 10%. In contrast, download statistics from SourceForge, the leading distribution center for Greenstone, show that of the average 1800/month Greenstone downloads approximately 80% are Windows binaries, 15% Linux binaries, and 5% source code [3].

Target audiences

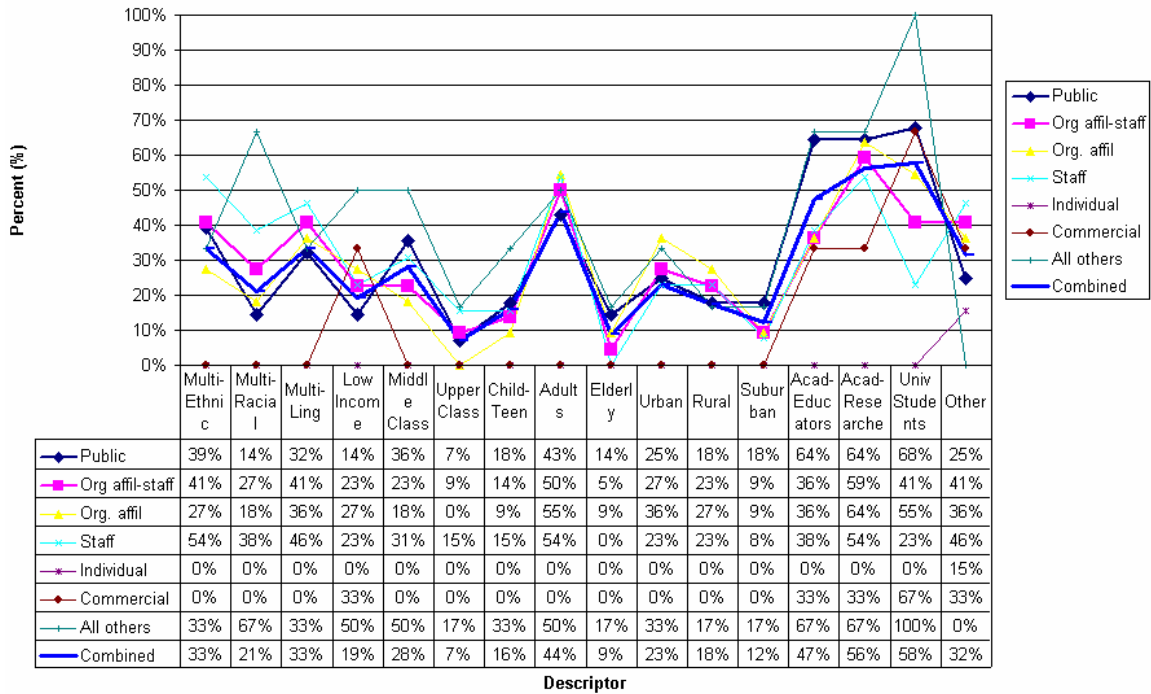
In order to get a basic understanding of the diversity of target populations for which Greenstone collections are built, survey participants were asked to select from a list of 15 descriptive terms to describe their target audience. Terms related to academia and higher education were most frequently selected, all at around the 50% mark (academic – educators 47%; academic – students 58%; academic – researchers 56%). Respondents also indicated that their collections were primarily for adults (40%). 16% of collections developed were for use by Children or Teens, and 9% by the Elderly. 33% indicated that the communities who use their collections are multilingual, 33% multiethnic and 21% multiracial. Just over a quarter of all target populations (28%) were described as middle class, 19% low income, and 7% upper class. Collections were more likely for urban and rural communities (23% and 18%) than for suburban communities (12%). Almost a third (32%) of all respondents chose other descriptive terms to describe the communities that use their collections, see Table 3 for a representative sample.

Figure 3 breaks down descriptor frequencies based on user groups: open to the public, open to organizational affiliates and/or staff only, open to organizational affiliates only, open to staff only, commercial organization, individual/personal use, all others, and combined.

Table 3. ‘Other’ descriptors for collection target audiences.

- | | |
|--|---|
| • extension officers (2) | • personal health |
| • farmers | • personal/individual use (2) |
| • federal & provincial govt. officials | • "public at large" |
| • in-house test bed | • public school educators |
| • low-budget organizations | • public school students |
| • multi-cultural | • researchers: govt., science, agriculture, NGO-based (4) |
| • multi-disciplinary academic | • users in developing countries |
| • national policy makers (2) | • in test phase, audiences would include... |
| • NGOs (3) | |
| • our client's customers, which include... | |

Figure 3. Descriptor selection by end-user group.



