World Digital Libraries
An international journal

Volume 3, Issue 2, December 2010

The Energy and Resources Institute
Contents

WDL (World Digital Libraries)
Volume 3, Issue 2 • December 2010

v Editorial

97 Project Naming: what makes it a “good digital collection”? Cobi Falconer

101 Editorial Board of Digital Library Journals: a social network analysis approach Monica Sharma and Shalini R Urs

115 Perceptions of South Asian LIS community towards open source software adoption in libraries Muhammad Rafiq and Kanwal Ameen

125 Agent-based web intelligent tutoring system Rohit Chawla and Priyanka Chawla

133 A survey of digital preservation strategies Feng Luan, Mads Nygård, and Thomas Mestl

151 Global e-library services in Novartis Knowledge Center: meeting worldwide information needs at Novartis Binu Chaudhuri, Leifang He, and Deborah Juterbock

159 Free/open source software for library and information management Rakesh Kumar Gupta and Shailesh Kumar Lohiya

169 Imperativeness of e-learning in a digital environment in an organization R K Bhatt

175 News
The present issue consists of eight articles addressing a variety of issues in the broad area of digital library. Topics covered are of contemporary nature and of very high practical utility. Articles include research interventions on the social networking aspect of digital library, digital library management software, digital library services, digital preservation, knowledge centre, and e-learning.

Muhammad Rafiq and Kanwal Ameen have presented the perceptions of South Asian LIS community towards the open source software for digital libraries. Rakesh K Gupta and Shailesh K Lohiya describe the features and availability of different open source software in the field of digital library. The reasons for the Library and Information Science professionals in India, to prefer to use the open source software for management of digital library over paid software, have been clearly brought out in their paper. Cobi Falconer, in his paper, describes project—Naming—which is a digital collection project for the Library and Archives in Canada.

The importance of building a good digital collection by the National Information Standard Organization has been discussed by the author. Shalini R Urs and Monica Sharma describe the social network of the editorial boards of various DL journals. Rohit Chawla and Priyanka Chawla explore the practical application of artificial intelligence to the next generation of web-empowered systems. Feng Luan, Mads Nygard, and Thomas Mestl have addressed emerging challenges for preservation strategies faced due to the fast obsolescence of the technology.

Binu Chaudhuri, Leifang He, and D Juterbock describe how the new e-model evolved a new role for Novartis Knowledge Center service group and how this service group contributes to the e-content lifecycle management process. It also provides the details on how the service group empowers end users to locate needed resources.

I do hope that the readers will find the issue very interesting and useful.
The following is an assessment of Project Naming, which is both a digital collection (work-in-progress) and a digital project through Library and Archives Canada (LAC). I shall demonstrate how this collection/project exemplifies the high-level principles outlined in A Framework for Guidance for Building Good Digital Collections (2006) by the National Information Standards Organization (NISO) in relation to the digital collection development and selection policies theme outlined in our class syllabus. In brief, Project Naming involves digitizing photographs of the Inuit pre-mid-20th century, which were originally snapped by several non-Inuit professional photographers and various federal government-related personnel such as the National Film Board of Canada, Department of National Health and Welfare, Royal Canadian Mounted Police, and Department of Indian and Northern Affairs (LAC - Photo collections). Many of the Inuits portrayed in the photographs were never personally identified—for example, ‘Eskimo playing drum’ taken at ‘Spence Bay’ has been revised through Project Naming to ‘Iharrataittuq Itirujuk, the father of Nilaulaaq Aglukkaq in Taloyoak (formerly Spence Bay), Nunavut’ (LAC - Richard Harrington collection: Eskimo playing drum). Before phase-I of the project (2001–2004), these photographs sat quietly in the former National Archives of Canada (now LAC). Several of these photographs were scanned and transferred to CD-ROMs. The CD-ROMs were to be taken up north by Inuit youth to be viewed on laptops by their elders. (LAC - Voices from Nunavut: Project Naming: Always on our minds).

