© The Energy and Resources Institute 2011

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers, TERI Press, The Energy and Resources Institute, New Delhi 110 003, India.

The publisher does not assume any responsibility for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, instructions or ideas contained in the material herein.
Contents

WDL (World Digital Libraries)
Volume 4, Issue 1 • June 2011

1 A simulation study for a number of simultaneous users of an institutional knowledge repository system
Gayatri Doctor

9 Carbon footprint of the knowledge industry and ways to reduce it
Prof. Gobinda Chowdhury and Prof. Michael Fraser

19 Comparison of open source software for digital libraries
Priti Rani Rathour and Ashok Kumar Sahu

33 Vision 2021 and public libraries: an action plan to allow Bangladesh to go digital
Md Nasiruddin

49 Education for digital libraries: library management perspective
Elena Macevičiūtė

63 The digital repository of the Swiss Federal Archives: a case study
Jean-Marc Comment

69 Information search behaviour in the German Education Index
Carola Carstens, Marc Rittberger, and Verena Wissel

81 News
A simulation study for a number of simultaneous users of an institutional knowledge repository system

Gayatri Doctor, Associate Professor, Faculty of Technology Management, CEPT University, Ahmedabad
gayatridoctor@yahoo.com

Abstract
Use of an emerging technology for the creation of Institutional Repositories to facilitate knowledge sharing and management among the faculty in academic institutions in India is becoming a necessity even with constraints like infrastructure, manpower and funding. The paper describes a pilot institutional knowledge repository implementation at a Business school in India and presents a simulation conducted to ascertain the number of simultaneous users who can search and view/access the contents of the institutional repository with a minimal server configuration.

Keywords
Institutional Repository, Knowledge Sharing, Knowledge Management
**Introduction**

Use of information technology to facilitate knowledge sharing and management among the faculty in academic institutions in India is becoming a necessity to improve the quality of education. Institutional Repositories are emerging technologies for capturing structural intellectual capital, knowledge sharing and management in academic and research institutions, especially in a developing country like India (Hayes, 2005). Institutional repositories may include pre-prints, technical reports, working papers, thesis & dissertations, conference proceedings, teaching materials etc. …the intellectual capital of the academic institution (Drake 2004). Institutional Repositories are digital collections capturing and preserving the intellectual output of a single or multi-university community (Crow, 2002).

A business school’s intellectual capital is the sum of its human capital (talent- skills and knowledge of the people- faculty & students), structural capital (intellectual property – the published scholarly material of its faculty, methodologies, software, documents – technical reports, question banks and other knowledge objects) and customer capital (client relationships- students, corporate) (Stewart, 2001).

Creation of a pilot institutional repository at Icfai Business School, Ahmedabad was initiated to enable knowledge sharing; capture, maintain and provide visibility to the structural intellectual capital of the faculty and business school (Doctor & Ramachandran, 2007).

Infrastructure, manpower and funding are constraints in any academic institute especially in a developing country like India. The objective of the paper is to present a simulation done at a business school in India to determine the number of simultaneous users who can search and view/access the contents of an institutional repository with a minimal server configuration depending on the type, number & size of documents being accessed.

**Pilot institutional knowledge repository**

The pilot institutional repository server is installed on a minimal configuration of a Pentium IV @ 2.4 GHz with 256 MB RAM, 40 GB Hard disk. It is available on the Institute Intranet @ 100 MBps for access by the full time faculties at the Institute.

The pilot institutional repository installation is implemented using Open Source Software - Fedora Core 4 Linux, Apache Tomcat 5.5.17, Apache Ant 1.6.5, J2EE 1.4 SDK Java platform, Postgres SQL 8.1.3, and DSpace Version 1.3.2 (Doctor & Ramachandran, 2008 (a)). A screenshot of the homepage is shown in Figure 1.

Considerations for implementing the Institutional Repository like type of documents faculty would contribute, motivation to contribute were identified with a survey conducted among the faculty of the business school (Doctor & Ramachandran, 2008 (b)). Populating the pilot institutional repository was started with the scholarly published materials like articles, journal, conference papers and research reports. Figure 2 shows the growth of the repository.

The Institutional Repository was slowly growing, in size (number of records) and usage (concurrent number of users). The types of documents that the Institutional Knowledge Repository consisted of were in Word, PDF or PPT format. The size of these documents varied.
A simulation study for a number of simultaneous users of an institutional knowledge repository system

from 100KB to 2.5 MB. Faculties would search and view the documents available in the Institutional Knowledge repository (Doctor, 2007).

The need for a simulation arose to ascertain the number of simultaneous users that the current minimal configuration could support depending on the type, number & size of documents being accessed. The pilot institutional repository server on a minimal configuration and a performance & load testing tool were considered for the simulation.

Simulation

The Simulation Software considered was Mercury Load Runner 8.0. LoadRunner is a performance and load testing product by Hewlett-Packard (since it acquired Mercury Interactive in November 2006) for examining system behaviour and performance, while generating actual load. LoadRunner can emulate hundreds or thousands of concurrent users to put the application through the rigors of real-life user loads.

Load testing typically consists of five phases: planning, script creation, scenario definition, scenario execution, and results analysis as shown in Figure 3. Working in LoadRunner involves using three different tools which are part of LoadRunner. They are Virtual User Generator (VuGen), Controller and Analysis.

The Virtual User Generator (VuGen) allows a user to record and/or script the test to be performed against the application under test, and enables the performance tester to play back and make modifications to the script as needed. The Virtual User Generator captures end-user business processes and creates an automated performance testing script, also known as a virtual user script. Once a script is prepared in VuGen, it is run via the Controller. Each run is configured with a scenario, which describes which scripts will run with how many virtual users. The Controller organizes drives, manages, and monitors the load test. Once a scenario is set and the run is completed, the result of the scenario can be viewed via the Analysis tool. This tool takes the completed scenario result and prepares the necessary graphs.

In order to simulate real user behavior, three Data sets were considered involving users searching & viewing files of different sizes. The different Data Sets are shown in Table 1. In

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Total Files</th>
<th>Size of Files</th>
<th>File Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>100 KB</td>
<td>Word</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500 KB</td>
<td>PDF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MB</td>
<td>PPT</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1.5 MB</td>
<td>Word</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 MB</td>
<td>PDF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5 MB</td>
<td>PPT</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>100 KB</td>
<td>Word</td>
</tr>
<tr>
<td></td>
<td></td>
<td>500 KB</td>
<td>PDF</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 MB</td>
<td>PPT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5 MB</td>
<td>PPT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 MB</td>
<td>PPT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5 MB</td>
<td>PPT</td>
</tr>
</tbody>
</table>
Data Set 1, the file sizes considered are small, in Data Set 2 the file sizes considered are large and in Data Set 3 the file sizes considered are small and large. Typical user behavior was searching and viewing the files sequentially. Virtual User Scripts for searching and viewing files in the three Data Sets were generated and executed such that they emulated real user behavior. Figure 4 is a flowchart depicting a typical user behavior.

On the basis of the faculty usage of documents during teaching, research and other activities, six possible scenarios were considered. During teaching activities, the usage of Word & PPT documents was prominent, whereas during research activities, usage of word & PDF documents was prominent. The scenarios considered can be seen in Table 2.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>All users are searching only Word files</td>
</tr>
<tr>
<td>2</td>
<td>All users are searching only PDF files</td>
</tr>
<tr>
<td>3</td>
<td>All users are searching only PPT files</td>
</tr>
<tr>
<td>4</td>
<td>One user is searching Word files, one User is searching PDF files and one User is searching PPT files. Word-PDF-PPT (1-1-1)</td>
</tr>
<tr>
<td>5</td>
<td>Two users are searching Word files, two users are searching PDF files and one user is searching PPT files. Word-PDF-PPT (2-2-1)</td>
</tr>
<tr>
<td>6</td>
<td>Two users are searching Word files, two users are searching PPT files and one user is searching PDF files. Word-PPT-PDF (2-2-1)</td>
</tr>
</tbody>
</table>

**Table 2** Different scenarios

Figure 4 Flow chart depicting user behaviour

For each Scenario, the average time taken, transactions passed & transactions failed of 5 runs were recorded. Depending on the type, number and size of documents that are being accessed by faculty, the number of simultaneous users accessing the Institutional Knowledge Repository Server available on a minimum configuration were to be identified.

**Resultant graphs**

Considering each Data Set and all the six scenarios, the time taken, the transactions passed, the transactions failed and when the server hangs were recorded. Resultant graphs for all six scenarios of the three Data Sets were generated. Figure 5 depicts the time taken, transactions passed, transactions failed with varying number of users simultaneously accessing the pilot.
Figure 5 Scenario 1 – Word Files

Figure 6 Summary of Scenarios 1, 2, and 3

Summary graphs

The recording of the number of transactions that failed and when the server hangs are important parameters to determine the maximum number of simultaneous users that the system can support.

Summary Graphs for the Scenarios of Data Set 1, Data Set 2 and Data Set 3 with recording of the number of transactions that failed and when the server hangs were generated.

Data set 1

Figure 6 is a summary graph of the recordings of the transactions failed and when the server hangs for Scenario 1, 2 and 3 for Data Set 1. It can be observed that the server hangs when 35 simultaneous users are accessing the repository system.

Figure 7 Summary of Scenarios 5 and 6

Thus, irrespective of the type of files being accessed, with 3 files of size 100KB, 500KB and 1 MB the Institutional Repository Server hangs with 35 simultaneous users for the given data set of files.

Data set 2

Figure 8 is a summary graph of the recordings of the transactions failed and when the server hangs for Scenario 1, 2 and 3 for Data Set 2.
Figure 9 is a summary graph of the recordings of the transactions failed and when the server hangs for Scenarios 5 and 6 for Data Set 2. Thus, irrespective of the type of files being accessed, with 3 files of size 1.5MB, 2 MB and 2.5MB the Institutional Repository Server hangs with 30 simultaneous users for the given data set of file set of files.

Data set 3

Figure 10 is a summary graph of the recordings of the transactions failed and when the server hangs for Scenario 1, 2 and 3 for Data Set 3. Figure 11 is a summary graph of the recordings of the transactions failed and when the server hangs for Scenarios 5 and 6 for Data Set 3.

From these summary graphs, it is observed that irrespective of the type of files being accessed, with 6 files of size 100KB, 500 KB, 1MB, 1.5MB, 2 MB and 2.5MB the Institutional Repository Server hangs with 19 simultaneous users for the given data set of files.

Inferences

The number of simultaneous users accessing the Institutional Knowledge Repository depends on the type, the size and the number of documents being accessed. Table 3 denotes the type, number and size of documents that the simulation takes into consideration. Table 4 depicts the number of users accessing the server, the transactions failed for the different datasets & user scenarios.

It can be seen from Table 4 when Data Set 1 & 2 are considered with 3 files, irrespective of the type & size of the file, 10 – 20 users can simultaneously access the server and the number of transactions failed are approximately <5. The Institutional Knowledge Repository server hangs when 30 -35 users simultaneously try & access the server.

When Data Set 3 is considered with 6 files, irrespective of the type & size of the file, 10 – 15 users can simultaneously access the server and the number of transactions failed are approximately <5. The Institutional Knowledge Repository server hangs when 18-20 users simultaneously try & access the server.

It can also be observed that irrespective of the type of file, number of files and size of files only...
A simulation study for a number of simultaneous users of an institutional knowledge repository system

Table 3  Document types, number, and size

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>3</td>
<td>Small (100KB, 500KB, 1MB)</td>
</tr>
<tr>
<td>PDF</td>
<td>6</td>
<td>Large (1.5 MB, 2MB, 2.5 MB)</td>
</tr>
<tr>
<td>PPT</td>
<td></td>
<td>Small + Large</td>
</tr>
</tbody>
</table>

Table 4  Transactions failed

<table>
<thead>
<tr>
<th>User scenario</th>
<th>Data set</th>
<th>Users</th>
<th>TF=0</th>
<th>TF &lt; 5</th>
<th>TF &lt; 10</th>
<th>Server hangs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td></td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>1</td>
<td>3</td>
<td>21</td>
<td>24</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>21</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>15</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>5, 6</td>
<td></td>
<td>1</td>
<td>3</td>
<td>20</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

three users can access the system simultaneously and have the transactions failed as zero.

The simulation exercise helped determine that about 10-20 simultaneous users with transaction failures < 5 are supported with the current configuration that the pilot Institutional Repository was built on with moderate usage.

In Business School like IBS Ahmedabad where there are only 25 full time faculties, the number of faculties simultaneously accessing & searching the Institutional Repository server would definitely be less than 20. Thus, in the initial stages of the Institutional Repository development and the pilot testing, this is minimum configuration is sufficient.

Infrastructure like disk space, memory, network access etc, manpower and funding are always the initial bottle necks in any implementation. In an academic institute, especially in a developing country like India, determining, justifying and sustaining requirements is an important aspect. With the growth of the Institutional Repository, in size (number of records) and usage (concurrent number of users) the need to support more users would arise.

Conclusion

Even with constraints like infrastructure, funding a business school with a small number of faculties can initiate knowledge sharing by implementing an Institutional Knowledge Repository with a minimal server configuration.

The advent of improved Open Source Software Versions, the growth in size, usage and moving to an Online Repository (available for access on the Internet) would involve challenging tasks for migrating the pilot Repository server to a system with a higher configuration.

The simulation experiment can be repeated with a server with a higher memory like 512MB or on a machine with a higher configuration. Depending on the number of faculties and the requirements of the Business school upgrading the memory to 512MB can be considered.

Knowledge sharing benefits the faculties, the institution and promotes the welfare of society. Thus, it is critical that business schools in India should capture the intellectual capital, facilitate knowledge sharing among the faculty by initiating the creation of an institutional repository.

References


Doctor G. 2008. **Determining the number of simultaneous users of an Institutional Knowledge Repository at a Management Institute in India.** *VINE – The journal of information and knowledge management systems* 38(3): 334–347


Icfai Business School (IBS), retrieved on 13 August 2010. Available at http://www.ibsindia.org


Mercury Interactive, retrieved on 13 August 2010. Available at http://en.wikipedia.org/wiki/Mercury_Interactive
Carbon footprint of the knowledge industry and ways to reduce it

Prof. Gobinda Chowdhury, Director, Centre for Information and Knowledge Management, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia
Gobinda.Chowdhury@uts.edu.au

Prof. Michael Fraser, Director, Communications Law Centre, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia
Michael.Fraser@uts.edu.au


Abstract

The current practices of content creation, access, and super distribution followed by the global content industry are neither economically nor environmentally sustainable. Since research studies on the carbon footprint of the content industry are few and far between, adequate data are not available. This paper shows that the current practices in the production of printed content produce a massive carbon footprint. It is argued that these emissions can be significantly reduced by developing a digital content supply network. Features of a digital content network are briefly discussed along with an indication of how this can help reduce carbon footprint of information services. The paper also points out areas of further research in this direction.
Introduction

Since the beginning of the Gutenberg era, the information industry has followed a typical business model, where content is produced as bundled products such as a book with several chapters, a journal issue with a pre-determined number of papers, and so on. A certain number of such bundled products—copies of books, copies of a certain issue of a journal, and so on—are produced by publishers in order to attain economies of scale. Such information products are either purchased by consumers for a set price or are purchased and managed by libraries for use by their clients. This model continued till about 15 years ago, with some minor changes taking place during the past four decades or so due to the arrival of remote online databases, which provide access to individual articles of journals, conference proceedings, and so on, via subscription. Throughout this period, the basic model for content access was either through purchase of bundled content by the consumer or access through libraries, barring some exceptional types of information resources such as free newspapers, institutional reports and publications, pamphlets, and so on. Essentially, consumers were either required to pay for the content or access content through libraries. The latter played a key social role in the democratization of access to knowledge.

Although it is not often highlighted as one of the major players in national economies, the size of the content industry and its contribution to the economy is quite significant. A somewhat dated figure reveals:

- **Total annual revenues of US book publishers** amounts to $26.8 billion (2004 figures) (Eco-Libris 2007);
- **Total annual revenues of European book publishers** amounts to Euro 22.3 billion (2004 figures) (Eco-Libris 2007);
- In the UK, the publishing industry, which is the second largest in Europe, has a turnover of over £18.4 billion, with its 8000 plus companies employing around 164,000 people and contributing to over 8% of GDP (http://www.berr.gov.uk/whatwedo/sectors/publishing/index.html);
- The estimated value-addition of the copyright industry in Australia is approximately 100 billion Australian dollars, which is about 10% of the GDP, while the annual compound growth of the industry is 4.7%, as compared to 3.6% of GDP (Price Waterhouse Coopers 2008); and
- Nielsen Book Scan reports that in 2008, 61 million books were sold retail in Australia, with a total value of Aus$1.21 billion (Nielsen 2009).

In 2007, the value of the global book market was estimated to be $127.5 billion, which is estimated to rise to $160.7 billion in 2012 (up 26% from 2007). In 2008, the US book publishers’ net revenues touched $40.32 billion (Healey 2009). If we consider the entire content industry, the size of the market will be several times larger. Yet, the processes of content creation, distribution, and access follow an extremely traditional business model. In this model, consumers have two choices: either to purchase priced content, which is edited or peer-reviewed and is thus reliable in terms of value; or to access free content, which may or may not be edited or peer-reviewed and hence may or may not be of a certain quality. Nevertheless, in most cases, the consumers have to acquire the entire work, even if they want to use only a small portion of the item. Consumers often wish not only to access granular content, but also to aggregate granular content elements and then super distribute the aggregated content package(s). Taking publishing as an example, consumers must either:

- Buy books and subscribe to journals, and then photocopy or scan the granular elements they actually want for physical aggregation and for physical distribution; or
- Go online from one proprietary site to another, and get access to the digital content
in different ways, on different terms and conditions, and then attempt to repurpose and distribute extracts of content within the limits of the various different digital rights management systems and licences.

Till date, the overall process of content access, use, and super distribution, which is very common in academic institutions like universities, remains very cumbersome. Complex and onerous statutory and commercial blanket licensing schemes have been developed to address this market failure. In response, a contested grey economy of copies has been built by one section of the consumer community to form a legitimate market for content that meets consumers' needs.

In this paper, it is argued that the prevailing business model is not environmentally sustainable, and that while other industries and businesses are concerned with reducing their carbon footprint, the processes of knowledge creation, access, and distribution and the content industry that supports these activities are not concerned with properly addressing and mitigating the impact of the industry on matters related to environment and ecology. The lack of detailed studies on the carbon footprint and environmental impact of knowledge-intensive activities hints at the hypocrisy amongst scholars, academics, and researchers who talk about and study the carbon footprint of every sphere other than the knowledge sector. This paper proposes a model of new content service network that will help reduce the carbon footprint of the content industry and of knowledge-intensive activities. However, this initial study has to be supported by more detailed ones on the environmental impact of knowledge creation, distribution, and access.

**Environmental impact of the content industry**

Studies on the environmental impact of the content industry are few and far between. However, the data gathered from these studies reveal some massive figures.

According to certain estimates, 10.2 kg of CO$_2$ is generated by each printed book over its lifetime (Ritch 2009). Therefore,

- 235.6 million books sold annually in the UK (Booktrade.info 2010) would generate 2.4 million tonnes of CO$_2$;
- 1 billion books produced annually in Germany (Space Daily 2009) would generate 10.2 million tonnes of CO$_2$; and
- 4.15 billion books produced in the USA (Eco-Libris 2007) would generate 42.3 tonnes of CO$_2$.

Thus, the annual book production in Germany and USA combined would generate 52.5 million tonnes of CO$_2$.

To put these figures into perspective, the CO$_2$ emission from power plants—the world’s most significant contributors to greenhouse gases—in various countries are as follows (Center for Global Development 2007):

- US: 2.8 billion tonnes;
- China: 2.7 billion tonnes;
- Russia: 661 million tonnes;
- India: 583 million tonnes;
- Japan: 400 million tonnes;
- Germany: 356 million tonnes;
- Australia: 226 million tonnes;
- South Africa: 222 million tonnes;
- UK: 212 million tonnes; and
- South Korea: 185 million tonnes.

(Note: The figures above are in US tonnes, where one tonne is equivalent to 2,000 pounds, as opposed to a metric tonne, which is equivalent to 2,204 pounds).

Another way of understanding the environmental impact of the content industry could be to make a comparison with the following figures:

- The average carbon footprint per person in the UK is 10 tonnes (www.climatechangewales.org.uk/public/?id=112) and
More than 30 million trees are cut down annually to produce books sold in the US alone (Eco-Libris 2007)

While the total carbon footprint figure for the entire journal publication industry is not available, a study commissioned by a leading journal publisher Reed Elsevier noted that the total carbon footprint for producing the journal *Fuel* in 2007 was ‘just over 40 tonnes of carbon dioxide, and an ecological footprint of almost 21 global hectares’ (http://www.reedelsevier.com/corporateresponsibility08/PDFFiles/fuel-footprint-study-exec-sum.pdf). The figure in terms of the total carbon footprint of printed journals will be staggeringly high if the total number of all the print journals in the world is multiplied by this figure, and the transportation and distribution costs of print journals around the globe are also taken into account!

**Environmental impact of copying**

The actual carbon footprint figures are cause for greater concern if other factors are taken into consideration. Let us take photocopying as an example. Photocopying under statutory licences is a legal activity, although it is well understood that the total number of copies that are made under the fair dealing provision of the copyright law, and those that are copied illegally, will be several times the figures of legally licensed copy pages. Even these figures are alarming. Copyright Agency Limited (CAL), Australia estimated that in 2008–09, over 2 billion pages were copied legally under licences in Australian educational, government, and some business institutions (CAL 2009). Assuming that an average book contains 300 pages (Kozak 2003), this amounts to about 50,000 tonnes of CO₂ emission.