Support structures and mechanisms

One of the main objectives of the User Survey was to gather feedback from users and developers about the adequacy of current support structures and mechanisms, and how support might be improved or augmented in the future. Current support includes an active Greenstone Users’ mailing list, which is primarily for those who install and use Greenstone to build digital collections, a Developers’ mailing list, and more recently a Spanish language mailing list and one for version 3 of the Greenstone software. There are also users and developers manuals available in several languages, example collections online, an archive of the users’ and developers’ mailing lists, demonstrations and workshops at conferences, and there have been a number of workshops and training courses sponsored by various organizations in a number of different countries. Half of the 40 survey questions focused on issues related to support and awareness of Greenstone.

Comparative ratings for support resources by type

When asked how they learned to use Greenstone and to rate the usefulness of different learning tools, including the manuals, mailing lists, consulting with more experienced users both inside and outside their organization, and training courses, respondents gave emailing lists and manuals the highest marks. On a scale of 1-3, emailing lists received an average rating of 2.44; manuals, 2.37; the Greenstone website 2.00; consulting within the respondents’ organization 1.71; consulting outside the respondents’ organization 2.04; and attending a Greenstone training course, 1.69. While it should be remembered that survey participation was promoted primarily through the

email lists, survey participant comments throughout the survey cite both specific and general experiences with the email list and are overwhelmingly positive.

It should be noted that some respondents included suggestions for improvements to the mailing list archives as an aspect of the email lists and some as an aspect of the website. In fact, this should have been expected because the availability of the email list archives through the website presents a fluid boundary between these two support resources. The lack of a distinct division between the email lists and the website may have influenced the ratings of the usefulness of these resources, but it is unclear how each was affected.

Email lists

When asked what could be done to improve the mailing lists, many who responded (59%) indicated that the lists were good or helpful in their current format, 21% suggested some sort of improvement. Regarding the list itself, users suggested (1) a more cooperative model of answering questions - that all list members reply to postings for help rather than relying exclusively on NZDL developers at the University of Waikato; and (2) regional or country mailing lists. Subsequent to the survey, the NZDL team has explicitly requested greater participation from mailing list participants, a call to which some list members have responded. Other suggestions regarding the mailing list pertained to the list archives available through the Greenstone site. A number of users (10) indicated that they thought information from the lists should be made more accessible through additional channels - either by indexing and/or giving subject headings to the list archives available through the website, or another means such as by extracting information from the email lists for inclusion in the manuals or online FAQ. Others (5) indicated they would find an online forum helpful.

Website

Comments related to website improvements fall into four basic categories: (1) Website functionality, especially as related to access to materials and navigation (24%); (2) requests for specific content (13%; e.g. information on Sun Solaris and "a section with support materials for giving conferences and courses on Greenstone"); (3) general content suggestions (16%; e.g. a searchable knowledge base of FAQs, addition of more example collections); and (4) aesthetics - suggestions related to the look of the Greenstone website (5%). Some respondents (21%) thought improvement was not needed.

Manuals

More people (74%) responded to questions regarding the manuals than they did to those for the website or email, and, as noted earlier, the manuals were rated highly as a learning resource (2.37/3.00). Suggestions for improvement to the manuals can be grouped according to three of the four classifications used for website comments: (1) Almost half (46%) of those who responded thought that the structure and/or functionality of the manuals could be improved, and the majority of these respondents (86%) suggested that some sort of tutorial, "how to" or "cookbook" type of format would be especially helpful. Several respondents indicated that bookmarks in the PDF files,

indexing, or the creation of other access points to the content would be useful. (2) A significant number of respondents suggested specific content to be included in the manuals (30%). Common requests were for more detailed information regarding macros and interface design customization. (3) Over half of those that suggested improvements to the manuals made general comments for content improvement. Included here are also the comments indicating 'how-to' documents would be helpful, for this is presented as both a function of the content and the structure of presentation. A significant number of respondents to this question (20%) commented that the manuals were too technical in nature, and even more (41%) that more detailed information is needed.

Based on majority opinion, it seems that the ideal manual would be more concise, more detailed, with more access points, and (for most respondents) less technical. It would also include stepwise explanations and tutorials and comprehensive reference lists detailing configuration and customization options. In support of these needs, the NZDL Development Team made a set of step-by-step tutorials available via the Greenstone website subsequent to survey data collection.

Training courses

The low ratings of training courses and the low number of respondents who reported attending a Greenstone training course is significant because many survey respondents thought training courses and workshops should be expanded. The survey question that elicited this rating did not differentiate between type or format of the course, and perhaps these factors (or even specific instances of training courses) had a negative impact on training course ratings. It is also possible that the benefits participants receive from a course are more valuable as introductory, networking, or other opportunities, and the amount learned in the course is not a good measure of its overall value.