The collections principles are part of the Framework of Guidance for Building Good Digital Collections (2004), which is a guide rather than a standard or best practice for digital projects/collections to follow especially when applying for funding from granting organizations like the IMLS\(^1\) (Cole 2002). Although it appears that Project Naming has not received funding from the IMLS or the NSF\(^2\), it is interesting to see how this digital collection/project might have been influenced by these principles (via LAC)\(^3\), which may have contributed to some of the project’s success. In this vein, Project Naming follows collections principle number one in that there is an agreed upon collection development policy created by LAC, which is based on the broad collecting mandate established by the Library and Archives of Canada Act’ (LAC -

---

\(^1\) The IMLS collections principles were referred to in our class in week three—the first edition (2002) (Rasmussen).

\(^2\) I am making this assumption based on the Project Naming website.

\(^3\) An indicator of this influence is exemplified in how LAC refers to their collection development mandate document as a ‘Collection Development Framework’ (LAC - Collection development framework).
Digital collection development policy. Concepts of LAC’s collection mandate include items having ‘heritage value’, materials published/created in Canada, Canadian creator(s), and has Canadian subject matter (LAC - Collection development framework). The photographs digitized through Project Naming fulfill all these concepts in that the photographs are of Canadians, mostly taken by Canadian photographers, and document a period of space and time that many Canadians today may not be aware of.

Project Naming follows collections principle number two in how the collections are described for physical and intellectual accessibility—‘so that a user can discover important characteristics of the collection’ (NISO 2004). Each photograph (after digitization) has an item-level archival description, which is accessible online through the site. These item-level descriptions include: ‘scope, format, restrictions on access, ownership, and any information significant for determining the collection’s authenticity, integrity and interpretation’ (NISO 2004). For example, the new/revised information added through Project Naming in square brackets, as well as a note indicating that the revised information has been gathered through this project.

Project Naming follows collections principle number three in the way it is ‘sustainable over time [and has] continued usability beyond the funded period’ (NISO 2004). On the one hand, digitizing and dissemination of the photographs appear to be based on funding, but the intellectual gaps that are currently being filled will last indefinitely. Funding for the website and project was made possible through the Canadian Memory Fund of the Canadian Culture Online (CCO) Department of Canadian Heritage, Department of Culture, Language, Elders and Youth (CLEY), and the Government of Nunavut (LAC - Acknowledgments). It is not clear how long the funding will last or if more donors will get involved, but the project’s sustainability is represented in how each digitized item has its own description, which appears to have created a dialogue (awareness) all over Canada. According to the website, ‘the naming continues’, and new features are being added such as new searchable images, educational resources, and various webcasts (LAC - Introduction: What’s new in phase II?).

Project Naming follows collections principle number four in that the collection on its website ‘is broadly available and avoids unnecessary impediments to use’ (NISO 2004). For example, information available on the website can be read in English, French as well as in Inuktitut. On the one hand, it can be said that the technology to view the photographs is not necessarily accessible to the elderly and to those who inhabit areas where computers are not commonplace. On the other hand, having the youth transport several photographs to a very remote region using laptops demonstrates ‘adaptability’ and ‘effectiveness’ because the photographs can be viewed quite easily in either a communal or intimate setting. The youth are also able to communicate at a language and cultural level with their elders using a device (laptop) that is communicable to Western culture. In this way, a youth (with a laptop) may be viewed as mediating between various binaries of north/south, past/present, non-Western/Western, old/young, memory/medium and isolation/inclusion.

Project Naming follows collections principle number five in the way LAC ‘respects intellectual property rights’ (NISO 2004). This is demonstrated in detail within each of the item-level descriptions, which contains information about property rights that existed before the items were digitized. The creators (either name

---

4 During this period, it may also be argued that the Inuit were not Canadians per se, but rather wards of the state (based on personal knowledge).

5 This awareness is generated through TV documentaries, publications, and conferences.
of the photographer or government agency) are listed under every photograph. Copyright information and terms of usage are also mentioned.6

Project Naming follows collections principle number six of providing (some form of) ‘measurement of use’ to make statistical comparisons (NISO 2004). I am not sure if the website tracks usage statistics. However, what appears to be more important is that the project itself can be measured on the basis of how many photographs have been scanned and how many have been identified. Even if some people’s names are still unknown, a photograph is still ‘updated’ to the best knowledge of its participants by stating its possible location (using indigenous geographical names) and detailing what type of activity is taking place in the photograph. For example: ‘Three Inuit girls in front of a tent’ (former title) was revised to ‘Enookoolook...Pond Inlet’ (LAC - Photo collections).