According to the HP carbon footprint calculator (HP Eco solutions 2009) a large office photocopier (Canon imageRUNNER 3300) consumes 2,736 kWh electricity (assuming that it prints 100,000 pages per year), and emits 3,087 kg of CO₂. As per this calculation, the CO₂ emission for photocopying 2 billion pages will be (at the rate of 3,087 kg per 100,000 pages) 617,740 tonnes.

This figure may vary depending on the type of photocopier and its frequency of use, special facilities like air conditioners, maintenance and repair, and so on. It should be noted that there is an environmental cost involved even when a photocopier is idle. According to one calculation (RISO 2009), the cost of printing 100,000 pages in a month would be $107.88 per year, but the cost of electricity use in idle and sleep mode would be $382.50. The study also states that ‘a mid-volume (20–44 copies per minute) copier in a low-volume office can use 70% more energy per page than an efficient low-volume (under 20 copies per minute) copier.’

However, even using the rough estimate, the environmental cost of photocopying 2 billion pages per year under the CAL license yields a massive 617,770 tonnes of CO₂. It must be noted that this figure of 2 billion photocopies include only those pages that are copied under the license—the number of pages copied for super distribution as permitted within the CAL licences. This does not include the pages that are copied by people for their personal use within the fair dealing provisions and other exceptions in the copyright law. In reality, the number of copies made in this manner will be at least equal to—if not more than—copies made under the CAL licenses. So, even if we make a conservative estimate, the CO₂ emissions for photocopying could be 1.2 million tonnes or more.

To put this into perspective, in Australia, the per capita CO₂ emission in 2007 was 18.8 tonnes (http://www.iea.org/co2highlights/CO2highlights.pdf). Taking into account Australia’s population of 22 million, and considering the figure of 1.2 million tonnes of CO₂ emission from photocopying alone, it may be noted that the per capita CO₂ emission from photocopying alone was 0.055 tonnes, or 55 kg per person.
Now, let us look at this scenario from a different perspective. CO₂ emissions from legal licensed photocopying in Australia are equivalent to the annual CO₂ emissions from four power plants. This observation is based on figures from Carbon Monitoring for Action (CARMA) (http://carma.org/region/detail/18) data, which state that the 766 power plants in Australia together generate 224 million tonnes of CO₂ per year. As mentioned earlier, this calculation does not include CO₂ emissions from photocopying done in the wider corporate sector, nor does it include copying of copyright-free content, copying under exceptions or illegal copying. Taking photocopying of all kinds of knowledge-related activities in all the sectors in Australia into account, it may be noted that the total carbon footprint from photocopying content (books, journals, reports, and so on) is at least equal to that produced by 6–8 power plants of average capacity.

**How can emissions be reduced?**

Now, the question is, can this massive emission be reduced, and can we create a more environment-friendly system for knowledge creation, distribution, access, and reuse? In order to answer these two questions, one needs to identify the major contributors to CO₂ emissions in the prevailing supply chain for knowledge creation, distribution, access, and reuse. The two biggest contributing factors in producing CO₂ emissions in these fields are:

- The number of printed copies of content: books, journals, newspapers, conference volumes, and so on. As noted earlier in this paper, the number of books published in Germany and the USA contribute to 10.2 million and 42.3 million tonnes of CO₂, respectively.
- Photocopying of content: both legally and illegally, which makes a significant additional contribution to the carbon footprint.

How can we reduce these two activities and yet improve the process of knowledge creation and knowledge use that are the two cornerstones of a knowledge society? In the rest of this paper, we propose the architecture of a novel content supply chain that will help improve the economic and environmental sustainability of the content industry, and will thus ensure a more environment-friendly knowledge creation and knowledge access infrastructure.

**Design of the new content service network (CSN)**

Figure 1 shows the basic architecture of a new content service network, which will enable users to choose digital information from a myriad of information channels and sources, with sufficient levels of granularity, required for their use; and produce new information packages for future reference and distribution to a designated audience through a variety of media, including institutional or personal computers, handheld devices, on-demand print, and so on, within the framework of a new and easy-to-implement business model. The network model, once fully built and implemented, will provide all these and many more facilities, thereby creating a demand-driven, real-time content service model that will benefit content creators/producers, as well as consumers. Overall, the new content service network will play a key role in enabling people to make smart use of information.

The proposed CSN will have five major characteristics that will make it unique and among the currently available models in the content market place:

- It will not replace the current print-based or Internet-based model of content supply. Rather, it will be an additional channel, which will facilitate granular content discovery, access, aggregation, and super distribution.
- It will not build a repository of centrally held content resources. Instead, a single window is proposed for seamlessly accessing content directly from the databases where the content...
is already held, such as the author or publisher databases.

- Neither does the system suggest that every piece of content used should be paid for, nor does it recommend that every piece of content be free. It builds an infrastructure that will allow the content creator/producer to decide whether or not the content should be free or paid-for, and the prices can be adjusted by them throughout the lifetime of each piece of content. It will be a content supermarket, where the consumers will be able to pick and choose the content—fee-based, as well as free, side by side—according to their preference.

- It will be a social infrastructure designed to create a level playing field for all types of content creators and producers to make their content available. Content creators and producers would register their content to be accessed through the CSN with a trusted third party, which will be a not-for-profit organization. The system will be built on a set of open standards and protocols, and its operation will be transparent.

- The CSN will include key players such as libraries and various other stakeholders, who add value to the content supply chain. It will create an environment where librarians and other stakeholders (such as market intelligence agencies) can add value in the supply chain, and play a key role in helping consumers get access to quality information through direct and transparent online transactions.

The proposed CSN will be significantly different from the currently available systems and approaches in many ways. It will:

- Create mechanisms for access to—and aggregation of—granular content, including copyright, out-of-copyright, and Creative Commons content currently available from heterogeneous information services, through a single window and a single process, in an on-demand and real-time mode.

- Allow the consumers to lawfully access, aggregate, store, and super distribute content.

- Have an underlying transparent business model that will facilitate all the legal rights management and commercial aspects of the transactions, and thus will hide from the consumers all the complexities of content access, rights management, and the supply chain.

- Create a win-win situation for both consumers and creators/publishers by generating a transparent mechanism for payment and/or attribution for access and super distribution of content.

It is proposed that the content service network will be developed and tested in a B2B transaction model. Once the technology—especially the relevant digital rights management tools and standards—is developed, it can be opened up for B2C transactions. It should be emphasized that the proposed network does not in any way suggest that consumers should pay for all kinds of content. Neither does it suggest that the payment has to be made separately for each transaction. On the contrary, it proposes a framework where the consumers will be free to choose different types of content—free as well as fee-based—in granular form, as required. Consumers will be able to add value, share, and create new knowledge as envisaged in digital library frameworks like the DELOS digital library reference model (DELOS 2006). Further research will help develop the precise systems for pricing and payment for content use and super distribution. Payments may be affected on an e-commerce transactional or subscription model. The other novelty of the network will be that it will provide a one-stop shop for all kinds of activities related to content discovery, access, aggregation, value-addition, use and super distribution, and so on. All these activities can be performed within a transparent legal and business model. The consumer will be able to
monitor and adjust the overall cost for every transaction involving their selection of content.

**CSN: a conceptual view**

The new knowledge infrastructure will create an environment where bundled content (produced by publishers) will co-exist with granular content. It will be up to the consumers to decide which content to use under specific circumstances. For example, some consumers may wish to read an entire book, a complete journal article, a complete issue of a journal or even a complete volume of a journal; while others might wish to use specific sections of books, journal articles, and so on for a specific purpose. Consumers should be able to have on-demand access to the content marketplace and pay for what they use, rather than shell out a pre-determined fixed charge for each whole document, even if only a small portion of it is actually used in a given situation.

The new infrastructure will not only facilitate creation of new content packages by aggregating whole or granular content from a variety of channels and sources, but it will also allow consumers to add their own value in a number of ways. First, the aggregated content package will be created using a deep understanding of the context and consumers—using the knowledge of context. For example, in an academic environment, a lecturer accessing and aggregating content online for super distribution to students will apply the structure and strategy/policy of the university, faculty, and...
department; the nature of the specific course; nature and characteristics of the students; and so on. Secondly, the users (in this case, the academicians, who are accessing the content for aggregation and super distribution) can add value by incorporating their own content in the form of notes, comments, interpretations, and so on. Once the content has been aggregated, the academy can then add its own branding, form, style, and so on. Therefore, the new product will not be just aggregated content as such (as happens in the current environment of printed course packs or electronic course reserves), but will be designed and rendered relevant content packages aggregated from the widest range of resources, and targeted for a specific context and environment of downstream users.

The CSN will have the following characteristics:

- Digital content will remain with the content creator/provider, who will register their content and its granular segments with the CSN.
- Content in granular form will be tagged and indexed using appropriate tools, technologies, and metadata, in addition to its existing metadata. This calls for specific research to find out which currently available tools, technologies, and metadata can be used unchanged or with some modifications/improvements to suit the specific needs of the new knowledge infrastructure.
- The CSN will have access to granular content, while the actual degree of granularity (chapters, sections or paragraphs, tables, figures, and so on) will be determined by market demand.
- Users will be able to search, locate, and access granular content. They will also be able to pick and mix specific granular content to suit their specific needs and the needs of their target audience (for example, a group of university students) with the help of appropriate tools, interface, and support systems. Again, specific research and evaluation studies will show, which of the currently available tools, technologies, and metadata can be used unaltered or with some modifications/improvements to suit the specific needs of the new knowledge infrastructure.
- Users will be able to prepare aggregated content packages in real time, as per their specific needs.
- The aggregated content packages can be made available for interactive use with the help of various Web 2.0 technologies and tools.
- The aggregated content can be packaged/rendered.
- The aggregated content can be saved for future use and/or distributed to a specified target audience.
- The rights for super distribution to the target audience will be managed with the content transaction.

All the above activities can be performed within an easy-to-use and transparent legal and business framework. Thus, the new CSN will establish many-to-many connections among the content creators and consumers. In other words, a given consumer will have access through one interface to a large pool of granular content produced by a variety of creators and publishers—large and small, local and overseas, copyrighted and copyright free. It will be a virtual marketplace, where content creators/providers and consumers will make transactions in real time to access or use content.

In the new CSN, content creators/providers/publishers will have the ability to set the price, if any. However, it need not be fixed throughout the life cycle of the content. Instead, the price can be negotiable and adjustable depending on the frequency and volume of access, as well as on super distribution. Different prices may be set for different categories of customers or customers in different locations (even in different countries). There can be a differential pricing system in place that will benefit both the consumers and creators/providers of digital content.

It is anticipated that on a global scale this will lead to a multicurrency payment system as an element of the e-business model. All
the transactions will take place through a single window in real time, thereby relieving the consumers from dealing separately and differently with multiple content producers/providers and complex different copyright practices and laws.

**Environmental benefits**

One of the most significant indirect social benefits of the CSN will be the significant reductions in carbon footprint. Although further studies are necessary to gather data on the carbon footprint of the content industry in each country—and comparative data can only be produced once the CSN is fully developed and implemented—it can be argued that the new model will be able to reduce the carbon footprint figures significantly because:

- it will enable consumers to move increasingly towards use of information in digital form because of its convenience in access, downstream use, interactive use, novel sharing, and peer review facilities provided by the Web technology. In some cases, content will need to be printed, but it will be print-on-demand only, as opposed to the current book and journal industries’ practice of printing, overprinting, sale or return, and pulping;
- there will be significant reductions in the physical movement of content for distribution in hard copies because content will only be printed as and when required. Consequently, there will occur reductions in return and pulping of unsold copies;
- less space will be required to store printed materials in libraries, bookstores, warehouses, and so on, where quite often special measures have to be taken for conservation of printed materials; and
- carbon footprint will be significantly reduced because within the new virtual supply chain model there will be less requirement for travel by people associated with the content industry.

Some critics argue that reading a book online produces more—or at least as much—carbon footprint as a printed book. Their studies calculate CO$_2$ emissions for the Internet, the computers involved, and the entire network infrastructure. However, this is not justifiable because every government, especially in the developed world, is investing huge resources towards building a broadband network. The network and its underlying infrastructure will operate whether or not it is used for online content transaction and use. In this regard, it might be mentioned that a comparison of CO$_2$ emissions from printed and digital books has been provided by Chowdhury (2010).

**Summary and conclusion**

The current model of content creation and distribution in print form is certainly damaging for our environment. Therefore, alternative and environment-friendly systems for content creation, access, and use should be developed to save the environment. The new content service network proposed in this paper will reduce carbon footprint by eliminating the level of CO$_2$ that is currently generated during the production and distribution of printed content. The proposed digital content service network will create a demand-driven content utility service system for institutional users that will allow them to search, find, access, aggregate, store, use, and super distribute granular content from a variety of free and fee-based content producers and services. It will also create a level playing field, thereby ensuring that small content creators and publishers have the same opportunity as others in making their products available through the virtual content marketplace. Thus, it is envisaged that the CSN will create a social infrastructure for content creation and use. The ultimate power, with regard to content creation and use, will rest with the general public (content creators and users), as opposed to a specific industry or business in the content supply chain. Although, further
research is necessary to develop standards and protocols for digital content creation and access through the network, the new model will support better content transaction and uses, thereby facilitating knowledge-intensive activities in a much more environment-friendly way.

References


CAL (Copyright Agency Limited). 2009. Annual report. Copyright Agency Limited


Comparison of open source software for digital libraries

Priti Rani Rathour, Assistant Librarian, Apeejay Satya University, Sohna, Gurgaon
priti_librarian@rediffmail.com

Ashok Kumar Sahu, Senior Librarian, Institute for International Management & Technology, Udyog Vihar, Gurgaon, Haryana
aksahu@iimtobu.ac.in

Abstract
This study is an examination of features of the four most popular digital library open source software packages against a set of predetermined criteria that is deemed to be essential for the development of a digital library. While analysing and identifying the similarities, differences, strengths, and weakness of the open source software packages, it is indicated that open source digital library software still lacks certain functionalities that are perceived to be important. Each software package has its own individual strengths and weaknesses that will appeal to organizations and stakeholders with different needs. The study is expected to help those library professionals, seriously interested in implementing a digital library by providing a checklist to evaluate how well a particular software fits into their specific implementation requirements. Certain good features of the four open source software packages have also been pointed out via a comparison analysis.
Introduction

Open source software (OSS) refers to any software that is free, and hence, is often confused with freeware and shareware. In contrast to open source, freeware is software that is released free of cost in binary format only. Its licenses usually prohibit modifications and commercial redistribution. On the other hand, shareware is software that is released free of cost in binary format but only for a limited trial period, after which users are encouraged to purchase the software.

The availability of the source code in open source software allows users to modify and make improvements to it. Such contributions can originate from a diverse talent pool of programmers. Thus, open source software tends to have more functions—being developed by users of the software themselves—as compared to commercial software, where a vendor’s priority (usually profit generation) might not be in line with the needs of users. Further, because the source code is accessible and modifiable, contributions also lead to improvements in the functionality of the software. In addition, updates can be obtained at low or even no cost, and there are no royalties or license fees. Moreover, there is less likelihood of being dependent on a single software provider or being trapped into long-term software support contracts, which restrict flexibility in implementation (Surman and Diceman 2004).

However, open source software is not devoid of disadvantages. One of the common complaints is the lack of formal support and training that a commercial software package would offer (Caton 2004). Often, support is provided through mailing lists and discussion forums. In addition, open source software is also not known for ease of use as the focus is usually on functionality. Consequently, open source adopters will have to take greater personal responsibility, in terms of leveraging staff expertise to implement and maintain their systems, including hardware and network infrastructure (Poynder 2001).

Nevertheless, open source is increasingly considered as an alternative to commercial digital library systems, mainly due to dissatisfaction with functionality (Breeding 2002). Another factor is the increasing budget cuts that libraries face (Evans 2005). The cost of software production and maintenance is also rising dramatically. As a result, open source digital library software, with its free access and satisfactory level of functionality, are steadily gaining ground in both usage and interest.

The Open Source Software (OSS) model makes the source code available to users, who can then make the necessary changes to it to tailor it according to their own requirements. With many OSS applications now available for library and information management, organizations now have another option for acquiring and implementing systems, as well as access to opportunities to participate in OSS projects. Examples of such systems include Greenstone, DSpace, E-prints, and so on. OSS is extremely popular with technically sophisticated users, who are often also the software developers. This study highlights the comparison, features, function, and usability of OSS like Greenstone, E-prints, Fedora, and DSpace.

Over the last several years, the rapid growth of OSS has captured the attention of research librarians and has created new opportunities for libraries. OSS can benefit libraries by lowering both initial and ongoing costs, eliminating vendor lock-in, and by allowing greater flexibility (Naik and Shivalingaiah 2006).

Objectives of the study

The main objectives of this study are to:

1. Determine the features that characterize ‘good’ open source digital library software and develop a checklist from these features. Here, we identified the essential categories of features a digital library software should possess. Examples include content
management, searching, and metadata standards. These categories were then subdivided into supporting features. From these, a checklist that covered all such features was developed.

- Evaluate the features of four open source digital library software packages against the checklist. Scores were assigned to each software, depending on its adherence to the features mentioned in the checklist.
- Analyse the strengths and weaknesses of each of the four digital library softwares. After compiling the scores, a comparison of the features was conducted with the objective of identifying similarities, differences, strengths, and weakness of each (Goh D H, Chua A, Khoo D A, et al. 2006).

Four most popular digital library software packages were selected for the study. These were DSpace, Eprints, Fedora, and Greenstone (Jose 2007).

**Digital library software packages**

Many open source software packages are available for organizations and individuals alike to create digital libraries. However, an easy-to-use instrument to evaluate these digital library software packages does not exist. The present work attempts to develop a comprehensive checklist for assessing digital libraries. Its flexibility allows users to tailor it to accommodate new categories, items, and weightage schemes to reflect the needs of different digital library implementations (Goh D H, Chua A, Khoo D A, et al. 2006).

Four digital library packages were selected for our evaluation:

- **Greenstone** (www.greenstone.org) has been developed by the New Zealand Digital Library Project at the University of Waikato, and is distributed in cooperation with UNESCO and the Human Info NGO. Greenstone is a tool for building libraries that aims to empower users—particularly those in universities, libraries, and other public service institutions—to build large distributed digital library collections. The current installed base is unknown, but the number of downloads of the software appears to be large. It is an open source software that can handle multilingual documents, and has a search and browse facility under GNU General Public License (Laxminarsaiah and Rajgoli 2005).

  Greenstone is a tool for creating and managing digital library collections. It runs on Windows as well as UNIX.

  The Greenstone digital library software builds collections with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt automatically. The system is extensible: software ‘plugins’ accommodate different document and metadata types.

- **DSpace**: Massachusetts Institute of Technology (MIT) and Hewlett-Packard (HP) have jointly created DSpace software as a digital repository to manage intellectual output of multidisciplinary research and development organizations. DSpace supports digital preservations and planning and managing institutional repository in a large institution. It also allows workflow and customization and supports community/collection-based content and submission by different user communities (Laxminarsaiah and Rajgoli 2005).

  It is based on a three-layered architecture, namely application layer, business layer, and storage layer. The application layer covers the interface to the systems, the Web and user and interface and batch loader, in particular. The business layer is where lies the DSpace specific functionality, workflow, content management, administration, and search and browse modules. The storage layer is implemented...
using the file system, as managed by PostgreSQL databases.

The system is primarily written in Java, and uses only free software libraries and tools, including PostgreSQL, RDBMS, Java servlet, Apache and tomcat, Lucene search engines, XML tools, and RDF tool. Collections within communities consist of items, which are, in turn, composed of one or more bit streams, or physical files of digital materials. DSpace item is a single bit stream. For example, a digital image encoded as a TIFF file or a digital document encoded as a PDF file.

It is a groundbreaking digital institutional repository that captures, stores, indexes, preserves, and redistributes the intellectual output of a university’s research faculty in digital formats. It manages and distributes digital items made up of digital files (or bit streams), and allows the creation, indexing, and searching of associated metadata to locate and retrieve the items. DSpace also supports submission, management, and access of digital content (Naik and Shivalingaiah 2006).

- **EPrints** (www.eprints.org) is a generic archive developed by the University of Southampton. The first version of the software was publicly released in late 2000. The objective behind the creation of EPrints was to facilitate open access to peer-reviewed research and scholarly literature through OAI. However, EPrints also serves as an archive for other electronic documents such as images and audio (Goh D H, Chua A, Khoo D A, et al. 2006).

  EPrints is intended to create a highly configurable Web-based archive. Its primary goal is to be an open archive for research papers, but it could be easily used for other things such as images, research data, audio archives—in fact, anything that can be stored digitally by making changes in configuration. It works on Linux and needs MySQL, Perl modules and Apache Web server. The software can be installed by any institution across the world. By its integrated advanced search, extended metadata, and other features, the software can be customized to meet local requirements (Laxminarsaiah and Rajgoli 2005).

- **Fedora** (www.fedora.info) has been jointly developed by the University of Virginia and Cornell University. Its first version was released in 2003. The objective of Fedora 1.0 was to create a production quality system using XML and Web services to deliver digital content. Fedora supports digital asset management, institutional repositories, digital archives, content management systems, scholarly publishing enterprises, and digital libraries. The system is designed to be a foundation upon which full-featured institutional repositories and other interoperable Web-based digital libraries can be built. It currently has a distributed installed base of more than 360, with collection sizes of 10 million objects. As an open source software, Fedora offers organizations a flexible service-oriented architecture for managing and delivering digital content. At its core lies a powerful digital object model that supports multiple views of each digital object and the relationships among digital objects. All functions of Fedora, both at the object and repository level, are revealed as Web services. These functions can be protected with fine-grained access control policies. This unique combination of features makes Fedora an attractive solution in a variety of domains. Some examples of applications that are built upon Fedora include library collections management, multimedia authoring systems, archival repositories, institutional repositories, and digital libraries for education.