Local and regional support structures

In four separate questions that had to be answered in order to submit the survey, respondents were asked whether there was: (1) enough awareness of Greenstone in their country, (2) adequate training available on Greenstone in their country, (3) adequate awareness of existing Greenstone collections in their country, (4) adequate support for developing Greenstone collections. By a solid majority, 68%, 74%, 76%, and 65%, respectively, the response to each of these questions was 'No'. Only 21% of respondents indicated that they were aware of support organizations in their country or region, and just less than half of these respondents indicated what the organization was. One other question survey participants were required to answer was whether there was adequate support for adapting Greenstone to the needs of their country. Response was split, with 43% indicating there was enough support for adapting Greenstone to their country and 55% that there was not. While a larger proportion of respondents felt they did not need more support for adapting Greenstone to national needs, suggestions from those who did were very similar to suggestions gathered in response to the first four questions. Subsequent to the Survey, several 'Internationalization' efforts have been launched that try to address region-specific issues in training, awareness-building, and support.

When asked how shortfalls in awareness, training, and support might be overcome, ideas offered in reply to each of these questions show some overlap, which might suggest that participants also see overlap in the issues of awareness of the software, training, support, and awareness of collections built with Greenstone. Many people commented on the needs to (1) increase awareness, promotion, and diffusion through professional organizations, presentations and workshops, grassroots networking, alliances with Information and Library Science and Computer Science departments and awareness of example—or exemplary—collections; and (2) establish local resources, especially national or regional trainers or centers, user groups, mailing lists and other means of locating and communicating with other Greenstone users. With respect to languages, some suggested that there is a need to create documentation, mailing lists, training kits and/or example collections in their local languages, including Arabic, German, Romanian and Spanish. There were also a few suggestions to improve language handling capabilities of Greenstone: fuzzy stem searching in Spanish, handling of Chinese characters and Bangla text, Arabic handling with the CDS-ISIS plug-in, and enabling multilingual collections. Notably, nine respondents commented that they would or have been working to meet training and other needs at the local, regional, or national level.

Regarding awareness of Greenstone, increased diffusion and exposure to the software was cited several times as being important. In particular, the need for adoption by different types of organizations was noted. Several respondents indicated that they thought awareness would increase if some high profile organizations adopted Greenstone and if this were recognized. Others indicated that the key was for more organizations similar to their own—geographically or organizationally—to adopt Greenstone as well as increasing visibility at conferences, workshops, and other public forums. Training also received much comment. Suggestions for types and modes of training span the spectrum. Some thought that the development of a training kit in their local language would be enough to enable local organizations to provide training on local, regional, or national levels. Others indicated needs for informal user groups; still others a need for more formal mechanisms such as workshops, courses, local centers and local trainers.

At the other end of the spectrum, some participants indicated that there are broader issues that complicate the diffusion of Greenstone in their regions. Issues raised include a lack of familiarity with digital libraries or open source software in general and also the small size of the Information and Library Science academic community and lack of digital libraries courses. In some cases, lack of local resources and geographic and linguistic isolation were cited as impediments to the development of support structures.

Conclusion

Up until now there has been no formal mechanism to learn about who is using Greenstone, where, in what capacity, and their needs for support. The Greenstone User Survey has given us the opportunity to render a rough sketch of Greenstone users and developers, and their needs for support. It has also suggested some new avenues for the development of support mechanisms, and for further research. The ability to build support structures is crucial to the continued diffusion and sustainability of Greenstone. In practical terms, steps have already been made to provide some resources and services that survey respondents thought useful, including the provision of tutorial exercises, more

information about learning and training opportunities, and encouragement for a broader spectrum of list members to be actively involved in answering email list questions. Additionally, several possible contacts, organizations and resources were suggested for developing localized support structures and mechanisms.

Data gathered with the survey not only provides valuable insight into at least one spectrum of the Greenstone user base, but also suggest future directions for gaining a greater understanding of Greenstone users. The survey captured many concrete measures to promote the awareness of the software, assist in community development, and help establish and improve current support mechanisms. However, though Greenstone is a global resource, many respondents focused on the need to develop local people, organizations, resources, networks and communications. The development of personal connections within Greenstone users might be a valuable resource on which to build future research. It would also be helpful to learn more about how Greenstone is used by individuals and organization who develop library collections, and the target audiences for whom they develop the collections.

References

1. Greenstone User Survey (2004).
<http://www.ils.unc.edu/~sheble/greenstone/survey.html>.
2. Greenstone Digital Library Software website (2006). <http://www.greenstone.org>.
3. Zhang, A.B., Witten, I.H., Olson, T.A. and Sheble, L. (2005) "Greenstone in practice: Implementations of an open source digital library system." *Proc American Society for Information Science and Technology Annual Meeting*, pp. 769-794, Charlotte, North Carolina, October.