Project Naming also follows collections principle number seven in that it ‘fits into the larger context of significant related national and international digital library initiatives’ (NISO 2004). Part of the digital initiatives at LAC is to be collaborative and provide outreach locally (Inuit), nationally (Canada), and internationally through online access (LAC - Digital collection development policy). This digital collection appears to be useful to a broad range of disciplines, including Canadian history, photography, anthropology, psychology, and political science, among others.

Questions raised in The Digital Library Toolkit (2003) article that relates to Project Naming include: whether there is a need for a digital library (SunMicrosystems 2003, 7) and is it worth digitizing archival materials (SunMicrosystems 2003, 16). In the case of the (archival) photograph collection sitting stagnant at LAC, the answer to these two questions is a resounding ‘Yes!’ The need may be viewed as imperative because the information needed to ‘update’ these photographs is extremely time sensitive in the sense that elders are passing away. Perhaps, the only way to view a large amount of materials in a way that would protect the pictures from ‘floating around’ is to keep them on a CD-ROM rather than in a non-digital form. On the other hand (for sake of discussion), I wonder if the Inuit, particularly the elders, are aware that these photographs are available online and can be viewed by anyone across the globe. I would also like to know whether it is possible for remote communities to ‘access their albums’ online from where they live because I myself find it difficult to log on to the site from my own computer (takes several minutes). This delay may be due to the images being uploaded onto the site. Another query I have is whether the additional contextual information will ever be added to the archival descriptions similar to a colophon. For example, the interview process, where it took place, how the person identified the individual (issues surrounding reliability and authenticity), and so on. While the project appears to be sensitive to Inuit needs first and foremost, I also like the way it still raises several questions which, needless to say, are difficult to avoid through digitization programmes.7 On the other hand, the project’s attention to detail in being a ‘good digital collection’ is undeniable, and is clearly demonstrated in mirroring the digital collections principles outlined in the framework.

---

6 I find this issue of property rights interesting because the photographs are not returned. I wonder whether Inuits can take back a photograph of themselves if they so wished because of the way they were portrayed in the photographs and treated by the government during the time period represented in the photograph.

7 For example, I am interested in whether the website itself is constrained by their funding agencies in that the pictures of Inuit displayed on the site do not make the government’s actions look questionable.
References


Library and Archives Canada. 2007. **Homepage**. *Project Naming*. Available at: http://www.collectionscanada.ca/inuit/index-e.html. (Last accessed on 1 October 2007).


Library and Archives Canada. 2007. **Photo collections**. *Project Naming*. Available at: http://www.collectionscanada.ca/inuit/020018-1030-e.html (Last accessed on 1 October 2007).


Rasmussen E. 2007. **Digital Libraries – Week 3**. PowerPoint presentation at SLAIS, UBC.

Editorial Board of Digital Library Journals: a social network analysis approach

Monica Sharma
Department of Library and Information Science
University of Mysore
monirsharma@yahoo.com

Shalini R Urs*
International School of Information Management
University of Mysore
shalini@isim.ac.in

World Digital Libraries 3(2): 101–114

Abstract

We present here our findings of a study, part of larger research focusing on understanding the structure and composition of the academic community of Digital Library (DL). Using social network analysis, the present study analyses the structural features of editorial boards (EBs) of leading journals of DL. Following the premise that boundaries and directions of a field, especially an evolving one such as digital libraries is shaped by journals and their editors, we undertook a study of the editorial boards of the top 56 journals in the field of digital libraries. Techniques from social network analysis are applied to the data to reveal the structures and relationships within the EBs and journals of the DL community. The paper reveals the structure of the scientific grouping as well as the key players in the system and finds the ‘diversified convergence’ of fields. The focus is on the stars of the DL community as well as journals. We have compared the results with our previous studies carried out on different communication platforms of the DL world. Our results show that while computer science is the common thread, library and information science is the field that dominates, both in the scheme of things of top-ranking journals as well as editors.