**Evaluation of open source software**

To effectively evaluate digital library softwares, a framework is necessary to guide the planning,
controlling, and reporting of the evaluation. Common elements among software packages need to be examined so that suitable conclusions can be drawn. To accomplish this objective, evaluation instruments are needed, of which, several types are available including (Punter 1997):

1) static analysis of code: for structural measurement or anomaly checking;
2) dynamic analysis of code: for test coverage or failure data;
3) reference tools that compare the software product;
4) reference statistical data; and
5) inspection with checklists.

In designing a digital library, there is no decision more important than the selection of quality software that forms the platform for service delivery. The variety of choices available makes the selection somewhat daunting. The key is a careful definition of the nature of information in the library and how it will be used. In the present work, a review of the digital library literature yielded the following five broad requirements that were used as our evaluation criteria (Dobson and Ernest 2000):

- **Content management:** This requirement is related to the ease with which content is created, submitted, reviewed, organized, and versioned within a digital library. It also encompasses the provision for searching and browsing functions such as metadata search, full-text search, and hierarchical subject browsing. Additionally, content encoded in a variety of popular formats, including text (ASCII, UNICODE, RTF), image (TIFF, GIF, JPEG), presentation (Adobe PostScript and Adobe PDF), structured formats (HTML and XML), audio and video (Real, MP3, AVI, and MPEG) ought to be supported.

- **User interface:** The user interface requirement covers the flexibility in customizing the interface to suit the needs of different digital library implementations as well as the support for multilingual access. With multilingual access, the user is able to specify the language for the digital library’s user interface and the cataloging information stored within it (Witten and Bainbridge 2002).

- **User administration:** This requirement is concerned with the range of functions needed to manage users of a digital library. For example, access to content in a digital library needs to be restricted through password authentication, IP filtering, and proxy filtering. Also, usage patterns have to be monitored and reported. When usage patterns are analysed, the needs and interests of digital library users can be better understood (Jones, Cunningham, and McNab 1998).

- **System administration:** This requirement is related to the back-end maintenance of a digital library. Automatic tools are useful, particularly for large digital libraries where maintenance work is labour-intensive (Cordeiro 2004). Functions such as automated content acquisition, harvesting, and automatic metadata generation (including named entity recognition and automatic subject indexing/classification) make digital library maintenance much easier. Moreover, a digital library needs to support preservation standards, as well as persistent document identification so that the transfer of digital materials from one hardware/software configuration to another would not compromise reference citations and other links (Hedstrom 2001).

- **Other requirements:** A digital library needs to be interoperable with other systems to which it is connected. This allows each system to evolve independently without sacrificing the ability to communicate with each other. At least two basic interoperability protocols should be supported, namely, Z39.50 and OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Additionally, the digital library must be compliant with standards established for digital library
collection and services. Examples of standards include XML for representation of information; XHTML for Web pages; GIF, TIFF, and JPEG for images; Unicode for multilingual support and information interchange; and Dublin Core and MARC 21 for metadata. Finally, a digital library needs to provide mechanisms through which digital library administrators and developers can obtain system support and help. Such mechanisms include documentation, manuals, mailing lists, discussion forums, bug tracking, feature request systems, and formal helpdesk support.

Digital library evaluation checklist
Due to the lack of a universally accepted definition of a digital library, there is no common methodology for the selection of a good digital library software. With this in mind, the present study aimed to develop a simple-to-use instrument for evaluating digital library software with the following characteristics:

- **Comprehensiveness** The evaluation criteria should cover all the key areas involved in selection of digital library software.
- **Useability** The instrument should be simple to understand, and more importantly, easy to employ by a variety of users with or without a background in software evaluation.
- **Flexibility** In choosing a digital library software, different stakeholders may place different emphasis on various criteria. For example, a library administrator may be more concerned with the ease of submission of material, whereas a researcher may be more interested in better search functionality. The weights associated with each criterion should be easily modifiable to reflect different stakeholder needs.
- **Expandability** The instrument should be easily expandable to include additional factors and criteria such as new and emerging standards.

Our digital library evaluation checklist consists of 12 categories of items, each with varying degrees of importance: content management, content acquisition, metadata, search, access control and security, report and inquiry, preservation, interoperability, user interface, standards compliance, automatic tools, and support (Goh D H, Chua A, Khoo D A, et al. 2006).

**Methodology for evaluation**
The method of assigning weights to evaluation criteria was adapted from Edmonds and Urban’s methodology, who recommended the use of the Delphi technique. In the original technique, a committee anonymously assigns weights to each criterion, usually through a questionnaire. It then reviews the results, and if there is no consensus, the steps are repeated until a consensus is reached. In the present study, we modified the Delphi technique by having a group of four people trained in information science and familiar with digital library concepts assign weights to each category and its respective items independently. The total sum of the category weights was 100, while the total sum of the items in each category was 10. Discrepancies were then resolved via face-to-face discussions, in which each person provided justifications for the reasons behind his/her decisions. Pairwise comparisons were also conducted as part of the process of formulating appropriate weights for the checklist. In pairwise comparison, the relative importance of each criterion against every other criterion is determined, often in a group session that is preceded by individual assessment (Koczkodaj, Herman, and Orlowski 1997).

Next, the four selected digital library software packages were evaluated using the checklist. Scores were computed by considering the software as a group, using the same four people that developed the checklist, but on a separate occasion. In cases where the evaluators disagreed over whether a particular criterion was met in a software package, a majority vote was taken. In case of a tie, a consensus was arrived through
emailing the digital library software developers or consulting other sources (Goh D H, Chua A, Khoo D A, et al. 2006).

**Findings and analyses**

Table I shows the scores for each of the sub-categories of the four digital library softwares evaluated, while Table IA depicts the consolidated score of all the categories for the four digital library softwares.

**Content management**

This category involves procedures and tools pertaining to the submission of content into the digital library as well as management of the submission process. As shown in Table I (row 1), all digital library software, with the exception of Fedora, satisfied most—if not all—of the criteria. Fedora only managed a score of 4.50 out of 10. This comparatively poor performance is due mainly to a lack of submission support and review. Fedora only provides capabilities to insert content, but not features such as notification of submission status or allowing users to modify submitted content.

**Content acquisition**

Content acquisition refers to functions related to content import/export, versioning, and supported document formats. Table I (row 2) shows that all the selected digital libraries managed to fulfill this criterion. EPrints, in particular, achieved a full score of 10.

**Metadata**

As mentioned earlier, metadata support in digital libraries is vital when it comes to content

### Table I  Evaluation results for individual categories

<table>
<thead>
<tr>
<th>Sl No.</th>
<th>Category</th>
<th>Digital library software package (Sum of sub-category score)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>E-prints</td>
</tr>
<tr>
<td>1</td>
<td>Content management</td>
<td>8.00</td>
</tr>
<tr>
<td>2</td>
<td>Content acquisition</td>
<td>10.00*</td>
</tr>
<tr>
<td>3</td>
<td>Metadata</td>
<td>5.13</td>
</tr>
<tr>
<td>4</td>
<td>Search support</td>
<td>3.40</td>
</tr>
<tr>
<td>5</td>
<td>Access control and privacy</td>
<td>8.33*</td>
</tr>
<tr>
<td>6</td>
<td>Report and inquiry capabilities</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>Preservation</td>
<td>5.50</td>
</tr>
<tr>
<td>8</td>
<td>Interoperability</td>
<td>5.00</td>
</tr>
<tr>
<td>9</td>
<td>User interface</td>
<td>10.00*</td>
</tr>
<tr>
<td>10</td>
<td>Standards compliance</td>
<td>10.00*</td>
</tr>
<tr>
<td>11</td>
<td>Automatic tools</td>
<td>5.00</td>
</tr>
<tr>
<td>12</td>
<td>Support and maintenance</td>
<td>6.00</td>
</tr>
</tbody>
</table>

Note: * Indicates the highest category score

Method for computing sub-category score:

Number of check marks × sub-category weight ÷ No. of items in the sub-category = Sub-category score

indexing, storage, access, and preservation. However, the performance in this area was disappointing. As shown in Table I (row 3), most of the digital libraries in our study only supported a few metadata standards. While it is encouraging that at least core standards like MARC21 and Dublin Core were supported, emerging metadata schemas such as EAD and METS were missing in all, except Fedora (Guenther and McCallum 2003).

**Search support**

Search support refers to a range of searching and browsing functions such as metadata search, full-text search, and hierarchical subject browsing. Considering the importance of searching functionality, Table I (row 4) shows that performance in this respect across the four digital libraries was varied, ranging from a low of 3.40 to a high of only 8.71. In particular, E-prints’ poor performance was due to the absence of full-text search support, as well as metadata search. In addition, none of the software exhibited proximity searching capabilities.

**Access control and privacy**

Access control and privacy include the administration of passwords, as well as the management of users’ accounts and rights to specified locations within the digital library. Most of the digital libraries surveyed scored well for this indicator (see Table I, row 5), with E-prints being the best overall performer.

<table>
<thead>
<tr>
<th>Sl No.</th>
<th>Category</th>
<th>E-prints</th>
<th>Fedora</th>
<th>Greenstone</th>
<th>DSpace</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Content management (7)</td>
<td>5.6</td>
<td>3.15</td>
<td>6.30</td>
<td>7.00</td>
</tr>
<tr>
<td>2</td>
<td>Content acquisition (11)</td>
<td>11.00</td>
<td>10.38</td>
<td>10.08</td>
<td>9.53</td>
</tr>
<tr>
<td>3</td>
<td>Metadata (12)</td>
<td>6.15</td>
<td>8.40</td>
<td>6.90</td>
<td>7.50</td>
</tr>
<tr>
<td>4</td>
<td>Search support (14)</td>
<td>4.76</td>
<td>6.94</td>
<td>10.15</td>
<td>12.19</td>
</tr>
<tr>
<td>5</td>
<td>Access control and privacy (7)</td>
<td>5.83</td>
<td>4.31</td>
<td>5.48</td>
<td>7.50</td>
</tr>
<tr>
<td>6</td>
<td>Report and inquiry capabilities (3)</td>
<td>0.00</td>
<td>1.80</td>
<td>3.00</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>Preservation (5)</td>
<td>2.75</td>
<td>5</td>
<td>1.25</td>
<td>3.75</td>
</tr>
<tr>
<td>8</td>
<td>Interoperability (10)</td>
<td>5.00</td>
<td>6.00</td>
<td>8.00</td>
<td>5.00</td>
</tr>
<tr>
<td>9</td>
<td>User interface (8)</td>
<td>8.00</td>
<td>8.00</td>
<td>8.00</td>
<td>8.00</td>
</tr>
<tr>
<td>10</td>
<td>Standards compliance (10)</td>
<td>10.00</td>
<td>10.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>Automatic tools (4)</td>
<td>2.00</td>
<td>2.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>12</td>
<td>Support and maintenance (9)</td>
<td>5.50</td>
<td>9.00</td>
<td>9.00</td>
<td>9.00</td>
</tr>
<tr>
<td></td>
<td>Consolidated score</td>
<td>66.49</td>
<td>74.98</td>
<td>73.16</td>
<td>72.33</td>
</tr>
</tbody>
</table>

Method for computing category score:

\[
\text{Category score} = \frac{(\text{Sum of sub-category score}) \times (\text{Category weights})}{10}
\]

For example, Category score of ‘content management’ category of E-prints software is:

\[
\frac{(8 \times 7)}{10} = 5.6
\]
administration, having not only system-assigned passwords, but also the ability to select passwords and retrieve forgotten passwords. Fedora scored well on access management in its support for IP address filtering, proxy filtering, and credential-based access.

Report and inquiry capabilities
This category deals with usage monitoring and reporting. Table I (row 6) shows that Greenstone was the only software that fulfilled all the requirements in this category. While Fedora provided usage statistics, it did not offer report generation tools. Both E-prints and DSpace lacked report and inquiry capabilities.

Preservation
It refers to preservation of metadata and quality control measures to ensure integrity, and persistent documentation identification for migration purposes (Hedstrom 2001). Fedora was a clear winner in this regard, with its support for CNRI handles, digital libraries, quality control, and provision for a prescribed digital preservation strategy (see Table I, row 7).

Interoperability
Interoperability is concerned with the benefits of integrating distributed collections and systems. Our study revealed that Greenstone was the best performer (see Table I, row 8) in this respect. All the software surveyed supported OAI-PMH. However, Z39.50 was only supported by Greenstone. This may be due to the protocol being much more complex to implement.

User interface
This category deals with support for multilingual access, as well as the ability to customize the user interface to suit the needs of different digital library implementations. All the four digital library software surveyed obtained a full score (Table I, row 9), reflecting that these issues have been taken into consideration.

Standards compliance
Standards are important for the sharing of digital content and long-term digital preservation (Dawson 2004). Thus, this category was evaluated by looking for evidence of the usage of standards. As the only other category with a full score across all software (see Table I, row 10), there appears to be a demonstrated commitment to the use of standards. It should be noted, however, that such a commitment does not imply every conceivable standard should be adopted. The other evaluation categories should be consulted to determine which specific standards are supported by each digital library. For example, while most document and image format standards are supported by the four digital libraries, not all metadata formats are, with Dublin Core being the only one supported by all.

Automatic tools
This category refers to tools for automated content acquisition, harvesting, and metadata generation. In the context of digital libraries, automatic tools are useful for maintenance and can reduce labour costs, especially for large collections. Table I (row 11) shows that Greenstone and DSpace came up with full scores, while Fedora and E-prints did not fare that well.

Support and maintenance
Support and maintenance are important for all software systems. But open source software is often criticized to be lacking in these aspects. However, our results show that three out of the four digital libraries evaluated performed well in this category (see Table I, row 12) by offering documentation, manuals, mailing lists, discussion forums, bug tracking, feature request systems, and formal helpdesk support. Only E-prints fared relatively poorly, due to its lack of formal helpdesk support and documentation that was not updated.
Discussion

Consolidated scores

Figure 1 shows the consolidated scores of the four digital library softwares that were evaluated. Fedora emerged as the best performer (with a consolidated score of 74.98), followed closely by Greenstone (with a score of 73.16). These two were followed by DSpace and E-prints with scores of 72.33 and 66.49, respectively. It should be noted that the consolidated scores were obtained by summing all category scores after normalization by their respective category weights.

Fedora was the only software package that consistently fulfilled the majority of the criteria in many categories, and obtained maximum scores in five of the 12 categories. These five indicators were content acquisition, metadata, standards compliance, user interface, and support and maintenance. In fact, Fedora secured full scores in four out of the 12 categories. Fedora’s key strength is its support for preservation and standards, in which full scores were obtained. It also ranked highest in the metadata category due to its support for many metadata standards. Other than the lack of Z39.50 support, Fedora appears to be a good candidate as far as long-term digital preservation needs are concerned. Fedora is also easily installed on a Windows 2003 server machine, although more configuration work is required as compared to Greenstone. However, Fedora has limited support for automated tools and content management features.

Greenstone places great emphasis on end-user functionality. For example, usage reports and statistics help a library administrator determine bottlenecks and identify popular files accessed. User interface customizability allows different digital libraries to create interfaces that suit the needs of their stakeholders, while automatic tools simplify content management and acquisition. In addition, Greenstone attained nearly perfect scores in content management and acquisition, implying that it considerably helps ease the task of managing content in a digital library.

Due to the ease in its installation, Greenstone is close to Fedora. Packaged in a single executable installation, the digital library becomes operational on a Windows 2003 server machine in less than an hour. Documentation for Greenstone is also extensive. There is a wealth of online documents and tutorials available on the Greenstone Website, and a number of organizations even offer training courses. In a nutshell, we believe that Greenstone is the most user-friendly software for creating digital libraries among the four softwares evaluated.

Similar to Greenstone, DSpace also secured maximum scores in five of the 12 categories. Although it was close to Greenstone in total score, DSpace performed slightly worse because of its lack of report and inquiry features. An issue with DSpace not reflected in the checklist was the difficulty in installing the software. As compared to the smooth installation of Greenstone, DSpace took time on a newly set up Linux machine, due to the number of other required software packages that needed to be installed and properly configured, and the extensive knowledge of Linux that was required.

DSpace is the most popular among the digital library solutions available in the open source domain. E-prints is also widely used. Educational institutions dominate in the use of these packages. Though many institutions have implemented digital libraries, only about
half of these are available online. Open access of knowledge is possible only if these repositories are made online. India is benefiting a lot from the open source movement (Jose 2007).

E-prints was the worst performer with a total score of 66.49. Its advantage is that the software, in the study, was the only one to obtain a full score in content acquisition, and that it supports the widest range of document formats. On the other hand, E-prints lacks in usage reporting and inquiry capabilities. It is also only available on the Linux platform, and therefore shares the same installation problems faced with Fedora. However, its greatest weakness is the low score (3.40) under the search category. Only metadata searching is supported and full-text search is not available in E-prints.

**Conclusion**

Although the checklist developed in the current study aims to be comprehensive, it is just the first step in the development of an exhaustive evaluation tool. The current version has some limitations in the assessment of digital library software. For example, the checklist does not take into account factors such as hardware, time, manpower, money, and other resources, as these may vary depending on the implementing organization or individual. The availability of application programming interfaces for implementing new features was also not considered. In addition, the weights given to the various categories were assigned through a consensus process among four evaluators. Other stakeholders with different viewpoints and needs may require different weightage schemes. However, as discussed previously, the checklist is flexible enough to accommodate new categories, items, and weightage schemes.

Extensive research was conducted to extract requirements for digital libraries, which led to the definition of criteria for digital library evaluation, and from these, a checklist was created. Assigning scores to each digital library software package against our checklist further reinforced the differences among these digital libraries in accommodating diverse needs. From the results of the evaluation, we have come to the conclusion that open source digital library softwares currently available still lack certain functionalities perceived to be important. However, among our four candidate digital libraries, Greenstone was able to fulfill most of the crucial requirements because of its strong support in end-user functionality. However, it must be noted that each software package has individual strengths and weaknesses that will appeal to different organizations and stakeholders with differing needs. Therefore, those interested in implementing a digital library can use the checklist to evaluate how well a particular software suits their specific implementation requirements (Goh D H, Chua A, Khoo D A, et al 2006).

Each of the above-mentioned software systems is designed to meet the original requirements of developing a digital library. DSpace supports community-based content policies and submission process, and accommodates various kinds of digital document formats. E-prints is a useful digital library system with a considerable user community. But when there is a need for technical support and training in using the software, Greenstone was found to be the most suitable. Though many libraries in India are using Greenstone and DSpace, some are also using E-prints because of its immense potential and the fact that it can support numerous forms and formats (Laxminarsaiah and Rajgoli 2005).

The Open Archives Initiative (OAI) has gained momentum since 2000 when eprints.org was launched OSS incorporates an interface that makes it easy for people to create their own library. Collections may be built and served locally from the user’s own Web server or remotely on a shared digital library host. End users can easily build new collections styled after existing ones from material on the Web or from their local files (or both), while collections can be updated and new ones brought online at any
time. OSS has a lot of potential for libraries and information centres, and there are a number of projects—including Greenstone, DSpace, and E-prints—that demonstrate its viability in this context. It provides library staff an option to be actively involved in development projects. This involvement can take many forms, such as reporting bugs, suggesting enhancements, and testing new versions. Currently available OSS projects cover application areas ranging from the traditional library management systems to innovations like Greenstone and DSpace, which complement traditional systems. These concepts and their benefits and importance to libraries are being examined. Benefits include lower costs, greater accessibility, and better prospects for long-term preservation of scholarly works (Naik 2006).

Traditional libraries are limited by storage space, while digital libraries/repositories have the potential to store massive amounts of scholarly information and requiring very little space in the process. Moreover, the cost of maintaining an institutional digital library is much lower than that of its traditional counterpart. Digital libraries can adopt innovations in electronic and audio-video book technology. Considering the advantages like absence of any physical boundary, round-the-clock availability of information, multiple access to information resources, faster information search and retrieval, preservation and conservation of exact copy of the original documents, and low costs, various organizations have opted for digital library using open source softwares. With the development of a digital library, information access and retrieval has been increased at campus levels and also at the global level. The aim is to deliver the right information to the right reader at the right time.

References
Caton M. 2004. Sugar sale: just the SFA basics. eWeek 21(50): 48

Cordeiro M. 2004. From rescue to long-term maintenance: preservation as a core function in the management of digital assets. VINE 34(1): 6–16


TerraGreen promotes the concept of sustainable development. Launched in June 2004, this magazine from TERI is an effort to bring forth information and knowledge in the fields of energy, environment, and sustainable development. The magazine is in keeping with our mission to expand the base of environmentally conscious readers and popularize sustainability issues at the local level. TerraGreen aims to provide the readers with the necessary inputs to enable them to be a part of the process of change. The magazine stays away from all jargon, so that the educated, informed, yet lay readers are updated on all that happens around them everyday.

**SUBSCRIPTION FORM**

<table>
<thead>
<tr>
<th>Term (yrs)</th>
<th>No. of issues</th>
<th>Cover price</th>
<th>You pay</th>
<th>You save</th>
<th>% saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>₹ 400</td>
<td>₹ 400</td>
<td>₹ 80</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>₹ 600</td>
<td>₹ 600</td>
<td>₹ 120</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>₹ 900</td>
<td>₹ 900</td>
<td>₹ 180</td>
<td>20</td>
</tr>
</tbody>
</table>

*Free online access to those subscribing for 3 years.

Customer Code: 
(In case of renewal)

Name of the Customer (IN BLOCK LETTERS):

Designation: 
Company / Organization:

City: 
State: 
PIN: 
Country:

Email:

IF YOU WANT TO GIFT A SUBSCRIPTION:

Name: 
Contact No: 

Address: 
City: 
State: 
PIN: 
Country:

Contact No: 
Email:

PAYMENT PROCEDURE:

Enclosed a Cheque/Draft number: 
Drawn in favour of “TERI” payable at New Delhi for ₹.

For 1/2/3 year subscription of TerraGreen.

Buy online at [http://bookstore.terlin.org](http://bookstore.terlin.org)

FOR SUBSCRIPTION QUERIES CONTACT Kakali Ghosh, Asst. Executive (Marketing); e-mail: kakali@terlin.in

THE ENERGY AND RESOURCES INSTITUTE
Darbari Seth Block, IHC Complex, Lodhi Road, New Delhi – 110 003
Tel. 2468 2100 or 4130 4900, Fax 2468 2144 or 2468 2145, India +91 • Delhi (0) 11
www.terragreen.terlin.org
Vision 2021 and public libraries: an action plan to allow Bangladesh to go digital

Md Nasiruddin, Associate Professor, Information Science and Library Management, University of Rajshahi, Bangladesh
mitulnasiruddin@gmail.com

Abstract
In Bangladesh, the present government has clearly outlined in its election manifesto the following four strategies to help the country go digital by 2021: (i) developing human resources that are ready for the 21st century; (ii) connecting citizens in ways that are most meaningful to them; (iii) taking services to citizens’ doorsteps; and (iv) making the private sector and the market more productive and competitive through digitization (Bangladesh Awami League 2009). To attain the objective, the government of the country has given the highest priority to science, technology, and management education. Public libraries play an important role in establishing a knowledge-based society by improving the quality of education. Unfortunately, the digitization process is spreading slowly in most public libraries, which still have a long way to go before they can be termed ‘digital’ in the real sense of the term. As of today, Bangladesh has 68 government and about 865 non-government public libraries (NBC 2008), which are unable to fulfill the functions that a public library should ideally perform.

So, the question arises, why public libraries in Bangladesh are not able to perform their noble roles? The main objective of the study is to conduct a survey to assess the digital initiatives of various public libraries in Bangladesh. It will also attempt to ascertain the cause of decreasing number of users of public libraries. The findings of the survey will encourage the government and other relevant policy-makers in Bangladesh to formulate a strategic plan for digitization of public libraries in the country in a cohesive manner.
Introduction

In Bangladesh, the present government has already fixed its target of achieving ‘Digital Bangladesh’ by 2021. The idea sounds good, but the concept of ‘Digital Bangladesh’ is still not clear to the common man (Siddique 2009). The term ‘Digital Bangladesh’ implies that the government wants to transform the nation into a fully digitized one by 2021 through application of third-generation information and communication technology (ICT). In the education sector, ‘Digital Bangladesh’ visualizes that by 2021 all universities, colleges, high schools, and primary schools will be computerized.

ICT is intended to be used as teaching-learning aid. For instance, it is intended that after five years of schooling, all students should have regular access to computers with internet facilities. The overarching goal is to improve the quality of education. It is common knowledge that public libraries play an important role in establishing a knowledge-based society. Ensuring quality education in a country is impossible without modernization of libraries. After 39 years of independence, in Bangladesh, development has occurred in almost all fields, but unfortunately, public libraries have been given little importance. The use of ICT in public libraries is spreading very slowly, and hence, these libraries still have a long way to go before they can call themselves ‘digital’ in the true sense. In fact, the government public libraries of Bangladesh are not even able to act the way public libraries should. As of today, there are 68 government and about 865 non-government public libraries in the country. But none of them has reached the level of development required to effectively support education, research, and socio-economic development. This study will examine the present situation of public library digitization in Bangladesh, and ascertain the cause behind the fall in the number of public-library users. It also suggests some guidelines for strategies that need to be undertaken to ensure that public libraries—not only in Bangladesh but in other developing countries as well—can continue to meet the challenges posed by rise in use of information technology and the globalization phenomenon.

Sample

The survey covered 20 out of the 68 district public libraries. The choice of libraries was made based on the availability of data and their location. The following libraries participated in the survey by completing and returning the survey questionnaire: Dhaka Central Public Library, Gazipur Public Library, Manikganj Public Library, Narayanganj Public Library, Chittagong Divisional Public Library, Cox’s Bazar District Public Library, Comilla District Public Library, Brammanbaria District Public Library, Rajshahi Divisional Public Library, Sirajganj District Public Library, Natore District Public Library, Bogra District Public Library, Khulna Divisional Public Library, Jessore District Public Library, Jhenaidah District Public Library, Magura District Public Library, Barisal Divisional Public Library, Patuakhali District Public Library, Sylhet Divisional Public Library, and Maoluvi Bazar District Public Library.

Methodology

Two sets of questionnaires (both structured and non-structured) seeking data on use of ICTs and user satisfaction were developed and distributed among the librarians and users of the selected public libraries. It was found that out of six, only four divisional libraries used free Web-based email services such as Hotmail or Yahoo. Others had no opportunity to access email. This being the case, a decision was made to visit the sample libraries with the questionnaires to conduct personal interviews. The distribution and collection of questionnaires through a direct participatory approach proved to be very reliable and fast but not a cost-effective one as every district had to be visited frequently. A
A questionnaire consisting of 15 questions was designed to elicit the opinions from selected librarians and users. A questionnaire on user satisfaction was distributed among 90 users in six libraries, out of which 60 users responded. The relevant secondary sources of the study up to December 2009 have also been reviewed and analysed. Provision was made in the study to gather information on various variables like background information on the library, year in which the first computer was used, access to computer facilities, hardware and software specifications, satisfaction level of the user regarding automation services, and so on.

**Background information on the sample libraries**

- **Begum Sufia Kamal National Public Library, Dhaka:** This library is situated at a central location, namely, 10, Kazi Nazrul Islam Avenue, Shahabag, in the capital city of Dhaka on three acres of land. The library is housed within a 62,300 square feet, three-storied building. There is one general reading room, one science reading room, one reference reading room, book stacks, one separate reading room for children and adolescents, one seminar room, and an administrative room. There are eight residents (Not clear what is meant by ‘residents’ here) for its emergency staffs, two electric substations (500 KVA), and a big 500-seater auditorium in the campus. The total collection of the library is 162,440; of which the number of books and periodicals in Bengali is 77,856, 79,508 in English, and 5,076 under ‘others’. The library subscribes to 19 Bengali and eight English newspapers daily. Each month, it subscribes to 26 local and four foreign journals for its users. The total manpower of the library is 48, out of which 20 are professional, 16 are semi-professional, and 12 are support staff. On an average, 1,851 users visit the library every day. It provides round-the-clock services for general readers, 12 hours for reference and recreational services, and eight hours for children and adolescent readers.

- **District Government Public Library, Gazipur:** This library was established on 29 July, 1986 as a non-government public library. A year later, the Bangladesh government undertook a project to expand the public library systems at the district level. As a part of this project, the government acquired the land and provided fund for the library’s development. Finally, on 25 October 1989, the library was inaugurated as a district government public library by government. The library building is a two-storied one, and covers 2,744 square feet. It is located at Stadium Road, Rothkholo, Gazipur. The total collection is 16,716; of which the number of books and periodicals in Bengali is 14,900; and 1,816 in English. It subscribes to 12 local journals and 11 newspapers daily. There are five staff members, who cater to an average of 315 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

- **District Government Public Library, Manikganj:** This library was established within a four-storied rented building near the stadium of Manikgonj in 1965. Its area was 1,275.37 square feet. It is located at Sadar Road, Manikgonj. The total collection is 14,316; of which the number of books and periodicals in Bengali is 13,276, and 1,040 in English. It subscribes to 18 local journals and three daily newspapers. There are four staff members, who provide services to an average of 180 users per day. Of the staff members, one is professional, two are semi-professional, and one is support staff.

- **District Government Public Library, Narayanganj:** It was established in 1965 in a rented building at 312, New Chasara, Jamtola, Narayanganj. The total area was 1,934.78 square feet. The total collection is 13,666; of which the number of books and periodicals...
in Bengali is 12,269, and 1,394 in English. It subscribes to 11 daily newspapers and 17 local journals. There are five staff members, who provide services to an average of 185 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

- **Divisional Government Public Library, Chittagong:** This library was established in 1963 under the Second Five Year Plan (1960–65). After the liberation of Bangladesh, the Department of Public Libraries was formed in 1983 based on the recommendations of the Enam Committee for Educational Reform. The Divisional Government Public Library, Chittagong is located at K C Dey Road. In order to expand the premises, a separate four-storied building with 46,000 square feet area was constructed between 1983 and 1989. The total collection is 65,868. The library subscribes to 24 local journals and 11 daily newspapers. There are 26 staff members, who provide services to an average of 1,066 users per day. Of the staff, four are professional, 14 are semi-professional, and eight are support staff.

- **District Government Public Library, Cox’s Bazar:** This library became functional on 2 September, 2000. It is located at the Circuit House Road of the central point in Cox’s Bazar district on 0.31 acres of land. It is a government building with a total space of 3,400 square feet. The total collection is 11,565, of which the number of books and periodicals in Bengali is 10,000, and 1,565 in English. It subscribes to 11 national daily newspapers and 18 local periodicals. There are five staff members, who provide services to an average of 200 users per day. Of the staff, four are professional, 14 are semi-professional, and eight are support staff.

- **District Government Public Library, Comilla:** Housed within a rented building, covering 1,640 square feet area, the library is providing services since September 1982. It is located at Sadar Road, Comilla. The total collection is 36,602; of which the number of books and periodicals in Bengali is 25,000 and 11,102 in English. It subscribes to 10 local periodicals and 11 local newspapers. There are five staff members, who provide services to an average of 529 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

- **Divisional Government Public Library, Brahmanbaria:** This library was established in 1964 in a rented three-storied, 1525 square feet building, at Sadar Road, Brahmanbaria. The total collection is 12,521; of which the number of books and periodicals in Bengali is 11,000 and 1,521 in English. It subscribes to 15 national daily newspapers and 10 local newspapers. There are five staff members, who provide services to an average of 200 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

- **District Government Public Library, Rajshahi:** Even though this library has been providing services since 1983, but it was formally inaugurated only on 8 December, 1993. It is located within a government building, and covers an area of 3.8 acres across 16,720 square feet. It is located at Medical College Road, Rajshahi. The total collection is 48,148; of which the number of books and periodicals in Bengali is 38,000, and 10,148 is in English. It subscribes to 21 national daily newspapers and 18 periodicals. There are 20 staff members, who provide services to an average of 529 users per day. Of the staff, four are professional, eight are semi-professional, and eight are support staff.

- **District Government Public Library, Sirajganj:** The library was established in 1984, and as of today it is situated within a rented building at central Sirajganj at Sadar Road. The total collection is 10,977; of which the number of books and periodicals in Bengali is 8,000, and 2,977 in English. It subscribes to none national daily newspapers and 15 local newspapers. There are five staff members, who
provide services to an average of 200 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

- **District Government Public Library, Natore:** This is also a rented building with 1139.47 square feet spaces, which became operational in 1998. It is located at Sadar Road, Natore. The total collection is 3,981, of which the number of books and periodicals in Bengali is 2,000, and the rest 1,981 in English. It subscribes to eight national daily newspapers and eight periodicals. There are five staff members, who provide services to an average of 200 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

- **District Government Public Library, Bogra:** Though the library was established in 1982 by local elites, the government acquired the land and renovated the library building in 2004. It is located at Thana Road, Bogra. The total collection is 24,596, of which the number of books and periodicals in Bengali is 20,400, and the rest 4,596 is in English. It subscribes to 11 periodicals and 12 national newspapers. There are five staff members, who provide services to an average of 200 users per day. Among them, one is professional, two are semi-professional, and one is support staff.

- **Divisional Government Public Library, Khulna:** The Divisional Government Public Library, Khulna is located at Boyra, within a two-storied building across 30,660 square feet. It was established on 25 June, 1964. The total collection is 87,080, of which the number of books and periodicals in Bengali is 80,000, and 7,080 in English. It subscribes to 24 local periodicals and eight local newspapers. There are 13 staff members, who provide services to an average of 1,048 users per day. Of the staff, two are professional, five are semi-professional, and six are support staff.

- **District Government Public Library, Jessore:** The library was established on 1 September, 1982. Before 1971, it was known as ‘Information Centre’ under the Ministry of Information and Radio. In 1982, the government of Bangladesh renovated the building as a public library and brought it under the Ministry of Cultural Affairs. The library is located at Gorib Sha Road, Jessore across 3,500 square feet. The total collection is 17,389, of which the number of books and periodicals in Bengali is 14,000, and the remaining 3,389 in English. It subscribes to 11 national daily newspapers and nine periodicals. There are four staff members, who provide services to an average of 200 users per day. Among them, one is professional, two are semi-professional, and one is support staff.

- **District Government Public Library, Jhenaidah:** The library was established in 1982 in a rented building with 3,411 square feet space. It is located at Thana Road, Jhenaidah. The total collection is 12,278, of which the number of books and periodicals in Bengali is 10,000 and 2,278 in English. It subscribes to seven national daily newspapers and six periodicals. There are four staff members, who provide services to an average of 200 users per day. Of the staff, one is professional, two are semi-professional, and one is support staff.

- **Divisional Government Public Library, Magura:** The library building was established in 1999 on 0.33 acres of land, and with a total area of 3,400 square feet. It is located at Sadar Road, Magura. The total collection is 16,439, of which the number of books and periodicals in Bengali is 13,000 and the remaining 3,439 in English. It subscribes to 11 national daily newspapers and 16 periodicals. There are four staff members, who provide services to an average of 200 users per day. Of the staff, one is professional, two are semi-professional, and one is support staff.

- **Divisional Government Public Library, Sylhet:** On 15 April, 2002, the then Finance Minister of Bangladesh declared that the government intended to establish a divisional
public library in Sylhet, and accordingly they had chosen the site near the stadium of the Sylhet city. In December 2005, on behalf of the government of Bangladesh, the Finance Minister inaugurated the four-storied building situated on 0.50 acres of land. The total collection is 41,124, of which the number of books and periodicals in Bengali is 31,000, and 11,124 in English. It subscribes to seven national daily newspapers and 15 periodicals. There are nine staff members, who provide services to an average of 1,171 users per day. Of the staff, two are professional, four are semi-professional, and three are support staff.

- **District Government Public Library, Maoluvi Bazar**: This library has been functioning since 1 September, 1982 in a building rented by the Maoluvi Bazar Municipality. The total collection is 11,827, of which the number of books and periodicals in Bengali is 9,000, and 2,827 in English. It subscribes to five national daily newspapers and nine periodicals. There are four staff members, who provide services to an average of 200 users per day. Of the staff, one is professional, two are semi-professional, and one is support staff.

- **Divisional Government Public Library, Barisal**: The library was established in 1965. With the upgradation of Barisal as a Division in 2002, it was converted into the Divisional Government Public Library on 16 May, 2005, and is located at College Road, Barisal. The total collection is 40,994, of which the number of books and periodicals in Bengali is 35,000 and the rest 5,994 in English. It subscribes to 15 national daily newspapers and nine periodicals. There are 11 staff members, who provide services to an average of 690 users per day. Of the staff, two are professional, five are semi-professional, and four are support staff.

- **District Government Public Library, Patuakhali**: The District Government Public Library, Patuakhali was established on 1 November, 1982. Subsequently, the government acquired 0.34 acres of land and renovated the building on 6 June, 2002 to ensure its proper functioning. The library is located at Circuit House Road, Patuakhali. The total collection is 22,967, of which the number of books and periodicals in Bengali is 18,000, and 4,967 in English. It subscribes to 11 periodicals and 14 newspapers. There are 5 staff members, who provide services to an average of 200 users per day. Of the staff, one is professional, three are semi-professional, and one is support staff.

**Findings of the study**

Librarians of the 20 sample libraries were asked to indicate the availability of hardware (servers, personal computers, laptops, and printers), software (operating systems and application software) and the year in which the first computer was used in the library. The analysis and interpretation of the same are presented below:

Out of the 20 libraries covered, the earliest use of computer was in 1994 at the Central Public Library. As shown in Table 1, three libraries started using computers in 1998, two in 2002, and three in the 2006. The remaining 11 still do not have access to computers.

Table 2 shows that majority of public libraries included in the survey have no access to computers, even personal computers. Only Dhaka Central Public Library has one server, but even that one is hardly functional because there is no LAN. The survey found that 25 computers were purchased under the Bangladesh National Scientific and Library Information Network (BANSLINK) project. BANSLINK was inaugurated on 29 June 1998, with the initial objective of having 15 nodes across the country in 2000 (Mannan 1990). However, due to various problems, both technical and lack of funds, the objective was not attained. Few other divisional public libraries do have a limited number of computers, but these are not capable of automating all the library services. One
Table 1  Year in which the first computer was installed

<table>
<thead>
<tr>
<th>Name of public library</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh Central Public Library, Dhaka</td>
<td>1994</td>
</tr>
<tr>
<td>District Government Public Library, Gazipur</td>
<td>NA</td>
</tr>
<tr>
<td>District Government Public Library, Manikganj</td>
<td>NA</td>
</tr>
<tr>
<td>District Government Public Library, Narayanganj</td>
<td>NA</td>
</tr>
<tr>
<td>Divisional Government Public Library, Chittagong</td>
<td>1998</td>
</tr>
<tr>
<td>District Government Public Library, Cox’s Bazar</td>
<td>NA</td>
</tr>
<tr>
<td>District Government Public Library, Comilla</td>
<td>2006</td>
</tr>
<tr>
<td>District Government Public Library, Brahmanbaria</td>
<td>NA</td>
</tr>
<tr>
<td>Divisional Government Public Library, Rajshahi</td>
<td>1998</td>
</tr>
<tr>
<td>District Government Public Library, Sirajganj</td>
<td>NA</td>
</tr>
<tr>
<td>District Government Public Library, Natore</td>
<td>2006</td>
</tr>
<tr>
<td>District Government Public Library, Bogra</td>
<td>2006</td>
</tr>
<tr>
<td>Divisional Government Public Library, Khulna</td>
<td>1998</td>
</tr>
<tr>
<td>District Government Public Library, Jessore</td>
<td>NA</td>
</tr>
<tr>
<td>District Government Public Library, Jhenaidah</td>
<td>NA</td>
</tr>
<tr>
<td>District Government Public Library, Magura</td>
<td>NA</td>
</tr>
<tr>
<td>Divisional Government Public Library, Barisal</td>
<td>2002</td>
</tr>
<tr>
<td>District Government Public Library, Patuakhali</td>
<td>NA</td>
</tr>
<tr>
<td>Divisional Government Public Library, Sylhet</td>
<td>2002</td>
</tr>
<tr>
<td>District Government Public Library, Maoluvi Bazar</td>
<td>NA</td>
</tr>
</tbody>
</table>

A major observation regarding access to computer facilities is that certain public libraries have very few computers, while most of the district public libraries have no computer facility.

**Hardware specifications**

Table 3 (below) shows that the situation of maximum district public libraries is woeful, as they have not started the use of computers in this digital age. Only six divisional library services and three district libraries out of 20 public libraries use a combination of Intel Pentium-based personal computers. These range from 486-based PCs to Pentium IV—in some cases with specifications of a 1.8 GHz processor, 128 MB RAM, 30 GB hard disk capacity, and a CD-ROM drive. Eleven libraries, namely, Gazipur Public Library, Manikganj Public Library, Narayanganj Public Library, Brahmanbaria Public Library, Sirajganj Public Library, Jessore Public Library, Jhenaidah Public Library, Magura Public Library, Patuakhali Public Library, and Maoluvi Bazar Public Library indicated that they have none of these options available, as they do not have access to computers.

The survey revealed that only divisional public libraries have access to personal computers, which were purchased for the purpose of...
<table>
<thead>
<tr>
<th>Public libraries</th>
<th>Servers</th>
<th>PCs</th>
<th>Laptops</th>
<th>Laser printer</th>
<th>Inkjet printer</th>
<th>Dot-matrix printer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh Central Public Library, Dhaka</td>
<td>1</td>
<td>30</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>District Government Public Library, Gazipur</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Manikganj</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Narayanganj</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Divisional Government Public Library, Chittagong</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>District Government Public Library, Cox’s Bazar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Comilla</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>District Government Public Library, Brahmanbaria</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Divisional Government Public Library, Rajshahi</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>District Government Public Library, Sirajganj</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Natore</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Bogra</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Divisional Government Public Library, Khulna</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>District Government Public Library, Jessore</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Jhenaidah</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Magura</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Divisional Government Public Library, Barisal</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>District Government Public Library, Patuakhali</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Divisional Government Public Library, Sylhet</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>District Government Public Library, Maoluvi Bazar</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 3: Type of processor used in personal computers

<table>
<thead>
<tr>
<th>Public libraries</th>
<th>Pentium IV</th>
<th>Pentium III</th>
<th>Pentium II</th>
<th>Pentium I</th>
<th>Pentium Celeron</th>
<th>Pentium Pro</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh Central Public Library, Dhaka</td>
<td>√</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Gazipur</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Manikganj</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Narayanganj</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Divisional Government Public Library, Chittagong</td>
<td>√</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Cox’s Bazar</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Comilla</td>
<td>X</td>
<td>X</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Brahmanbaria</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Divisional Government Public Library, Rajshahi</td>
<td>√</td>
<td>X</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Sirajganj</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Natore</td>
<td>X</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Bogra</td>
<td>X</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Divisional Government Public Library, Khulna</td>
<td>√</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Jessore</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Jhenaidah</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Magura</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Divisional Government Public Library, Barisal</td>
<td>√</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Patuakhali</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Divisional Government Public Library, Sylhet</td>
<td>√</td>
<td>√</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>District Government Public Library, Maolvi Bazar</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
digitization of library functions. But as there are no related infrastructural facilities available, full-fledged digitization services in the libraries could not be initiated. The availability of CD-ROM drives on most of the machines also implies that they could be used to provide access to various CD-ROM-based products, including educational resources such as multimedia encyclopaedias, databases, and full-text documents.