* Part of the work accepted as a poster paper at 11th International Conference on Asia-Pacific Digital Libraries, ICADL’2008, 2-5 December, Bali, Indonesia
Introduction

Preamble
In the making and shaping of a science, academic publications—especially peer-reviewed journals—play a very important role. Research efforts have consistently shown that it is possible to unravel various structural features of research communities by studying the network structures of their communication systems such as journals. The landscape and the direction of research are molded by two key players—the editors and the authors in these journals. Their sphere of influence is quite enormous. Our research is in the direction of understanding the centrifugal forces in the shaping of the field of digital libraries using Social Network Analysis (SNA) approaches and tools. In this paper, we present our research study of the structural features of editorial boards (EBs) of leading journals on digital libraries (DL). Based on the premise that the community structure of the EBs reflects the knowledge structure of the field, we draw inferences and conclusions on the dynamics of the DLs.

Science of networks and networks of science
A new science of networks has emerged in the recent past. It is a multi-disciplinary academic pursuit with interesting possibilities. Disciplines such as sociology, applied mathematics, physics, and computer science have contributed tremendously to the science of networks (Newman 2003). The phrase ‘Six Degrees of Separation’ coined by Stanley Milgram in the early empirical study of ‘Small World Problem’ (Milgram 1967) was immortalized by John Guare in a 1990 play. However, it is perhaps best known for the popular game ‘Six Degrees of Kevin Bacon’ (Oracle of Bacon). The international best seller, The Tipping Point by Malcolm Gladwell has turned the concept of impact of social networks into folklore (Gladwell 2003). A sense of excitement surrounds this fast developing field, with new papers appearing almost daily and an unprecedented degree of integration across many disciplines (Watts 2004).

The network of science refers to the collection and connections of scientists and scientific organizations. In order to understand the dynamics of scientific research, it is important to examine the formal and informal communication networks of researchers. As Kuhn notes that a paradigm transforms a group into a profession or, at least, a discipline and the paradigm guides the whole group’s research, and it is this criterion that most clearly proclaims a field as science (Kuhn 1970). ‘Invisible Colleges’—a term coined by Price in 1961 to describe informal networks of scientific specialists—have been found to influence the growth of specialties (Gresham 1994). Today, the use of SNA for the study of academic networks has lent a fresh dimension to the exploration of research trends and paradigms.

Given that journals play a critical role in the structure of science, the influence of the two key players—the editors and authors—in the sociology of science is fairly significant. In a competitive academic environment, scholars pay attention to the quality and quantity of their published research and the number of citations to their work. According to Faria, editors enjoy market power and they play the game as leaders in shaping the structure of any discipline (Faria 2000). While editors play the role of leaders, authors are the followers. By directing the editorial processes and policies of journals, the EB of journals shapes the community and the discipline.

SNA has emerged as a key technique in information sciences, organizational studies, economics, as well as a popular topic of speculation and study. SNA could be defined as the mapping and measuring of relationships and flows between people, groups, organizations, computers, and other information processing entities. The nodes or actors in the network are the people and groups while the links show
relationships or flows between the nodes. SNA provides, both a visual and a mathematical analysis of human relationships (Krebs 2007).

The aim of this paper is to explore the structural features of the EBs of the major journals in the field of DL, with a view to analyse the shape, network of journals, key editors, and other such features of the network.

Submissions to prestigious journals are channeled through editors, associate editors, and boards of editors in various forms for the purpose of evaluation. Editorial positions exert an impressive degree of power over the discipline, both in terms of the dissemination of information and in the success or failure of colleagues. Kaufman argues that the board of editors has some power in shaping the editorial processes and policies of scientific journals. Because of their importance, it appears reasonable that the positions on the boards are held by "persons who have the confidence and trust of their colleagues in the journal’s areas of coverage for the journal to be successful in attracting quality submissions and in building and maintaining a reputation for quality. Thus, selection as an editor or member of an editorial board is a considerable honour that reflects one’s standing in the profession as evaluated by his or her peers" (Kaufman 1984). Unsurprisingly, members of the editorial board usually place this information in evidence in their curriculum vitae; and it is universally recognized that to be a member of an editorial board is a signal of the esteem reserved to a scholar by the academic community. Though this paper does not present extensive discussion on the importance of editors and information exchange between editors, it is understood from the above discussions and research work that editors play an important role in shaping the policies of the journals. This paper is an attempt to study the first level analysis of the structure of EBs and journals, without any focus on the kind of information exchanged between the editors.