Software: operating systems
It should be noted from Table 4 that all the divisional public libraries indicated that they use Windows 2000. In early 2000, they used Windows 95. Microsoft Windows, especially Windows 95/98, is the most common operating system installed on PCs and laptops. Windows 95/98 was never really a network management system, and lacks the advanced functions found in systems like Windows 2000. However, given that purpose-built network software systems require skilled manpower to manage, and are also expensive in terms of user license fees, it is not surprising that these libraries have settled for Windows 2000. Most of the public library services cannot afford to employ skilled IT staff, and therefore, cannot even use free open-source UNIX systems like Linux or any integrated software (Khan 1993: 45) (Not mentioned under references). The popularity of Windows-based operating systems is mainly due to the fact that it is bundled with new Intel-based PCs, which was not the case in the past when it was necessary to purchase the operating system separately. Regarding the large number of machines with Windows 95/98, this can be attributed to the fact that most of these machines were purchased in 1998, when Windows 95/98 was the most commonly pre-installed operating system on PCs.

<table>
<thead>
<tr>
<th>Public libraries</th>
<th>Operating system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dhaka Central Public Library</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Chittagong Divisional Public Library</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Rajshahi Divisional Public Library</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Khulna Divisional Public Library</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Barisal Divisional Public Library</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Sylhet Divisional Public Library</td>
<td>Windows 2000</td>
</tr>
<tr>
<td>Comilla</td>
<td>— *</td>
</tr>
<tr>
<td>Natore</td>
<td>— *</td>
</tr>
<tr>
<td>Bogra</td>
<td>— *</td>
</tr>
</tbody>
</table>

*Comilla, Natore, and Barisal Public Libraries did not indicate which operating system was installed on their PCs.

Library automation scenario
Among the libraries that have automated some of their functions, the earliest instance was Dhaka Central Public Library in 1997. The following table depicts the current scenario of partially automated libraries in Bangladesh.

As can be seen from the table above, out of 20 libraries, only nine have access to computers, while the remaining 11 plan to have access to computers in future. Apart from Dhaka Central Public Library, no divisional or district public library has automated its major functions. The six divisional libraries are in the process of implementing digitization as per commitment of the present government to make Bangladesh go digital. Very soon, five divisional libraries, namely, Chittagong Divisional Public Library, Rajshahi Divisional Public Library, Khulna Divisional Public Library, Sylhet Divisional Public Library, and Barisal Divisional Public Library will receive funds from the government to automate some of their basic functions. While the aforementioned five libraries will be...
### Table 5: Library automation scenario

<table>
<thead>
<tr>
<th>Public libraries</th>
<th>Acquisitions and budgets</th>
<th>Cataloguing</th>
<th>Circulation</th>
<th>Serial control</th>
<th>OPAC</th>
<th>Management information</th>
<th>Inter-library loans</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh Central Public Library, Dhaka</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Gazipur</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>District Government Public Library, Manikganj</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>District Government Public Library, Narayanganj</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Divisional Government Public Library, Chittagong</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Cox's Bazar</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>District Government Public Library, Comilla</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Brahmanbaria</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Divisional Government Public Library, Rajshahi</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Sirajganj</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>District Government Public Library, Natore</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Bogra</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Divisional Government Public Library, Khulna</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Jessore</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>District Government Public Library, Jhenaidah</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>District Government Public Library, Magura</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Divisional Government Public Library, Barisal</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Patuakhali</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Divisional Government Public Library, Sylhet</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>District Government Public Library, Maoluvi Bazar</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
integrated with a central public library, other district public libraries like Gazipur Public Library, Manikganj Public Library, Narayanganj Public Library, Brahmanbaria Public Library, Sirajganj Public Library, Jessore Public Library, Jenaiah Public Library, Magura Public Library, Pathukhali Public Library, Maoluvi Bazar Public Library, and Cox’s Bazar Public Library will be covered in the next phase.

In addition to digitization of basic library functions—such as acquisitions, cataloguing, circulation, and access to OPAC—some library services are also using computers for related activities. For instance, at the Natore Public Library computers are being used for stock control and stock-taking; for accounting, community information, and book selection at the Comilla Library; and book reservations and access to MS Access database at Bogra District Public Library. The application of computers to library-related activities in some of these institutions is very limited, and often on a very small scale. For example, Bogra Public Library has two small catalogues (one is for collections from 2007–09 and the other is on older collection of books) compiled on computers using WordPerfect 5.1. Chittagong Divisional Public Library has developed three databases of part of the reference section of the main library. Rajshahi Divisional Public Library is developing a database on periodicals. Khulna Divisional Public Library is using computers for cataloguing and management information. Barisal Public Library Service is using computers for the maintenance of accounts, library statistics, and bibliographic control; while Sylhet Library is using computers for cataloguing.

Those libraries, which have not yet automated any of their functions, and those using computers on only a small scale, have plans to automate in future, with the exception of district public libraries at Barisal, which does not have an immediate plan to do so. The reasons are: their existing manual systems are adequate and digitization would affect no significant service improvement to users; they suffer from unreliable power supply that could cripple the library’s operations; and that no funds are available to meet the high cost of initial software installation and future updates. The major reason given for automating functions only at the main or central library was a lack of funds to extend digitization to all the libraries within the system. Other reasons provided were lack of ICT experts and inadequate supporting infrastructure. For example, absence of continuous electricity and telephone facilities where district libraries or service points are located.

**Satisfaction level of the user on the digitization services**

It is apparent from Table 6 that out of 50 respondents, majority of the users (90%) are not satisfied with acquisition and budget control services; while 86% are unhappy with cataloguing services, as most of these are manual services. Circulation, serial control, OPAC, management information systems, and even inter-library loan services of these libraries are not automated to a minimum satisfactory level. Some of the divisional public libraries use tailormade database through MS-Access or Excel for conducting routine jobs such as inter-library loan service, budget control, inventory control, staff information, and so on.

**Suggestions by users for improvement in public library digitization**

Success of any library depends a lot on its users, particularly scholars who are involved in research. This is because they have the better grasp of the knowledge and information related to their respective fields. Their wisdom can play a vital role in the improvement of library services, including its digitization. With this objective in mind, the users of public libraries included in the study were asked to offer suggestions. Out of 60 respondents, 10 did not offer any suggestions. From Table 7, it can be inferred that altogether 20 popular suggestions have emerged.
### Table 6 User satisfaction on library digitization

<table>
<thead>
<tr>
<th>Automated services</th>
<th>Satisfied</th>
<th>Not satisfied</th>
<th>Responded</th>
<th>Not responded</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisitions and budgets</td>
<td>5 (10.00%)</td>
<td>45 (90.00%)</td>
<td>50</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>Cataloguing</td>
<td>7 (14.00%)</td>
<td>43 (86.00%)</td>
<td>50</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>Circulation</td>
<td>3 (7.50%)</td>
<td>37 (92.50%)</td>
<td>40</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>Serial control</td>
<td>1 (2.78%)</td>
<td>35 (97.22%)</td>
<td>36</td>
<td>24</td>
<td>60</td>
</tr>
<tr>
<td>OPAC</td>
<td>1 (2.04%)</td>
<td>48 (97.96%)</td>
<td>49</td>
<td>11</td>
<td>60</td>
</tr>
<tr>
<td>Management information</td>
<td>2 (4.88%)</td>
<td>39 (95.12%)</td>
<td>41</td>
<td>19</td>
<td>60</td>
</tr>
<tr>
<td>Inter-library loans</td>
<td>3 (6.00%)</td>
<td>47 (94.00%)</td>
<td>50</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>Others</td>
<td>2 (4.35%)</td>
<td>44 (95.65%)</td>
<td>46</td>
<td>14</td>
<td>60</td>
</tr>
</tbody>
</table>

### Table 7 Suggestions by users for improvement in public library digitization

<table>
<thead>
<tr>
<th>Keywords of the suggestion offered</th>
<th>Size of sample</th>
<th>No. of respondents</th>
<th>No. of non-response</th>
<th>No. of users offered</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full digitization</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>7</td>
<td>14.00</td>
</tr>
<tr>
<td>Introducing LAN</td>
<td>60</td>
<td>45</td>
<td>15</td>
<td>5</td>
<td>11.11</td>
</tr>
<tr>
<td>Provision of online jobs</td>
<td>60</td>
<td>30</td>
<td>30</td>
<td>1</td>
<td>3.33</td>
</tr>
<tr>
<td>User-friendly service</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>1</td>
<td>2.00</td>
</tr>
<tr>
<td>Provision of more computers</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>2</td>
<td>4.00</td>
</tr>
<tr>
<td>Barcoding of library materials</td>
<td>60</td>
<td>25</td>
<td>35</td>
<td>1</td>
<td>4.00</td>
</tr>
<tr>
<td>Online information on new arrivals</td>
<td>60</td>
<td>30</td>
<td>30</td>
<td>1</td>
<td>3.33</td>
</tr>
<tr>
<td>Access through websites</td>
<td>60</td>
<td>45</td>
<td>15</td>
<td>4</td>
<td>8.89</td>
</tr>
<tr>
<td>Online catalogue service</td>
<td>60</td>
<td>35</td>
<td>25</td>
<td>1</td>
<td>2.86</td>
</tr>
<tr>
<td>Good software</td>
<td>60</td>
<td>40</td>
<td>20</td>
<td>1</td>
<td>2.50</td>
</tr>
<tr>
<td>Circulation control</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>1</td>
<td>2.00</td>
</tr>
<tr>
<td>Users training</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>1</td>
<td>2.00</td>
</tr>
<tr>
<td>Increase in trained manpower</td>
<td>60</td>
<td>52</td>
<td>8</td>
<td>8</td>
<td>15.38</td>
</tr>
<tr>
<td>Provision of more funds</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>6</td>
<td>12.00</td>
</tr>
<tr>
<td>Increase in internet facility</td>
<td>60</td>
<td>47</td>
<td>13</td>
<td>4</td>
<td>8.51</td>
</tr>
<tr>
<td>Introducing e-libraries</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>1</td>
<td>2.00</td>
</tr>
<tr>
<td>Network with other libraries</td>
<td>60</td>
<td>45</td>
<td>15</td>
<td>3</td>
<td>6.67</td>
</tr>
<tr>
<td>Provision of more space</td>
<td>60</td>
<td>55</td>
<td>5</td>
<td>2</td>
<td>3.64</td>
</tr>
<tr>
<td>Acquisition control</td>
<td>60</td>
<td>50</td>
<td>10</td>
<td>1</td>
<td>2.00</td>
</tr>
<tr>
<td>Collection development</td>
<td>60</td>
<td>53</td>
<td>7</td>
<td>2</td>
<td>3.77</td>
</tr>
</tbody>
</table>
Out of 52 respondents, eight (15.38%) have suggested ‘increase in trained manpower’. So, it is evident that to improve library digitization, first and foremost, trained manpower is needed. The next suggestion is ‘full digitization’, which was made by seven respondents (14.00%) out of 50, who want their libraries to be fully automated. It is a healthy sign, which acts as a motivating factor for libraries to take up computerization for better and effective use of library resources. In order to get the libraries computerized, a considerable amount of fund is required at the primary stage; otherwise the project cannot succeed. The suggestion ‘provision of more funds’ for library computerization came as third choice, with six suggestions (12%) out of 50. It is followed by suggestions like (a) introduction of LAN, (b) provision of online journals, (c) provision of more computer facilities in the library, (d) use of good software, and (e) increase in internet facility.

**Major findings**

Use of commercially available integrated library management systems is absent in public libraries of Bangladesh. Some public libraries in the country use free CDS/ISIS bibliographic database management software available from UNESCO. The deployment of automation services in a couple of divisional public libraries is being done on an ad hoc basis, since these libraries do not have the requisite automation strategies in place. Lack of budget is the major barrier towards digitization of public libraries. In most cases, funding for the acquisition of digitization services is not provided by the government, unless the library in question is a central public library. Public libraries plan to establish internet connectivity. Some divisional public libraries are providing computers for internet access by library staffs, while others have been trying to develop their respective library websites. However, the provision of library and information services via the Web has not yet taken place. Central public libraries also plan to set up a cyber café as a way of providing access to digitization facilities to their communities, and also as means to raise funds to sustain internet connectivity.

**Suggestions**

The study has also revealed that due to lack of adequate financial support, physical facilities and skilled manpower, public libraries of the country are not functioning properly. This reflects the absence of an organized public library and information system in the country, and the slow progress of resource-sharing activities. There is unanimous agreement on the importance and priorities of public library digitization. However, there exist various constraints in implementing any library digitization plan in the country. These include lack of clear conception, sufficient budget, willingness of the authorities concerned, qualified and trained staff, commitment of the library professionals, modern technological facilities, adequate information resources, physical and infrastructural facilities of the libraries and library standardization. In spite of these unfavourable conditions, a start has to be made to establish real and active online public library and information network in the country. It needs concerted as well as integrated efforts of administrators, politicians, academicians, researchers, and information professionals to ensure that information resources of one organization is accessible and made available to others who need them. After all, the ultimate aim of public library digitization is to provide relevant and authentic information expeditiously to the user as and when required.

**Conclusion**

The tremendous growth of information and communication technology has opened the door for libraries to make use of technological facilities for the betterment of user services. But, even today both government and non-government public libraries are unable to utilize information technology to its full
extent. For this reason, users are not satisfied because they need up-to-the-second and instant information. The public library system in Bangladesh still present a sorry picture. As the profession of a librarian is now attracting people from different social backgrounds, it is expected that over the years, this predominantly young group of professionals will be able to organize and digitize the entire public library systems of Bangladesh.

Public library digitization in Bangladesh is essential because the process can avoid duplication in collection building. It will also allow sharing of available resources effectively, and will ensure maximum utilization of information resources while fostering information exchanges, and in developing nationwide centralized database on different fields. It also saves labour, time, and money on the part of the information user and ultimately contributes to national development. Sustainability is more important than starting. If we fail to manage sustainable digital public libraries with our own resources, the dream of making Bangladesh go digital by 2021 will be harmful rather than beneficial for the country.

References


Geoinformatics for Climate Change Studies

P K Joshi • T P Singh

August 2011 • ISBN 9788179934098
• 492 pages • Size 160 mm x 240 mm
• Hardback • ₹995

Climate change is increasingly being considered a critical topic in research and policy-making. Evidences related to climate change deal with spatial and non-spatial data, which can be utilized for policy formulation. Geoinformatics, which includes remote sensing, GIS, GPS, and ICT, provides the most relevant technology to monitor climate change-related variables at different dimensions and scales.

Geoinformatics for Climate Change Studies discusses the art of using this technology for investigating, monitoring, documenting, and understanding the impacts of climate change. This book provides information on the concepts and uses of geoinformatics, and focuses on filling the gap in the available literature on the subject by bringing together concepts, theories, and experiences of experts in this field.

Key features
- Contains comprehensive information on the latest developments in geoinformatics for climate change studies.
- Chapters range over a wide variety of application areas, including forestry, glaciers, sea level rise, and agriculture.
- Highlights the potential usage of geoinformatics in understanding climatic changes.
- A conceptual guide and ready reference for students, researchers, and teachers working in the field of geoinformatics.

Table of contents
- A GIS-based Framework for Modelling and Global Design of Earth Systems
- Geoinformatics for Climate Change Research
- Reactions of Mountain Glaciers to Climate Change—A Remote Sensing Approach
- Geoinformatics and the Mapping of Lands Vulnerable to Sea Level Rise
- Geospatial Tools to Assess Forest Ecosystems under Climate Change Trajectories
- Geoinformatics for Comprehensive Impact Assessment and Analysis of Climate Change for Integrated Water Resources Management
- Climate—Population—Energy Scenarios for 2050
- Geoprocessing for Soft Mapping of Sparse and Inaccurate Evapotranspiration Data
- Impact of Climate Variability on Human Health
- Geo-information-based Approach for Monitoring
- Climate-induced Land Degradation in Nigeria
- Assessing Land Surface Temperature over Bangladesh using MODIS Satellite Images—An Indicator of Climate Change
- High Resolution Climate Change Scenarios for Morocco for the 21st Century
- GIScience Tools for Climatic Change Related Natural Hazards and Modelling
- Spatial Data Infrastructure Convergence: Building Spatial Data Infrastructure Bridges to Address Climate Change
- Geoinformatics for Climate Change Adaptation and Disaster Risk Reduction
- Geoinformatics and Communication Technologies for Climate Change Hazards: Musing beyond the Technical Issues

The Energy and Resources Institute
TERI Press
Darbari Seth Block
IHC Complex, Lodhi Road
New Delhi – 110 003, India

Tel.: 2468 2100 or 4150 4900
Fax: 2468 2144 or 2468 2145
India +91 • Delhi (0)11
Email: teripress@teri.res.in
Web: http://bookstore.teriin.org

To purchase the book, visit our online bookstore at http://bookstore.teriin.org or send us your demand draft or cheque in favour of TERI, payable at New Delhi (outstation cheques are not accepted).
Education for digital libraries: library management perspective

Elena Macevičiūtė, Swedish School of Library and Information Science (SSLIS), University of Borås, SE-50190 Sweden
Faculty of Communication, Vilnius University, Universiteto 3, Vilnius, Lithuania
elena.maceviciute@gmail.se

Abstract
The paper examines three Master’s programmes in digital libraries: Digital Library Learning (Oslo University College, Tallinn University, and University of Parma), Digital Library and Information Services, and Digital services – culture, information and communication (Swedish School of Library and Information Science). The author uses her direct experience of developing one of the programmes, collaborating with another and evaluating the third, for comparison of their conceptual basis, goals, curricula, students’ admission and study process, teachers’ experience, and several aspects of the students’ satisfaction.

The author concludes that all three programmes build their programmes on the notion of change management, either through innovation, organizational change or project implementation. This perspective cultivated in LIS management for several decades since the introduction of computer technology into libraries unites all three programmes on the common foundation of LIS management.

Key words
Digital library, LIS education, information management, digital resource management, Europe
Introduction

Today we have a recognizable situation with digital librarianship: digital libraries have arrived without prior announcement and librarians have started working with digital library resources, creating digital library services, using Web 2.0 tools for interacting with their users before the LIS education started reacting to it. A similar situation was witnessed with the integrated library systems or electronic resources. As a previous student and a lecturer at the LIS department of a university I could witness the efforts to forecast the future at least in the institutions I was involved with. As a (Soviet) Lithuanian student at the beginning of 1970s, I was studying literature on MARC and electronic catalogues; later (in the middle of 1990s) I enthusiastically preached information service delivered to customers through any channel, thereby, encouraging the students to use telephone or emerging e-mail. However, the same does not help much when reality fails to provide opportunities to apply what students call “theoretical knowledge”, meaning something that does not exist in the workplace. And in a second, the scene changes: the future has arrived and requires more and different competencies instead of what was earlier anticipated. This is a constant catch-22 situation for many educational areas, not just the LIS. Nevertheless, the educators continue to have a vision for the future and try to develop flexibility and thirst for permanent learning among their students using the means available today.

The ICDL conference in India is one such conference in the world that is focussed on the development and research in digital libraries. Most of these conferences also include papers on digital library education, though it is usually a small proportion of the total content. Nevertheless, the importance of the topic is rarely questioned. More attention is paid to the place of digital library education in LIS and computer science (Coleman 2002; Meera 2010), to the basic elements of the digital library curriculum and the skills that it should be developed (Abrizah et al. 2009), or to finding general approaches for teaching digital librarianship (in the panel on digital library education in European Conference on Digital Libraries – ECDL2009).

Recently, there has been a growing interest in digital library education in Asian and African countries (Boamah 2009, Bakeri Abu Bakar 2009). A recent article published stated: “Only the Department of Computer Science at the National Chiao tung University in Taiwan offers a Master Programme of Digital Libraries in the world” (Bakeri Abu Bakar 2009: 458). However, this is not exactly the case. It is true that full Master’s programmes directed only towards digital libraries are very rare, though courses on various aspects of digital libraries proliferate in LIS and computer science departments. But, there are quite a number of these programmes that are already available for the students. I myself have direct experience of three Master’s programmes in digital librarianship in Europe delivered in Library and Information Science departments. They are listed here in chronological order of the start of the programme:

- International Master’s programme: Digital library learning (further referred to as DLL): Programme started in autumn 2007 (ERASMUS MUNDUS joint degree programme by Oslo University College, Tallinn University and Parma University).
- Digital libraries and information services (further DLIS): first experimental intake of students in autumn 2008, official start of the programme – autumn 2009. (Swedish School of Library and Information Science)
- Digital services – culture, information and communication (further DiSe): programme started in autumn 2009 (Swedish School of Library and Information Science)

The three programmes are still at the nascent stage, but these have been advertised, and have been recruiting students for over four years.
now. The goal of this paper is to present these programmes to a wider audience. Concentrating on the curriculum structures and the common concept underlying them, I would introduce the experiences of the students and teachers and highlight the strengths of digital library education as well as difficulties met on the road.

**Contemporary higher education for digital librarianship**

The surveys of digital library education were conducted by Spink and Cool (1999), Saracevic and Dalbello (2001), and Liu (2004). These authors formulated approximately the same research questions and used the same methods to collect the data. In fact, the two later investigations may be treated as the continuation of the first one, though Saracevic and Dalbello had a wider perspective and investigated more issues.

Saracevic and Dalbello (2001) included the results of the survey into the context of a wider investigation. They were looking into more general questions of why universities should teach digital library topics, what these studies should consist of and how they should be taught. This survey also used a wider understanding of digital library studies and investigated courses that were related to the digital libraries (but did not include the words ‘digital library’ or ‘electronic library’ in the title).

The three surveys used the same methods of data collection: e-mail survey and analysis of the material provided on the websites of the higher education departments and in schools.

The surveys provided data on the number of schools offering modules and study programmes on digital libraries in different countries. In 1999, the authors of the surveys found at least 20 modules at the HEIs (higher education institutions) in eight countries (Spink & Cool 1999).

Saracevic and Dalbello received data from the USA, Canada, Australia (2 schools), and UK (1 school). As they have widened the meaning of digital library courses for the USA and Canada, it was observed that: “Of the 56 LIS programs, 47 (89%) include digital library in some form or another or to some degree in their curriculum; for 5 (9%) programs, this cannot be determined and 4 (7%) programs show no presence of digital library education. This analysis shows that digital library education is included in the curriculum of most LIS programs. However, the degree of inclusion varies widely. When considered closely, the inclusion ranged from a full course or courses on digital libraries to metadata standards being covered in organization of information courses to cases in which digital library focus is relegated to continuing education. Participation of the students in a local digital library project is another way of including the digital library topics within LIS education.” (Saracevic and Dalbello 2001).

The Internet survey in 2004 revealed that “42 educational institutions offered courses on digital libraries. (It is possible that more institutions offered courses but did not post any information about the courses on the web.) Most of the educational institutions were located in North America, and some countries in Europe and East Asia. As Spink and Cool (1999) only located 20 institutions offering courses on the subject, it appears that since 1999, the amount of courses on the subject of digital libraries has more than doubled, from 20 in 1999 to 42 in 2003.” (Liu 2004)

While the first survey found that most modules on digital libraries were delivered at the graduate level, the other two also found courses at the undergraduate level. The number of schools providing more than one course on the subject is slowly growing. At present a number of universities offer full programmes in digital librarianship as was stated at the beginning of this article.

Most of the courses on digital libraries were found in the library and information studies departments, some in computer departments and a few in other departments (like, media).
Basic differences in teaching approaches can be traced along the lines of the type of schools that offer them:

“The Technocrats”... teach the mechanics of computer architecture and programming that will ‘house’ the digital libraries of the future (akin to library architects and designers), and “The Librarians-As-Guides”... teach the technical side of digital libraries only as an adjunct to operational/user-oriented understanding of the potential of digital libraries. In many cases, aspects of digital librarianship are incorporated within traditional library materials courses, for instance, library missions and collection policy.” (Liu 2004).

Saracevic and Dalbello identify four broad areas of application in the general domain of digital library education:

- **tools** (technologies and technology based processes);
- **environment** (the context in which digital libraries operate);
- **objects** (representation, structure and life cycle of documents in various formats), and
- **combined** (several areas of applications present without any one being distinctive).

They also state that “digital library education in the two areas, library and information science on the one hand and computer science on the other, do not have any relation. This exactly follows the practice of development and operation of digital libraries in library institutions on the one hand, and the research in digital libraries as funded by programs identified in the section on rationale, on the other hand. While they are in the same planetary system, one is from Venus and the other from Mars” (Saracevic and Dalbello 2001).

Since these big surveys were completed, supplements to them were added in many later articles on the digital library education. For example, Abrizah et al. have found that the DL education in LIS schools can be categorized in three types:

- a) an independent or full library course,
- b) an integrated digital library course with other LIS topics, and
- c) courses with close relation to DL processes.” (Abrizah 2009)

This paper is looking at the three full digital library programmes conducted by four European universities (in library and information science departments) and discusses the reasons behind some solutions taken by developers of DL curricula. Some effort to bring the two planets - Venus and Mars - together can be witnessed in some of these programs.

**Database for the paper**

The data for this paper was collected through different channels mainly during my own involvement in various activities related to these programmes. I have been involved in the development of the DLIS programme from the very conception of the idea in 2004. At present, I coordinate the delivery of the programme. The programme DiSe was developed a little later and with a different target group in mind, but the coordination of two similar programmes provided by the same department is important from the point of view of resources and mutual interests. At present, I am a member of the Steering and Operational group of this programme. The International DLL programme was within the circle of interest of the SSLIS for a long time as well. For a number of legal and administrative reasons the school could not participate in this Erasmus Mundus project, however, the proximity of interests and location brings us together quite often. In addition, I was invited to serve as a subject expert evaluating the Erasmus Mundus programme in 2010 and visited the Oslo University College and Tallinn University looking quite close into the documentation of the programmes, interviewing students and teachers, conversing with the heads of the universities about the programme.
Thus database for this paper consists of the conceptual documents and programme documentation, teachers’ experiences, students’ focus groups transcripts, data evaluation, and study tracks.

**Master’s studies in digital librarianship**

*The origin and nature of the Master’s programme*

All three Master’s programmes have originated roughly during the same period of 2003-2004 mainly in relation to previous experience in LIS education.

Parma University initiated the DLL Master’s after acquiring expertise in running a joint programme for working librarians with Northumbria University. A consortium of three universities (Oslo University College, Tallinn University and Parma University) received financing for the delivery of the Master’s programme in 2007. The programme has a website introducing it to potential students and in general (http://dill.hio.no/).

The international Master’s programme in DLIS – Digital Library and Information Services was initiated in 2004 by a group of senior academics at the Swedish School of Library and Information Science. Swedish universities were preparing for a big change according to Bologna agreement among the members of European Union. It was one of the elements in the introduction of new study structures and programmes and was seen as a possibility to widen the influence of Swedish library studies internationally. The survey of information specialists was conducted to understand the demand for skills on the labour market and the curriculum was developed in 2006-2007, but the University had to wait for the structural change in all Swedish Universities and the rights for delivery of Master’s programme (until then the Master’s programmes did not exist in Sweden). Thus the programme was announced to start in 2008 with a pilot group of students. The programme site is available at http://www.hb.se/wps/portal/bhs/digitallibrary.

In another two years, there was a possibility to apply for financing the development of Master’s programmes together with companies or public organizations from the big Swedish funding organization The Knowledge Foundation. SSLIS capitalized on that and acquired support for running another Master’s programme in digital services for cultural heritage for Swedish students. The programme was started in autumn 2009. The programme site is available at http://www.hb.se/wps/portal/bhs/.

*The goals of the Master’s programmes in digital librarianship*

All three Master’s programmes have developed courses for full-time two year Master’s programme. The universities and the departments dealing with these programmes as well as the developers of the curriculum have been and are active in discussing education for digital librarians (Dahlström 2009; Tamaro 2007) and investigating requirements for information specialists and digital librarians (Maceviciute 2005; Maceviciute et al. 2009). All three programmes have been developed pursuing the same idea—education for future librarians, however, all three followed a slightly different aspect of this education and this resulted in specific curriculum structures.

The DLL programme had to follow the requirements of ERASMUS MUNDUS programs and targeted the non-European countries in Asia and Africa seeking to satisfy their needs for research and leadership in modern librarianship. A consortium of three universities took into account the educational competence existing in the participating departments and built the curriculum on the basis of it.

The developers of the DLIS Master’s programme targeted a different audience. First, they wanted to attract working librarians and
specialists in different organizations who would need to update their professional competence to manage digital resources and develop digital information services. Therefore, the education was conceived in a distance half-time mode. It was planned to reach the worldwide audience and to attract new groups of students to the SSLIS, therefore, English was selected as the language of course delivery.

The DiSe programme had to satisfy the requirements of the funding agency (Knowledge Foundation) that mainly seeks to strengthen relations between academia and professional practice. Thus, the main goal of the programme was to satisfy the need for competence in a library sector related to digital heritage creation and maintenance. Professional organizations, such as the Swedish Royal Library and Gothenburg City Library took part in curriculum development and worked together to ensure that the students acquire not only theoretical knowledge but practical skills as well.

However, each programme was also affected by general university strategies of the parent institutions. For example, Tallinn University has a vision to increase a number of foreign students from 4% to 10% in five years. The main aim is to create international intellectual environment at home for the students and teachers that may be comparable to the one found by Tallinn students when they study abroad. To achieve all these objectives Tallinn University is using internationalization funds to finance international priorities and the ERASMUS MUNDUS programme was one of these opportunities. Similar considerations of international visibility guided the other partners in the project. Somewhat different university strategy influenced the two programmes delivered by SSLIS. The internationalization goals were less important even in the development of the English language programme. The orientation of the University of Borås toward professional practice and a goal to become a university of professions as well as the general context of re-organization of the whole higher education system in Sweden to meet Bologna requirements were driving and directing the creation of DLIS and DiSe.

The curricula of the Master’s programmes in digital librarianship

The curricula of all three Master’s programmes reflect the goals and the situation of the participants working on them. All three of them have a modular structure. Table 1 shows the sequence of the modules in the programmes.

For the DLL programme the modules were coordinated already during writing an application for ERASMUS MUNDUS. In 2007, a meeting for module coordination has taken place. The partners presented their modules in detail, agreed about learning objectives and outcomes, discussed contributions of each partner.

The similarities of the curricula are quite easy to observe. There are four basic building blocks that one can notice in each of them: users and use of digital information, technological issues related to creation, maintenance and access of digital collections, research related and theoretical modules, and organizational management issues. For the DLIS, the module structure and units were selected in relation to the professional demands of the Swedish labour market for information professionals (Maceviciute 2005) and the recommendation of the European LIS curriculum results (Manzuch et al. 2005, Georgy et al. 2005). Later the experience of the Digital Library Curriculum Project conducted by University of North Carolina at Chapel Hill and Virginia Tech (http://curric.dlib.vt.edu/) was taken into account. For the DiSe, the modules and the structure of the programme not only followed the requirements of the partners, but also of other Swedish institutions concerned with digitization and digital cultural heritage. The latter two programmes also followed the requirement of coordination and module
sharing between the two programmes at the same institution.

The differences between the study programmes are also quite obvious. DLL is mainly concerned with research in digital libraries (though they provide practical experience during the second year), while the other two programmes are much more practice oriented. Both Swedish programmes devote more time to the technology issues. DLIS especially emphasizes significant expertise in such areas as information retrieval and digitisation. They also introduce digital library technologies, databases and programming. Though competence and skills provided within these courses remain on the basic level in comparison to digital library courses provided by computer departments, it is quite extensive in comparison to the DLL and some other programmes offered by LIS departments in different universities. The DLIS and DiSe programmes also differ in approach to the practice: while DiSe devotes significant attention and a 15 ECTS module to workplace-related project and Project management, DLIS incorporates smaller practical projects into each module. Both programmes offer a Master’s thesis or an extensive examination project as options for the final examination, while DLL concentrates on traditional Master’s thesis.

Despite clear differences, the three programmes share few significant similarities. These similarities most probably are born from the common understanding of the requirements and situation of digital librarianship in Europe as well as of the educational traditions existing in LIS departments of the European and more specifically Nordic-Baltic higher education institutions. The originators of the programmes have carried out a labour market analysis of one or another kind before launching the programmes (the examination of existing LIS curricula for DLL, a survey of professionals for DLIS or the consultations with leading experts working in partner institutions for DiSe).

**Delivery of the Master’s programmes**

These differences between the programmes manifest themselves in certain aspects of similar modules and in the delivery of the Master’s programmes in digital libraries is specific in several respects.

Both Swedish programmes are delivered on distance and half-time mainly because they are targeting working librarians. In addition, both programmes target specialists spread over the vast areas in Sweden and in the world. According to the targeted audience DiSe programme uses Swedish, while DLIS—English languages. The main vehicle for the programme delivery is the e-learning platform PingPong used by the University of Borås. The teachers create an internal course site for each module. Most of the modules run over the whole term. The main building block of each module is a unit devoted to a certain topic. The modules in DLIS tend to be quite large, exhaustive and diversified. Some wider topics may be covered in several modules (e.g. the issues of the collection development are present in *Digital Library Management*, *Information Retrieval* and *Digitizing cultural heritage material* modules). The modules in DiSe are smaller as a rule. The other difference is also determined by the student population. While DiSe tends to organize two-three short (one-two days) meetings per term, DLIS organizes one longer (at least one week) residential period. Both programmes were running free of charge till present, but Swedish government has introduced tuition charge for the students from the non-European countries since autumn 2011.

The difference of these programmes in delivery from the DLL is evident. The DLL programme is full-time and is delivered at the three universities. The students move to a different university for each term (Oslo – during the first; Tallinn – second; Parma – third) and choose one of the partner institutions for their final course in writing the
Master’s thesis. In addition, students attend summer schools organized before the official start of the school year and have an internship period by the end of the third term. Students also may use the opportunities to attend international digital conferences organized in Europe. Significant resources required for this particular model of programme delivery are covered by ERASMUS MUNDUS project. The language of delivery is English. In addition, the students have a possibility to take local language course at the partner universities. However, it is interesting that the cohesiveness of the programme throughout the three universities is ensured by using the IVA e-learning platform of Tallinn University.

### Teachers and students in the Master’s programmes

#### Teachers

All three programmes have assembled a competent team of teachers with strong research profile and teaching experience. Quite a significant percentage of the staff working on these programmes conducts research project in the area of digital libraries, digital publishing, digitization and digital preservation, information retrieval, information literacy or other related topics. The most important structural element in each programme is collaboration among the teachers through a variety of formal and informal meetings, the coordination of various

---

Table 1 The modules in Master’s programmes for digital libraries

<table>
<thead>
<tr>
<th>DLIS</th>
<th>DiSe</th>
<th>DLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Digital library management (15 ECTS)</td>
<td>1. Digital media in the culture and information sectors (7,5 ECTS)</td>
<td>1. Research methods and theory of science (15 ECTS)</td>
</tr>
<tr>
<td>2. Users and information behaviour in digital environments (15 ECTS)</td>
<td>2. Technology of digital libraries (7,5 ECTS)</td>
<td>2. Digital document (15 ECTS)</td>
</tr>
<tr>
<td>3. Information retrieval for digital libraries - I and II (15 ECTS)</td>
<td>3. Users and information behaviour in digital environments (15 ECTS)</td>
<td>3. Information and knowledge management (15 ECTS)</td>
</tr>
<tr>
<td>5. Digitising cultural heritage material (15 ECTS)</td>
<td>5. Elective courses (22,5 ECTS)</td>
<td>5. Access to digital libraries (15 ECTS)</td>
</tr>
<tr>
<td>8. Masters’ thesis (30 ECTS)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note on colour coding:  - theory and research,  - use and users,  - technology related,  - project,  - management.
course elements and creativeness in programme delivery.

The coordination of the courses and activities between partner universities in the DLL demand also quite close collaboration of the teachers in Norway, Estonia and Italy and they participate in regular coordination meetings. Both teams of Swedish programmes work together, share teaching resources and are involved in each others governing structures.

Another important feature is the educational experience of most of the teachers and the communication style. Most of the teachers are regarded by the students on all programmes as professional, accessible, helpful and resourceful, though, of course, some personality clashes or dissatisfaction moments appear occasionally in surveys or quality monitoring exercises.

Admission and dropout

Modern digital library studies in English attract a considerable number of students. However, one should also bear in mind that educational programmes require a significant time to earn reputation. Sometimes the reputation is provided by external bodies, like in the form of accreditation. To some extent ERASMUS MUNDUS is a label that guarantees the quality of a programme to a large extent. Therefore, it should not be surprising that DLL attracts over 300 candidates for 12 scholarships available for the students from Asian, African and some European countries.

The number of students attracted by a programme delivered by a single institution would be subject to change in relation to a variety of factors, but in principle the number of applicants to a successful programme with good reputation should be growing and stabilizing at some point. The first admission to the DLIS programme was announced in 2008 and there were 32 applications. The department considered this to be a rather low number for a distance programme and decided to start an experimental run to which only 12 students were admitted. In 2009, the number of applicants increased to 55 and the programme was started officially. The third intake in the spring of 2010 attracted almost 200 applicants, for the autumn 2010 SSLIS received over 400 applications, and for the intake in spring 2011, around 1,742 applications were sent in.

The school has to select the best 30 students out of 500 eligible students. It is difficult to establish the factors that have influenced the number of applicants for the current intake and we will have to wait to find out if the trend proceeds.

The Swedish DiSe programme is in a different position. It neither offers an attractive scholarship nor appeals to the worldwide public. Therefore, the numbers of applicants remain on the level with other Swedish Master’s programmes in the LIS and similar areas—on an average around 30 students apply and are admitted to it annually.

The admission criteria for DLL are: academic record (40%), experience (20%), statement of purpose (30%), and references (10%). The DLIS and DiSe have chosen a more formal criterion: the number of academic credits. In borderline cases, the work experience is taken into account. For DLL and DLIS the English language proficiency is an important requirement.

Most of the applicants for DLL and DiSe have either previous degree in LIS or related disciplines, while DLIS attracts a significant number of students with computer science, economics and business as well as technology background.

The origin of the students is also quite different in DLL and DLIS. While the DLL is dominated by the students from Asia and Africa with a very low number of applicants from European countries, the DLIS attracts mostly European citizens and residents. However, the applicants also come from the USA and Canada, Australia and New Zealand, Africa, Middle East and Asian countries.

The on campus full-time programme of DLL over the three intakes has registered only two dropout cases. The high quality of students is
ensured by highly competitive admission process, which is one of the factors of high graduation percentage.

For distance studies a very high dropout rate of the students is one of the main problems. On an average only one-third of the students manage to finish the first course. The dropout rate slows down only during the second half of the studies. At present, there are five students following the last courses in the DLIS from 12 students who started in 2008.

It is not easy to deal with this problem, especially, regarding the economics of the programme delivery. Therefore, SSLIS has developed several methods to solve the problem and at present tests their effectiveness.

One of the solutions may be to concentrate the demand for the programme and to admit larger number of students every second year. Another is to take in the students every term and introduce flexibility into the system allowing the university to accumulate students from different intakes in one course. This approach was taken by DLIS and is tested right now. The high demand for the programme and severe competition may also result in admission of highly motivated and good students. This may solve the dropout problem without additional efforts, but it has to be proved as yet. Meanwhile the SSLIS tries to increase the flexibility of the programme for individual students by providing a possibility to take more courses per term than planned and finish sooner or on the contrary, slowing down the pace of individual studies. So far, the school has improved the economic foundation of each module, but the problem of the dropout still persists.

**Student satisfaction**

Despite high dropout from both distance programmes DLIS and DiSe, the level of students’ satisfaction registered in each course is quite high. The first courses on the programme are rated especially high for the delivery mode and the educational forms as well as for the content. The technologically oriented modules in DLIS that were expected to scare away the students with humanities orientation do not produce this effect on a worrisome scale, i.e., the dropout from these courses is not higher (sometimes even less) than from the other ones.

The content of the programmes has been seen as relevant and logical, building on the previous knowledge and/or practical expertise. It is also innovative and up to date. The students expect that the studies will enhance their possibilities of finding a good job or even an advance in career. In this respect it is important to note that the DLL students see their future in research and most of them after graduation plan further doctoral studies. They also have expressed the opinion that their careers in this profession may be inhibited by the lack of technology-related courses. The students in distance programmes, DLIS and DiSe, seem to cope quite well with technology subjects and some have expressed the opinion that they “have provided exciting challenges”. How useful these courses will prove to be in future, however, remains to be seen.

The course literature in all three programmes seems to be consistent with the trends established by the Digital Library Curriculum Project team: “there is no core set of readings assigned in DL courses, but there is a core set of authors whose works are assigned. This study also found that there is a core set of journals from which readings are assigned” (Pomerantz et al 2006). The literature is also constantly renewed. The students seem to be appreciative of the assigned readings and become active information seekers sharing their findings with course mates on a regular basis.

Another high satisfaction factor is the cohesiveness of the student groups. In the DLL programme, this cohesiveness is achieved quite quickly due to the proximity among the groups of individuals living in alien environments and experiencing studies and change of scenery together. In the distance programmes it is not
easy to achieve the same level of closeness. Residential periods seem to be the main element that helps to build relationships in the groups. Quite a significant number of students prefer to attend both residential periods per year, though only one is obligatory. However, the resources needed for this are quite beyond the financial capacity of our students from African countries. They especially value the interaction possibilities created on the PingPong platform, the blog of DiSe, by e-mail and other means.

**Emerging library management perspective and its manifestation in the programmes**

The three programmes have much in common with one another, in spite of having been devised according to rather different ideas.

Wilson’s (Wilson 2001) Venn diagram of the scope of librarianship (see Figure 1) was used as a guiding concept for the DLIS programme. Given the target market of practicing librarians, the focus of its curriculum planning was on the management of digital libraries as organizations. Consequently, the irregular area labelled D in diagram constitutes the core, since it identifies the organizational framework within which people are both served and (in the case of staff) managed, where content is organized, digitized and made available for use, and where systems are developed and employed to manage the digitized information resources (Macevicute et al. 2009).

The DELOS conceptual model of digital libraries is applied for structuring the DLL programme. Its basic scope is rather similar to the components in the Venn diagram. It includes the following elements: digital contents (including copyright issues), people (librarians and users), and systems. As the first one this model addresses mainly the global scope of the studies and is not related to country context.