### Background work and need of the present work

Despite its relative youth, DL has emerged as a discipline and an area of research in its own right. DL represent the meeting point of a large number of technical areas within the field of informatics and several other disciplines and fields beyond informatics, such as library sciences, museum sciences, archives, sociology, psychology, knowledge management, and so on. The interdisciplinary nature of this field has drawn the attention of information scientists and scientists from other fields such as computer science, and so on (DELOS 2007; Fox and Urs 2002).

In the past, studies have been carried out by researchers to understand the composition and structural features of the DL community as expressed in different communication platforms such as journals and conferences (Sharma and Urs 2007); journals (Sharma and Urs 2008); online database, that is, Citeseer (Sharma and Urs 2008); and social networking site namely LinkedIn (Sharma and Urs 2008). Baccini, Barabesi, and Marcheselli had studied the network characteristics of editors of statistics journals (Baccini, Barabesi, and Marcheselli 2009). Their study is based on the hypothesis that each editor possesses some power in the definition of the editorial policy of his or her journal. Malin and Carley (2007) carried out social network analysis of the EBs of bioinformatics and medical informatics journal to study how they integrate researchers from disparate communities (Malin and Carley 2007).

The present study is an attempt to expand our previous studies to understand the network dynamics of EBs of top-ranked journals in the domain of DL, following the premise that editors and authors are the leaders in shaping any academic community. We attempt to find the nature of ‘Diversified Convergence’ of DL as a discipline and its most central journals and editors in the field. The results of the study were
compared with our earlier studies in the domain of DL.

**Data sets and research methods**

The key factor in any SNA is to first identify the group and define its membership. We first had to identify the group of journals that justifiably represent the field and then define the selection process. We have used the Thomson Reuters’s Web of Knowledge (Thomson Reuters 2008) database for the identification of journals and selected the journals on the basis of publication count. We explain the process in detail in the following sections. We believe that given that Web of Knowledge covers the most significant literature in the fields of science and technology, the datasets would be not only representative, but also the ones to ‘direct’ the field.

**Selection of journals**

We did a topic search on DLs and retrieved 7770 papers in the field of science and technology from the Web of Knowledge. Keeping the record count threshold at 5, this list was pruned to include the top 250 journals and conferences. Further, these titles were segregated into journals and conferences, resulting in 143 journals after removing repetitions. Journals, which did not have any information on EB or did not have a website were also removed. Though D-Lib journal topped the 143 journals list, it was not among the 56 identified journals, as it does not have an EB. Finally, 56 journals with a record count of 7 or more were identified and this formed the dataset for the study of the network of editors. We have completed the macro-level analysis using the 143 journals and the SNA of the editor network of only 56 journals.

**DL journals and EB analysis**

Information on EB members of these journals was collected manually by visiting individual journal websites. A total of 1585 names were recorded. In some instances, names appeared more than once within a journal, which were also removed. There are a total of 1453 distinct editors from these 56 journals for which adjacency matrix was created using a Java programme. The 56 journals were given acronyms (Table 1) and then converted into adjacency matrix, which is acceptable by SNA software. An adjacency matrix of 56 x 56 journals was created, and subsequently, their network was created using SNA software.

To carry out the analysis and visualization of the DL journals network, UCINET and Netdraw (Borgatti, Everett, and Freeman 2002) were used. The distinct names of editors were converted into an adjacency matrix of 1453 X 56, wherein 1453 are editors and 56 are journals.

---

**Table 1** Journals listed by acronyms provided

<table>
<thead>
<tr>
<th>Record count</th>
<th>Journals (JL2)</th>
<th>Acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>127</td>
<td>Electronic Library</td>
<td>EL</td>
</tr>
<tr>
<td>94</td>
<td>Program-Electronic Library and Information Systems</td>
<td>PRG</td>
</tr>
<tr>
<td>91</td>
<td>Journal of the American Society for Information Science and Technology</td>
<td>JASIST</td>
</tr>
<tr>
<td>88</td>
<td>Library Hi Tech</td>
<td>LHT</td>
</tr>
<tr>
<td>83</td>
<td>OCLC Systems and Services</td>
<td>OCLC</td>
</tr>
<tr>
<td>79</td>
<td>Online Information Review</td>
<td>OIR</td>
</tr>
<tr>
<td>63</td>
<td>Journal of Library Administration</td>
<td>JLA</td>
</tr>
</tbody>
</table>

Contd...
<table>
<thead>
<tr>
<th>Record count</th>
<th>Journals (JL2)</th>
<th>Acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>Information Processing and Management</td>
<td>IPM</td>
</tr>
<tr>
<td>51</td>
<td>Science and Technology Libraries</td>
<td>STL</td>
</tr>
<tr>
<td>49</td>
<td>International Journal on Digital Libraries</td>
<td>JODL</td>
</tr>
<tr>
<td>47</td>
<td>Serials Librarian</td>
<td>SL</td>
</tr>
<tr>
<td>45</td>
<td>Library Review</td>
<td>LR</td>
</tr>
<tr>
<td>36</td>
<td>Interlending and Document Supply</td>
<td>IDS</td>
</tr>
<tr>
<td>33</td>
<td>Journal of Documentation</td>
<td>JD</td>
</tr>
<tr>
<td>31</td>
<td>First Monday</td>
<td>FM</td>
</tr>
<tr>
<td>29</td>
<td>Jodi—Journal of Digital Information</td>
<td>JODI</td>
</tr>
<tr>
<td>27</td>
<td>VINE</td>
<td>VINE</td>
</tr>
<tr>
<td>27</td>
<td>Library Collections Acquisitions and Technical Services</td>
<td>LCATS</td>
</tr>
<tr>
<td>24</td>
<td>Internet Reference Services Quarterly</td>
<td>IRSQ</td>
</tr>
<tr>
<td>24</td>
<td>Libri</td>
<td>LIBRI</td>
</tr>
<tr>
<td>24</td>
<td>New Review of Information Networking</td>
<td>NRIN</td>
</tr>
<tr>
<td>23</td>
<td>Journal of Academic Librarianship</td>
<td>JAL</td>
</tr>
<tr>
<td>22</td>
<td>Liber Quarterly</td>
<td>LQ</td>
</tr>
<tr>
<td>22</td>
<td>Multimedia Tools and Applications</td>
<td>MTA</td>
</tr>
<tr>
<td>21</td>
<td>Information Research—An International Electronic Journal</td>
<td>IR</td>
</tr>
<tr>
<td>21</td>
<td>Journal of Zhejiang University (Science)</td>
<td>JZU</td>
</tr>
<tr>
<td>20</td>
<td>Information Services and Use</td>
<td>ISU</td>
</tr>
<tr>
<td>20</td>
<td>Library Hi Tech News</td>
<td>LHTN</td>
</tr>
<tr>
<td>20</td>
<td>Reference Librarian</td>
<td>RL</td>
</tr>
<tr>
<td>19</td>
<td>Health Information and Libraries Journal</td>
<td>HILJ</td>
</tr>
<tr>
<td>17</td>
<td>Reference Services Review</td>
<td>RSR</td>
</tr>
<tr>
<td>16</td>
<td>Library Management</td>
<td>LM</td>
</tr>
<tr>
<td>15</td>
<td>Literary and Linguistic Computing</td>
<td>LLC</td>
</tr>
<tr>
<td>14</td>
<td>Technical Services Quarterly</td>
<td>TSQ</td>
</tr>
<tr>
<td>12</td>
<td>Journal of Information Science</td>
<td>JIS</td>
</tr>
<tr>
<td>11</td>
<td>Collection Building</td>
<td>CB</td>
</tr>
<tr>
<td>11</td>
<td>IEEE Transactions on Knowledge and Data Engineering</td>
<td>KDE</td>
</tr>
<tr>
<td>11</td>
<td>Issues in Science and Technology Librarianship</td>
<td>ISTL</td>
</tr>
<tr>
<td>11</td>
<td>Journal of Electronic Resources in Medical Libraries</td>
<td>JERM</td>
</tr>
<tr>
<td>11</td>
<td>Reference and User Services Quarterly</td>
<td>RUSQ</td>
</tr>
<tr>
<td>9</td>
<td>Library and Information Science Research</td>
<td>LISR</td>
</tr>
<tr>
<td></td>
<td><em>Contd...</em></td>
<td></td>
</tr>
</tbody>
</table>