The third DiSe programme attempted “to integrate three levels of scale – macro, meso (interpreted as national), and micro – in a master’s programme … an important task has been to collaborate closely with partners from libraries and companies with varying local and national agendas.” (Dahlström and Francke 2009).

Thus the underlying foundation for modules in each of the programme differs:

- DLIS focusses on the education of the managers and senior specialists of organizations (physical or virtual) organizing digital collections and providing digital services,
- DLL directs main attention to the leaders of innovation and change in libraries with specific focus on developing research abilities, and
- DiSe concentrates on providing competence for local institutions for digital media and digital collections management.

Nevertheless, the library and information service management perspective is quite clearly built into all three programmes. DLIS devotes only one course to the Digital Library Management, but most of the topics included into it are also present in other courses. The managerial perspective guides the depth and level of the technological courses and service development for the users. The problems for the final Master’s examination should also include a managerial aspect.
The DLL programme is clearly stating the LIS management perspective through the courses delivered at the Tallinn University: Information and knowledge management and Human resource management, including leadership issues. Both courses seem to be diverted from the digital library context and address more general management areas in organizations. However, as students stated in the interviews: “all the doubts of the necessity of such courses in the programme evaporate by the end of the courses; after all digital libraries are run by people and on information, therefore it is very important to manage both in the most appropriate way.”

The third programme puts strong emphasis on project management within the library by including a course and a workplace project. This is clearly related to the organizing of work with digital library resource collections in Swedish libraries. Thus, all elements of LIS management are introduced for this particular purpose.

So, in one way or another all three programmes build their perspectives as the notion of change management, either through innovation, organizational change or project implementation. This perspective cultivated in LIS management for several decades since introduction of computer technology into libraries unites all the three programmes as built on the common foundation of LIS management.

References


This unique collection of stories from across India, South Asia, and the world brings to you personal accounts of struggle, survival, trust, and hope for a better tomorrow. From the pollution-choked rivers in our cities, contamination in our food, to the carbon footprint of the US elections; from the promise of smokeless chulhas to the scenario in which we run out of oil; from the slow death of our historical heritage to the plight of the magnificent big cats, this thorough, complete, and meaningful anthology takes a broad sweep over the past few years to highlight and present the best and the biggest stories.

Table of contents (Selected)

Is sustainable development an oxymoron? (Dr R K Pachauri) • Is development passé? (Darryl D’Monte) • The citizen vs tiger debate (Malini Shankar) • Mahatma and the environment (John S Moolakkattu) • The Ganga’s last stretch (Julian Crandall Hollick) • If oil runs out... (Bhaskar Hazarika) • Can global poverty be eradicated? (Leena Srivastava) • Fear comes full circle in Kashmir (Faisul Yaseen) • For whom the bells toll (Anuradha Dutt) • Tainted tales (Ramesh Menon) and much more...

Key features

• Unique collection of articles on development and underdevelopment from India and abroad
• Contributions by well-known journalists and environmentalists
• Inspirational stories of struggle, survival, and hope
• Immensely readable and informative
The digital repository of the Swiss Federal Archives: a case study

Jean-Marc Comment, Swiss Federal Archives
Archivstrasse 24, 3003 Berne, Switzerland

Jean-Marc.Comment@bar.admin.ch

Abstract
The Swiss Federal Archives has a digital repository in production since September 2009. The author presents this digital repository. He discusses the challenges faced and also points out the open issues.

Key words
Long-term preservation, digital repository, format, geodata
**Introduction**

The Swiss Federal Archives is mandated by the federal law on archiving to store the documents of the Swiss federal administration. This also applies to the documents in digital form.

The Swiss Federal Archives has now introduced a Digital Repository and has trained its staff to manage the same. With the Digital Repository, the Swiss Federal Archives are now in the position to ingest, store and access digital documents from record management systems, databases and from the collections of files.

The Digital Repository set the prerequisites for the archiving of digital documents. For efficient management of the delivery of digital submissions at the Swiss Federal Archives, the government units must be enabled to deliver digital documents in data packets compliant with the structure and formats defined by Swiss Federal Archives and deliver them on a suitable medium to the Swiss Federal Archives.

We have developed additional tools in accordance with the requirements of the Swiss Federal Archives for helping the implementation of Records Management Systems in different administrative units of the Swiss federal administration. The Swiss Federal Archives provides technical assistance when needed. New examples may be seen on the website of the Swiss Federal Archives.¹

The conceptual basis for digital archiving at the Swiss Federal Archives is based on the ISO OAIS Open Archival Information System Reference Model (ISO 14721:2003). The reference model describes an archive as an organization where people and systems collaborate in order to preserve information and make it accessible to a designated community of users. The general architecture of the application landscape of the Swiss Federal Archives (see Figure 1) respects the OAIS reference model.

The principal components of the digital repository are the Digital Information Repository and the Infrastructure Digital Archive.

---

The government units submit through the archival interfaces a Submission Information Package. The digital information repository is informed of the arrival of new Submission Information Package on the transfer platform. The archivist makes quality tests of the submission information package with the help of the digital information repository and a software called Package Handler. When all quality tests are passed, the digital information repository builds an Archival Information Package. Each archival information package has a unique address and is stored in the Infrastructure Digital Archive in three different copies at three different locations.

The logical metadata concerning the archived data, the provenance of the data and the order of the data are transmitted to the archival information system. This archival information system contains the same metadata for digital and non-digital material. A subset of all metadata of the archival information system is on the Internet within the Findings Aids. When a customer searches through the Finding Aids and finds a digital file, at this time he must ask an archivist to search the right file in the digital information repository with the unique address. The digital information repository will then prepare a Dissemination Information Package. This dissemination information package will be at the disposal of the customer on the dissipation platform.

Applications

**Application of “Digital Information Repository”**

The company Tessella in UK has developed the digital information repository on basis of the software Safe Deposit Box (SDB). This application, based on the reference model OAIS, must allow the Swiss Federal Archives to:

- Ingest and store digital data from record management systems, relational databases and file collections.
- Preserve digital data over the time.
- Access digital data at each time.

You can recognize five of the main components of the OAIS reference model: Ingest, Preservation, Access, Data Management and Administration. The Package Workflow Control component allows the Swiss Federal Archives to manage the processes within the digital information repository. The reporting component allows the Swiss Federal Archives to get information on the archival units within the digital information repository. For example, how many documents of a special format are in the digital information repository? The Storage Adaptation is the component of the digital information repository which makes the link to the infrastructure digital archive.

**Infrastructure Digital Archive**

At the bottom of the figure are the two copies on tapes and the third copy on a Network Area Storage (NAS). This whole infrastructure (as shown in the figure) is used in a secure network. The archivist accesses this network with the

---

protocol https. PRONOM is a database of file formats.

**Standards**

**Digital archiving policy**

The Swiss Federal Archives have also published a policy for digital archiving. This policy explains to the stakeholders, to the public and to the staff of the government units how to deal with digital archiving. In this policy, the reader will find a general framework, the strategy and standards, the processes, the organization, and the release and change management. The digital archiving policy has been published on the website of Swiss Federal Archives.

**Standardization of the digital package**

The Swiss Federal Archives have decided to define the submission information package, the archival information package and the dissemination information package - in detail. All these three different packages have the same structure.

The definition of only one digital package for all the submissions to the Swiss Federal Archives will reduce the complexity. The specification of the submission information package has been published on the website of Swiss Federal Archives.

![Figure 4 The structure of Submission Information Packages](Image)

![Figure 5 The internal structure of a Submission Information Package](Image)

**Standardization of the digital formats**

The Swiss Federal Archives will also define and publish a list of file formats suitable for digital archiving. The list is very short and contains only eight file formats.

The eight formats are, namely:
- TXT for unstructured text data
- PDF/A for printable documents
- CSV for data in tabular form
- SIARD for relational databases
- TIFF for halftone images
- WAVE for audio data
- MPEG/4 for video data

---

The Swiss Federal Archives is confident and from previous experiences in pilot projects it is of the opinion that this list is sufficient for most of the formats founded in the Swiss federal administration. By the choice and definition of a file format, one of the most important requirements of the file format is the openness. The implementation of the format must be open documented. The Swiss Federal Archives can order, in case of disappearing of a viewer tool for one format, the implementation of a new viewer. Other requirements are the robustness and the market importance of the format.

In the case of relational databases, no open standard exists in the world. The Swiss Federal Archives decided to define a format for relational databases called Software for Invariant Archiving of Relational Databases (SIARD). This format is based on open standards like XML and has been published in 2008. It has been accepted as the database standard within the European project - Planets. During the year 2010, it has been also accepted by the National Archives of Germany, France and Canada and with minor changes by the National Archives of Denmark.

The Swiss Federal Archives has developed a software called SIARD Suite to produce and read the SIARD format. The software SIARD Suite can be ordered free of charge by the Swiss Federal Archives.

During year 2010, the Swiss Federal Archives conducted a survey among the customers of SIARD. The results were very positive. All customers have used the software without problem and wanted that the Swiss Federal Archives extend the amount of types of databases to migrate. At this time, the SIARD Suite can do the migration from ORACLE, MS SQL and Access database to the SIARD format. The survey has shown that the next database to migrate through SIARD Suite is mySql.

**Pilot projects**

In parallel to the definition of archival principles and the implementation of the digital information repository, the Swiss Federal Archives has run several pilot projects to gain confidence in the correctness of the archival principles and of the requirements of the digital information repository.

The requirements for these pilot projects were:
- a definition of standard formats for preserving the data;
- a choice of real and not test data;
- the whole life cycle of digital documents from appraisal to access must be covered by the pilot project; and
- all the internal staff must be involved, in particular the people who will deal with the data to be preserved in the future. When the pilot project is completed, the staff should have expanded their skills and be prepared to start preserving digital documents on a regular basis.

The Swiss Federal Archives has run in the past two pilot projects: DAp, which is an acronym for Database Archival Process and GAp, which is an acronym for GEVER\(^7\) Archival Process.

But in the Information World, the situation is still under process. So the Swiss Federal Archives has started a new preliminary study with the Federal Office of Topography about geographical data (geodata) coming from the Geographical Information System (GIS). The goal of this study is to examine the robustness of the submission information package specification and to define a file format for geodata.

The preliminary study was based on the principles of digital archiving used by the Swiss Federal Archives and the new federal legislation

---


\(^7\) GEVER means Records Management in German
governing the geoinformation. The study also took account of the international context. It was ascertained that no solution for the archiving of geodata currently exists.

The aim of the preliminary study was to define criteria not only for the development of an internal solution for the Swiss federal administration, but also to secure nationwide coordination of all aspects of geoinformation, including the archiving of geodata.

Studying a variety of models, it was observed that the existing digital archiving solution in use at the Swiss Federal Archives could (and should) serve as the basis for the archiving of geodata. The options of using the SIARD suite and format and the data warehouse model of the Federal Office of Meteorology and Climatology were examined, but both models were found to be inadequate for the declared purposes.

A first prototype of a method of delivering geodata to the archive was developed. It was based on the already applicable requirements for digital transfer. The components of a documentation system and procedures for the integration of metadata have already been identified. These are being integrated into existing standards: archiving metadata based on the criteria of the Swiss Federal Archives, and geo-metadata in accordance with the official GM03 model.

The preliminary study put forward proposals for formats that would be suitable for archiving (for both vector and raster data). These were selected on the basis of the criteria of the already existing catalogue of specifications of the Swiss Federal Archives concerning suitable formats for archiving, to which additional criteria specifically relating to geodata were added for the purpose of evaluating geo-formats.

You can read more about this preliminary study on the website of the Swiss Federal Archives.

Open issues

Interfaces to the federal agencies

Metamorphose is another new project started by the Swiss Federal Archives. The goals of this project include the following:

- to define the processes of closing files in Records Management System and to prepare the building of a submission information package.
- to implement these processes in the most common Records Management System of the Swiss federal administration.

We have finished to define the requirements and have started implementing them in the most common Records Management System of the Swiss federal administration.

New standards for file formats

A major problem faced is that of conversion to the right formats. So the Swiss Federal Archives is still working with government units, IT-suppliers and software producers to find automated solutions to this very problem.

But with the apparition of the new formats like Open Office XML (OOXML) and Open Document Format (ODF), we have now the production format which can also be of archival format. So if we choose these formats as archival format, we do not need to convert the produced documents. The Swiss Federal Archives has made in 2010 a preliminary study with the format XLSX. The first results of this study were very positive.

---


Information search behaviour in the German Education Index

Carola Carstens, German Institute for International Educational Research, Information Center for Education, Schloßstraße 29, 60486 Frankfurt, Germany

carstens@dipf.de

Marc Rittberger, German Institute for International Educational Research, Information Center for Education Schloßstraße 29, 60486 Frankfurt, Germany

Verena Wissel, German Institute for International Educational Research, Information Center for Education Schloßstraße 29, 60486 Frankfurt, Germany


Abstract
This paper presents results from an analysis of user interactions in the German Education Index (GEI) specialized search engine. The main interest of this study is to gain insight into the GEI users’ typical search behaviour to identify optimization potential for the further development of the GEI search engine. Based on a sample of query logfiles, our study reveals that the users apply advanced search functionalities and query reformulation tactics adequately, though not frequently, and that more support in their application might increase the effectiveness of searches in the GEI.

Keywords
Information search behaviour, search tactics, search strategies, logfile analysis, digital libraries
Introduction

In DL (digital library) research, the study of user interactions is a recent trend, as stated by Xie. Compared to general purpose search engines, digital libraries, as well as specialized search engines, commonly offer more sophisticated advanced search functionalities such as the search in different fields and the support of a Boolean search syntax.

While there exists a large body of research on search behaviour in general purpose web search engines, this is by far less in the case for DL and specialized search engines. To contribute to the study of user interactions in DL-like environments, this paper presents results from the analysis of user interactions in the GEI, a search engine specializing in education.

Related Work

The study of user interactions in retrieval systems can be subsumed under information search behaviour, which Wilson defines as being concerned with ‘the interactions between information user (with or without an intermediary) and computer-based information systems, one type of this being information retrieval systems for textual data’. (p. 263)

Bates and Harter enumerate tactics and strategies that can serve to analyse the way users interact with retrieval systems. Bates defines a search tactic as ‘a move made to further a search’, thus serving the realization of a certain retrieval goal. (p. 207) She presents a catalogue of 29 search tactics. Amongst others, she distinguishes two kinds of tactics – search formulation tactics and term tactics. The former refer to the design of the query structure. For example, adding a new facet to a query would fall into this category. Term tactics, by contrast, describe the selection of terms in the query formulation. For example, this includes the use of related, broader or narrower terms. As defined by Harter, different tactics may form part of a search strategy, the ‘overall plan or approach for a search problem’. (p. 170)

A distinction can be drawn between subject-search strategies and known-item search strategies: While the former serve to query for documents that cover a certain topic, the latter query by already known document facts.

As delineated by Jansen, the query logfile analysis is a well-established method of examining the users’ search behaviour, which has often been applied to web search engines. For example, Silverstein et al. and Rieh and Xie have analysed query logfiles of the Excite search engine, while Spink et al. have applied this method to gaining insights into the Altavista users’ search behaviour. The study by Rieh and Xie analysed the nature of query reformulations in the Excite search engine in depth. The authors identified parallel movements, such as the use of related terms, as the most common of all reformulation patterns (51.4%). Query specifications were less frequent (29.1%) but still more common than generalizations (15.8%), while the replacement with synonyms was rare (3.7%).

Spink et al. further state that many users (48.4%) submitted only a single query and consulted only a few documents. Another finding in their study refers to the use of advanced search functionalities such as Boolean operators, which is reported to be scarce. Spink et al. conclude that ‘most people use few search terms, few modified queries, view few web pages, and rarely use advanced search features’. (p. 226)

Jones et al. report a more frequent use of Boolean operators than Spink et al. in their analysis of search behaviour in a digital library environment. Nevertheless, Jones et al. found that the use of Boolean search was influenced by the default search configuration. During the experimental period, this configuration varied between Boolean and ranked search. In either case, users showed a tendency to maintain the pre-selected configuration.

Wildemuth and Moore conducted a study on search engine behaviour in a specialized search engine. They analysed 161 searches in
Information search behaviour in the German Education Index

Search behaviour in the GEI specialized search engine was already explored in a predecessor study by Carstens et al., aiming to assess whether search tactics and strategies listed by Bates and Harter are identifiable in the query log files of the GEI. This explorative analysis used a sample of long sessions that were expected to comprise complex searches and focused on the analysis of subject searches. As a result, 19 different tactics and 6 search strategies could be identified in the data set, illustrated by examples in the paper.

Research Interest

The study at hand builds on the above mentioned results of the study by Carstens et al., applies the query log file analysis to a more comprehensive data set and analyses it in more depth. For example, the number of hits is taken into account in the analysis, which allows to investigate query result-specific user reactions and to draw conclusions regarding the possible effectiveness of the searches. This work goes beyond the previous study in also examining characteristics of known item searches.

The study will thus reveal which tactics and strategies the GEI users typically apply. As stated in the related works section, several studies have demonstrated that users scarcely apply advanced search functionalities which might increase the effectiveness of their searches. We will examine whether this also holds true for the GEI; respective findings can serve to derive ideas for the further development of the GEI.

Moreover, the study will deliver insights into user interactions in a DL-like environment which offers sophisticated search functionalities for retrieving metadata documents.

The German Education Index

The German Education Index (GEI) specialized search engine, a part of the German Education Portal (GEP), comprises more than 700,000 bibliographic references for the domain of educational research, primarily in the German language (more than 80%). Its metadata documents relate to index terms, title, author/editor, institutions, abstract and source, while full texts are rarely indexed in the GEI.

The GEI is based on the Lucene open source search engine framework. Its underlying retrieval model is a combination of the Boolean Model and the Vector Space Model. The GEI implements both a simple and an advanced search mode, the latter being the default entry point that is accessible from the GEP. The simple search is implemented as a free text search across the above listed document fields.

In the advanced search mode, users can define more sophisticated queries than in the simple search (Figure 1). For example, they can restrict their search to certain fields (1) that can also be combined by the Boolean operators AND, OR and NOT. (2) Within each field, query terms are by default connected by the AND operator. (3) If desired, this pre-configuration can be changed by the users, thus leading to the combination of query terms by the OR operator. Moreover, a person and an index term register can be consulted to look up query terms for the

Figure 1 Advanced search mode in the GEI

respective search fields. The latter register also serves to identify synonyms of query terms which are used to automatically expand the queries.

A survey by Wendt and Patjens revealed that the majority of the surveyed GEP users have an academic background in an educational field and use specialized information primarily for research purposes.

**Research Method**

To gain a comprehensive insight into the search behaviour of GEI users, the query logfiles of one typical working day, October 15, 2009, are analysed. For this purpose, the logfile entries are grouped by their anonymized IP addresses, resulting in 870 IP-specific sessions. As shown in Figure 2, these may comprise several distinct search sessions whose consecutive query formulation steps are both timely and topically related, the latter being verified by a human assessor. This way, an amount of 1,823 search sessions have been identified.

Each search session may comprise one or more distinct search steps, as illustrated in Figure 2. For the purpose of analysis, each search step is classified as either belonging to the subject search or to the known item search category. By analysing how queries are reformulated from step to step, search tactics are identified and categorized by a human assessor. A sequence of such distinct tactics can suggest that a certain search strategy is applied.

In search session 2, presented in Figure 2, a query is specified by applying the tactic of subsequently adding query facets. The searcher may have decided not to include all query facets in the initial query in order to prevent over-specification. New query facets are then consecutively added to the query, following the most general query first strategy.

Harter refers to an initial query starting with the most specific facet in terms of the most specific concept first strategy. This strategy makes assumptions about the users’ expectations on the number of result documents and is therefore not directly identifiable by logfile analysis. However, the description of this strategy inspired Carstens et al. to define two new strategies, referred to as the most specific query first and most general query first strategies. Instead of making assumptions about the expected number of result documents, these strategies describe the query structure and the specificity of the query terms. Pursuant to the most general query first strategy, the searcher starts with the most general facet(s) and consecutively adds new facets that further specify the search and imply a reduction of the result list. The most specific query first strategy proceeds in the opposite direction. It starts with a specific query, often consisting of several facets, which are consecutively deleted or reformulated in more general terms. Examples of these strategies are given in Figure 3 which

---

Figure 2 Identification of tactics and strategies in query logfiles

<table>
<thead>
<tr>
<th>Step</th>
<th>Search session 1</th>
<th>Type</th>
<th>Tactic</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Klieme AND &quot;teaching patterns&quot;</td>
<td>K</td>
<td>K</td>
<td>Search by Document Facts</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Search session 2</th>
<th>Type</th>
<th>Tactic</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>education</td>
<td>S</td>
<td>AND-element added</td>
<td>Most General Query First</td>
</tr>
<tr>
<td>2</td>
<td>education AND fundamentals</td>
<td>S</td>
<td>AND-element added</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>education AND fundamentals AND task</td>
<td>S</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Search session 3</th>
<th>Type</th>
<th>Tactic</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>education AND standards</td>
<td>S</td>
<td>Briefsearch</td>
<td></td>
</tr>
</tbody>
</table>

---

Information search behaviour in the German Education Index

Apart from these sophisticated subject search strategies, the more simple quick approach defined by Chu should be mentioned: It describes a search without using Boolean operators. As query terms are by default combined by the AND operator in the GEI, the quick approach only refers to single-term queries.

The brief search strategy is often applied to get a first idea of the documents in a retrieval system. For this purpose, few query terms are combined with Boolean operators. Harter points out that this basic strategy often serves as an entry point to more complex strategies like the building blocks approach wherein distinct query facets are expanded with semantically related terms, as illustrated by search session 3 in Figure 2.

A further search strategy is called pairwise facets; it is applicable if all query facets are considered to be equally important, as explained by Harter. Following the pairwise facets strategy, only two facets at a time are combined and finally, the result sets of all facet combinations are merged.

### Results

Selected results from the log file analysis are presented in the following. They refer to characteristics and tactics of both known item and subject searches. Out of the entirety of 3,631 search steps recorded in the analysed search sessions, more than 33% are categorized as known item search steps, while subject search steps account for 67% of the search steps.

#### Search mode

An analysis of the search masks chosen by the users demonstrates that most of the search steps are conducted in the advanced search mask, namely 85%. This may be influenced by the fact that the GEI’s advanced search mask is directly linked to the German Education Portal entry page.

Although the advanced search mask is frequently employed, only the free text field is used in 47% of the search steps entered in the advanced search mode. The advanced search mask is thus frequently deployed in a simple search mode.

#### Strategies of subject searches

As can be seen from Figure 4, of all the subject-search steps, a large part (28%) is single-step quick approach searches. In special cases, quick approaches may also comprise more than one step, i.e., if a single-term query is reformulated, for example by changing the spelling.

### Brief searches

Brief searches are also comparatively common, accounting for 17% of the subject search steps. These are often single-step search sessions. In some cases, brief searches also

<table>
<thead>
<tr>
<th>Subject Search Strategies</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quick Approach</td>
<td>1) pedagogics</td>
</tr>
<tr>
<td>Briefsearch</td>
<td>1) deaf AND “language support”</td>
</tr>
</tbody>
</table>
| Pairwise Facets           | 1) “intercultural competence” AND pedagogics AND learning  
2) “intercultural competence” AND pedagogics  
3) “intercultural competence” AND learning |
| Most General Query First  | 1) education  
2) education AND fundamentals |
| Most Specific Query First | 1) “comparative tests” AND mathematics  
2) test AND mathematics |
| Building Blocks           | 1) “method competence” AND (library OR “school library”) |

Figure 3 Examples for subject search strategies
comprise two or three steps, for example if the query terms are replaced by a spelling variant. The search steps classified as unintended brief searches account for 3% of the search steps. In these cases, queries in which phrases or compound terms are entered without phrasing although this would possibly have been adequate. Due to the default configuration in the GEI, they are combined by the AND operator, thus being interpreted as a Boolean combination of terms.

The most general query first and most specific query first strategies consist of several query steps as the queries are consecutively generalized or specialised during the search session. Figure 4 illustrates that the length of these strategies ranges from 3 to 9 steps in the GEI data set. Around 13% of the 2,442 subject-search steps belong to a search sequence following the most general query first strategy, while 11% form part of a most specific query first strategy.

The combination of these two strategies is also common. This is illustrated by the fact that 26% of the subject search steps form part of a combined general and specific search strategy, either starting with a general or a specific query formulation. The pairwise facets strategy, by contrast, is rarely applied, and the building blocks approach could not be identified at all in this data set.

**Types of known item searches**

Figure 5 depicts the distribution of different types of known item search steps. A high percentage (26%) of the total amount of 1,188 known item search steps is precision-oriented where users possibly expect a small set of specific result documents. These are, for example, searches by title where typically only one document is viewed if any result documents are delivered.

Search steps classified as searches by document facts, by contrast, do not necessarily query for one specific document. Instead, already known facts are used to specify search criteria like the combination of author and year or topic and year. Twenty-eight per cent of the known item search steps belong to this category.

Searching only by author name is very common. This strategy accounts for 42% of the known item search steps. Roughly half of these search steps (54%) are incorporated in longer search sessions, while the remaining 46% occur in single-step search sessions. Although 74% of the single-step sessions by author name induce hits, in the majority of these search sessions no document is viewed (61%). This may indicate that the GEI is either used to merely assess if a certain author is listed in the database or to check if new publications by an author are listed.

However, users often seem to be uncertain about how to adequately formulate a query for...
an author name. The 501 search steps querying for author names comprise 92 two- or more-step sequences where author names are varied, for example, by spelling variants, changing the order of first and last names or consulting the person register. These reformulations thus account for a large number of the search steps by author and they show the need for user support. In the current GEI implementation, such support is offered by the person register which is consulted in roughly a quarter of the author searches.

**Search fields of known item searches**

In many of the cases where the users do not consult the person register, they enter their search query into the designated author search field. However, the free text field is used for this purpose nearly as often, as shown in Figure 6. If users search for a specific document by its title or parts of the title, these queries are most frequently entered into the free text search field while the title field is used in 19% of the cases, as illustrated in Figure 7. The author field is often used to specify these queries for specific documents. In 18% of the searches for specific documents, the author field is combined with the title field, and in 12% of these searches, it is employed in combination with the free text field.

In the queries for title searches, phrasing would usually be appropriate but this pertains to only 4% of the searches for specific documents. Title searches without phrasing probably lead to the expected result documents if the entered title is both long and specific. For short title queries, by contrast, large and probably unsatisfying result lists are likely, especially if the query is entered into the free text field.

**Search tactics**

The following Figure 8 shows that considering all of the 3,631 analysed search steps, most of the types of search tactics rarely occur.

While Bates distinguishes between term tactics (for example, the use of a synonym) and search formulation tactics (for example, the use of an AND-operator), findings from our predecessor study (Carstens et al.) inspired us to specify these tactics by analysing term and query structure characteristics in combination. For example, a certain term type like a synonym can be added to a query, it can replace another term or it can be deleted from the query. Consequently, these variants are recorded as distinct tactics, as shown in Figure 8.

Figure 8 shows that the replacement of terms with semantically related terms constitutes the most common category of tactics. More specifically, terms are most frequently replaced by related terms, followed by synonyms, broader terms and narrower terms whose total number of occurrence are very low, compared to related terms.

Figure 8 further illustrates that terms are rarely combined by the Boolean OR operator,
which would for example be an effective strategy for using synonyms: Instead, synonyms are only used to replace query terms. The rare combination of terms by the OR operator shows that the default Boolean AND operator in the GEI is hardly ever changed. A tactic that occurs relatively more frequently is the conversion of query terms,
which can be identified 84 times in the data set; for example, it describes the conversions between singular and plural forms or adjective and noun forms of a term. Apart from conversions, the correction of spelling mistakes and the use of spelling variants are term tactics with a comparatively high frequency. The keyword register is used 48 times for reformulating queries while the person register is consulted slightly more often for this purpose, namely 65 times.

Among the tactics that refer merely to the query structure, the addition of a facet is common (307 times), slightly more frequent than the deletion of a facet (295 times) and the replacement of a facet by a semantically unrelated new facet (213 times).

As depicted in Figure 8, users often change the search fields of a search facet to reformulate their queries (202 times). For example, they switch from the keyword to the free text field, from the free text to the title field or vice versa.

**How users react to few or zero results in subject searches**

The data set has revealed that about a quarter of the queries incur empty result sets. We therefore analysed users’ reactions upon encountering these situations. In 23% of the subject search sessions that start with an empty result query, users immediately end the search session. For the remaining search sessions, the application of the tactics presented in Figure 8 is examined. Users most frequently apply the following tactics to overcome empty result lists, ordered by decreasing number of appearance in the data set: delete a facet, correct spelling mistakes, change the spelling, replace a term by a related term, use register terms from the keyword register or replace a term by a broader term. The users thus apply adequate tactics to overcome queries with zero results.

To enlarge result sets with few results, the most common tactics are the replacement of a query term by a related term and the deletion of a facet.

If a query delivers only few results, it may be interpreted as either successful or unsuccessful, depending on the user’s retrieval goal: few hits may testify a high precision value but in case more relevant documents exist in the corpus, few hits may also indicate a low recall value. While the relevance of the result documents cannot be immediately deduced from the query logfiles, the user’s possible retrieval goal can be inferred from the number of inspected documents. If none of the few results is inspected, the query can be classified as unsuccessful. If at least one of the few results is viewed, this raises the probability of a successful precision-oriented search. In the GEI logfiles, search sessions with an initial query with few results (1–3 hits) are immediately quit in 44% of the cases but more than half of these searches can be interpreted as possibly successful precision-oriented searches at least one document is viewed. Nevertheless, query logfiles do not allow for assumptions about the relevance of the viewed documents. Inspection of a document can only be interpreted as an indicator of its possible relevance.

**How users react to large result sets in subject searches**

If users receive large result sets (more than 100 hits) in the initial subject-search query of a search session, more than half of these search sessions (65%) are immediately ended, either without having viewed any document (30%) or after having viewed at least one result document (35%).

In the search sessions with large result sets that are not immediately ended, different tactics are employed to reduce the result set in the first reformulation step. The most common tactic for reducing result lists consists in adding a new search facet. Furthermore, the replacement of terms by related terms is common, and narrower terms are also employed for this purpose, although less often. Another frequent tactic is
the change of search fields, for example, from the free text to the keyword field. Still, in many of these search sessions where initial queries with large result sets are reformulated (53%), no document is viewed.

**Interpretation of results**

Generally speaking, the majority of the search steps in the GEI consist of subject searches, but known item searches are also frequent, a high proportion of these relating to searches for specific documents or searches by author name. This can be explained by the GEI’s primary role as a bibliographic reference database: It is obviously often consulted to look for specific documents in order to assess their availability.

The study has shown that the advanced search mask is generally preferred by the users. Nevertheless, it is often used in a simple search mode, many searches in the advanced search mask only deploying the free text search field. This field is also frequently used for known item searches although in these cases, users sometimes also adequately employ specific search fields like the author and title fields.

In general, the GEI users do not often employ further advanced search functionalities like phrasing (for example for title searches), truncations or register terms. The users’ behaviour supports the findings by Wildemuth and Moore who state that retrieval effectiveness could be improved by fostering these functionalities. Even in DL-like environments, the users thus do not seem to fully exploit the potential of advanced search functionalities, as noted by Spink et al. with reference to a web search engine.

The fact that the advanced search mask is employed for a high fraction of the queries may be due to the GEP’s pre-configuration which offers the advanced mask as the default option. This assumption would be in line with the findings by Jones et al. who state that users tend to maintain search pre-configurations.

This could also explain why the pre-configuration of Boolean operators is rarely changed in the GEI’s advanced search mask. The analysis of term tactics revealed that query terms are hardly ever combined by the OR-operator. Instead, queries are generally reformulated by replacing terms, most frequently by related terms, which were identified as the most common term tactic by Rieh and Xie. Nevertheless, use of the OR-operator to combine original query terms with semantically related terms would possibly be more effective than the replacement of original query terms. Especially in queries that comprise several facets, the expansion of a facet with a related term may be more effective for increasing recall than its replacement with a related term. At present, however, the expansion of facets is difficult to define in the GEI, which may explain their scarce occurrence in the logfiles, as well as the non-use of the building blocks strategy in this data set.

Sophisticated search strategies are thus seldom and quick approaches account for most of the search steps. These single-term queries often deliver high number of results, which have been shown to frequently lead immediate terminations of the search sessions. The contrary situation of empty result sets has incurred similar reactions.

Although the users employ adequate tactics for specifying and generalizing their queries, the analysis has shown that the use of narrower terms (for specifying) and broader terms and synonyms (for generalizing) is rare. Instead, users tend to replace query terms by related terms or vary the number of query facets.

**Conclusion**

Based on the above stated results, two main areas for further developments in the GEI can be identified, i.e., the avoidance of empty result sets, as well as the prevention of immediate endings of possibly unsuccessful searches. While the adaptation of the retrieval algorithm and the implementation of further automatic query expansion mechanisms may serve the
former purpose, support in the application of search tactics and strategies may help users to overcome unsuccessful query formulations. A more supportive term suggestion functionality than the registers that are presently implemented and scarcely used might be serviceable. Based on these findings, we may hypothesize that the implementation of query expansion mechanisms may lead to an increase in retrieval effectiveness, as Carstens currently investigates.

The study has revealed search characteristics which are due to the GEI’s role as a bibliographic reference database where known item searches account for a vast amount of the queries. Moreover, the study showed that the advanced search functionalities offered in the GEI are adequately applied by the users. While these results are specific to the GEI, they are comparable to studies in other DL environments that may also be based on the Lucene search engine, apply similar search forms, support Boolean search syntax and offer advanced search functionalities.

References


News

NISO launches E-book Special Interest Group
The National Information Standards Organization (NISO) and its architecture committee are pleased to announce the creation of a Special Interest Group focused on E-books (the NISO E-book SIG). Simultaneous with the formation of the group, NISO is issuing a call for participation in the E-book SIG and its associated monitoring group. The E-book SIG will explore a range of industry best practices and standards related to the creation, distribution, discovery, delivery, and preservation of digital book content.

- The primary responsibilities of the group will be to continuously monitor and review the state of the industry for e-books and to suggest areas for new initiatives within NISO or areas where NISO can engage with other communities on e-book work underway outside of NISO.
- The group will also host thought leader meetings and commission relevant research to advance the state of the industry.
- The SIG will also foster cross-community dialogue in a given topical area and provide a place for the incubation of ideas even if no formal standards process within NISO is ever initiated as a result.

For more information, please see the full details at http://www.dlib.org/dlib/july11/07inbrief.html

Early European Books in Digital Library
The Digital Library has been expanded with Early European Books. This database is the result of a cooperation between ProQuest, the supplier, and different university libraries in Europe. All works printed in Europe before 1701, regardless of language, fall within the scope of the project, together with all pre-1701 works in European languages printed further afield. Early European Books builds upon and complements Early English Books Online. Early European Books offers full-colour, high-resolution (400 ppi) facsimile images scanned directly from the original printed sources. Each item in the collection is captured in its entirety, complete with its binding, edges, endpapers, blank pages, and any loose inserts, providing scholars with a wealth of information about the physical characteristics and provenance histories of the original artifacts.

Source http://www.library.leiden.edu

Google scholar citation (GSC) tool
Google formally launched GSC to provide a simple way for scholars to keep track of citations to their articles. Citing the fact that this represents “a new direction us,” GSC is currently in limited launch with a small number of users, although some researchers have been able to create profiles in the past weeks. The system is still in the early stages, but is clearly an essential citation tool for the future. Any research organization needs to follow this development closely—and researchers with significant publishing records should get in line to create their own profiles in what promises to be a key citation research tool of the 21st century.

For more information, please log onto http://scholar.google.com/intl/en/scholar/citations.html
Indian Academy of Sciences (IAS) repository

The Indian Academy of Sciences was founded with the aim to promote the progress and uphold the cause of science, both in pure and applied branches. It is estimated that the total number of articles published so far by all fellows in various national and international publications could be around 100,000. IAS has set up an open repository for research publications running on EPrints open archive software. Publications of the IAS fellows repository collects, preserves and disseminates in digital format the research output and other scholarly publications created by the fellows of the Indian Academy of Sciences. While Publications of the IAS fellows can be accessed by anybody, submission of documents to this repository is limited to the fellows of the Academy only. Presently there are over 50,000 articles and documents are already submitted in the repository and is growing very fast.

For more information, please log onto http://repository.ias.ac.in/index.html

Report on International Conference on Digital Library Management (ICDLM) 2011

The International Conference on Digital Library Management (ICDLM) 2011 was held from 11-13 January 2011 at Science City Auditorium, Kolkata. The theme of ICDLM 2011 was “Extending benefits of modern technology to public, academic and special libraries”. The Conference was jointly organized by The Energy and Resources Institute (TERI) and Raja Rammohun Roy Library Foundation (RRRLF). This prestigious event was inaugurated by Prof. Tapan Raychoudhury, Former Director, Delhi School of Economics, Delhi University on 11 January 2011. Dr P Anandan, Managing Director, Microsoft Research India had delivered keynote address on ‘Capturing and Experiencing Cultural Heritage in Digital Form’. The second keynote address was given by Prof N Balakrishnan, Associate Director, Indian Institute of Science.

The conference had 27 invited speakers from 11 countries including India and 375 participants including 244 registered participants which included more than 24 foreign participants from 14 countries from Asia, Africa, Europe, the United States of America, and Australia. A total of 125 papers/abstracts were received. These papers were reviewed by members of the national and international programme committee and a total of 35 papers were short-listed for presentation during the conference.

At the end, a committee of experts had deliberated following key recommendations for larger benefits of Indian libraries with special reference to public library system.

- Reorient libraries for a more customer focused resources and services
- Set up a National Digital Public Library of Indian Culture and Heritage
- Develop a National Policy on Creation of Digital Resources
- Build and strengthen capacity and competencies of public libraries professional staff in ICT applications
- Develop a platform to provide online access to integrated catalogue of large, medium and flagship public libraries
- Enable public libraries to manage their functions by using LMS hosted on a central server
- Build and strengthen information infrastructure in public libraries
Guide to authors

*World Digital Libraries* is an international peer-reviewed biannual journal. The journal seeks quality research papers that present original theoretical approaches. It also seeks experimental case studies related to digital library developments, maintenance, and dissemination of digital information focusing on research and integration of knowledge at the interface of resources and development. The journal will, therefore, keep readers abreast with the current developments and contain articles, reviews, current developments, and case studies, encompassing the following areas.

- Theoretical and methodological issues that relate to the interrelationships among electronic resources management, digital preservation, multiple access, multilinguality, copyright issues, and security aspects.
- Theoretical approaches as well as experimental case studies related to digital library development and maintenance.
- Initiatives towards digitization through lucid case studies.
- Current developments across the globe.
- Dialogues between the scientific community and society at large.

Articles should examine concepts, analyses, and case studies of important issues in the field.

Book reviews should be of recent publications in the field, to be reviewed by an independent reviewer.

Commentaries should discuss critical issues in the field.

**Submissions**

Authors are requested to send a soft copy (in Microsoft Word format) of their contribution to the editor, either in a CD or as an e-mail attachment.

All submissions will be peer-reviewed using the criteria of originality, accuracy, and quality of contribution in these fields.

**Presentation of manuscripts**

Articles must be original, in English, and should not exceed 8000 words. The main text should be double-spaced with headings and subheadings clearly indicated in the text. All tables, figures, and equations should be numbered in Arabic numerals and clearly cited in the text. All measurements should be in metric (SI) units. The manuscript should be arranged in the order given below.

- Short title (10 words is the desired maximum length), subtitle (if desired)
- Author’s name, affiliation, full postal address, and e-mail, telephone, and fax numbers (respective affiliations and addresses for co-authors should be clearly indicated)
- Abstract (not exceeding 200 words)
- Main body of the text, suitably divided under headings
- Acknowledgements, if any
- References
- Appendices (each on a separate sheet)
- Tables (each on a separate sheet)
- Figures (each on a separate sheet)

**Shorter items**

The following shorter items are also welcome and must be typed in the same way as major articles.

- Commentaries (research notes and short communications) and case studies (maximum 5000 words)
- Book reviews (maximum 1200 words)

**In-house style: references**

In the text, the surname of the author(s) followed by the year of publication of the reference should be given, for example, (Hall 1993). In case of several publications by the one author or by a group of author(s) in one year, use notations ‘1993a’, ‘1993b’, and so on. Up to three authors can be mentioned in text references; more than three authors should be limited to the first three authors’ names followed by ‘et al’. References must be listed alphabetically at the end of the
paper (double spaced) and should conform to the following style.

For journals
Davis G R. 1990
Energy for planet earth
Scientific American 263(3): 55–62

For books
Carmichael J B and Strzepek K M. 1987
Industrial Water Use and Treatment Practices

For chapters of edited books
Sintak Y. 1992
Models and projections of energy use in the Soviet Union
In International Energy Economics, pp. 1–53
edited by T Steiner

For grey literature
Togeby M and Jacobsen U. 1996
How conflicting goals concerning environment and transport influence the policy process?
Paper presented at the Conference on Transport, Energy and Environment,
3–4 October, Helsingor, Denmark

WBCSD (World Business Council for Sustainable Development) and UNEP (United Nations Environment Programme). 1998
Industry, fresh water, and sustainable development
Details available at <www.gm-unccd.org/FIELD/Private/WBCSD/freshwater.pdf>, last accessed on 9 January 2004

Footnotes
Authors are requested to use as few footnotes as possible, and keep their length to the minimum. Footnotes should be indicated in the text by superior Arabic numerals, which run consecutively through the paper. They should be grouped in order of appearance at the bottom of the concerned page in numerical order and must be double-spaced.

Accepted manuscripts
On acceptance, contributors are requested to provide the editor the final version of the article in soft and hard copy. Please observe the following instructions.
- Tables, figures, illustrations, should be on separate sheets.
- Retain a back-up disc for reference and safety.

Proofs
One set of proofs will be sent to the author before publication, which should be returned promptly within 48 hours of receipt. Authors are urged to check the proofs carefully as late corrections cannot be accepted.

Offprints
Apart from one free copy of the journal to the authors, 10 free offprints will be supplied to the first author. Further offprints and copies of the journal can be purchased at a reasonable cost, if ordered when sending the final copy of the article, or when returning the proofs.

Copyright
The responsibility for the contents of the paper rests with the authors, not with the editor or the publisher. Contributions are accepted for refereeing on the understanding that they have been submitted only to this journal and not to any other journal. Only when each author signs and submits the CTA (Copyright Transfer Agreement) can TERI Press publish the article. This CTA enables TERI Press to protect the copyright material for the authors, but does not affect the authors’ proprietary rights. The CTA covers the exclusive rights to reproduce and distribute the articles, including reprints, photographic reproductions, or any other reproductions of similar nature and translations, and includes the right to adapt the article for use in conjunction with computer systems and programmes, including reproduction or publication in machine-readable form and incorporation into retrieval systems. Authors are responsible for obtaining, from the copyright holder, permission to reproduce any figures for which copyright exists.