World Digital Libraries 3(2): 101–114
### Table 2: Subject labels of journals

<table>
<thead>
<tr>
<th>Subjects</th>
<th>JL1</th>
<th>JL2</th>
<th>Acronyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer science</td>
<td>143</td>
<td>56</td>
<td>MRSQ</td>
</tr>
<tr>
<td>Information science and library science</td>
<td>36</td>
<td>21</td>
<td>NLW</td>
</tr>
<tr>
<td>Education and educational research</td>
<td>30</td>
<td>19</td>
<td>WTC</td>
</tr>
<tr>
<td>Engineering</td>
<td>13</td>
<td>6</td>
<td>ACM</td>
</tr>
<tr>
<td>Arts and human</td>
<td>11</td>
<td>9</td>
<td>ID</td>
</tr>
<tr>
<td>Government and law</td>
<td>9</td>
<td>5</td>
<td>JS</td>
</tr>
<tr>
<td>Healthcare sciences and services</td>
<td>9</td>
<td>5</td>
<td>JAMIA</td>
</tr>
<tr>
<td>Operations research and management science</td>
<td>9</td>
<td>6</td>
<td>BSSL</td>
</tr>
<tr>
<td>Social issues</td>
<td>8</td>
<td>3</td>
<td>CH</td>
</tr>
<tr>
<td>Psychology</td>
<td>7</td>
<td>3</td>
<td>GIQ</td>
</tr>
<tr>
<td>Telecommunication</td>
<td>6</td>
<td>3</td>
<td>IEEE-CM</td>
</tr>
<tr>
<td>Behavioural sciences</td>
<td>5</td>
<td>2</td>
<td>INSPEL</td>
</tr>
<tr>
<td>Medical informatics</td>
<td>5</td>
<td>3</td>
<td>JDIM</td>
</tr>
<tr>
<td>Medical informatics</td>
<td>5</td>
<td>3</td>
<td>JNCA</td>
</tr>
<tr>
<td>Social sciences—other topics</td>
<td>2</td>
<td>1</td>
<td>MS</td>
</tr>
<tr>
<td>Science and technology—other topic</td>
<td>1</td>
<td>1</td>
<td>General and internal medicine</td>
</tr>
</tbody>
</table>
Data analysis

**DL journals network and analysis**

All the 143 journals (JL1) and the 56 journals (JL2) were categorized under various subject headings as given by the Web of Knowledge. These journals were presented under 26 subject headings. As can be observed from Table 2, there are many overlaps as most of the journals have been assigned more than one subject.

It can be observed that all journals have been classified under Computer Science. Of the rest, Information Science and Library Science has the highest number of journals, followed closely by Education and Educational Research. The DL discipline has converged from various other subjects varying from Social Issues, Law, and Administration to Health Care and Geography. Study on the DL group of LinkedIn has indicated the dominance of members from the Computer Science, Library and Information Science, and Libraries domains (Sharma and Urs 2008). Thus, there is strong evidence to show the diversified convergence of DL from various disciplines confirming the multi-disciplinary nature of this domain.

**Journals network: components**

Figure 1 shows the network of 56 journals connected with each other if the editors are common to a pair. Out of 56 journals,
10 journals were found to be isolates. EB members of these journals are not members of any other journal’s board. The number is high as compared to statistical journals (Baccini, Barabesi, and Marcheselli 2009) in which only 4 journals have degree 0. The journals having degree 0 do not have editorial relationship with other journals so their editorial policies can be considered independent of each other.

Centrality of journals
Centrality measure of the journals is a very powerful metric to identify the most central and powerful journals. Baccini and his colleagues report that the degree centrality of statistics journals varies between 0–35 values. Only one journal has direct connection with 35 other journals. Maximum journals with the frequency of 9 fall in the degree of 4. Our results show that the highest degree centrality is 15, which is low compared to statistical journals (Figure 2 a). Library Management and Library Review have highest degree of centrality (15), followed by Online Information Review (13).

The journal with highest “betweenness” acts as a gatekeeper controlling the interaction between the connected journals. Even though Library Management and Library Review have highest degree, but when it comes to keeping the most central position in the network and acting as a bridge, it is Information Processing Management that emerges as the subject area with the highest centrality. Library Management and Library Review follow Information Processing Management. Table 3 (Figure 2 [a] and [b]) shows the top ten ranked journals according to degree and “betweenness” centrality.

Valued network analysis
The simple journal network shows the links between the journals irrespective of the number of editors common to a pair. To find the strength of these links, a valued network of these journals was created wherein the strength of the ties is the number of editors common to a pair of journals, and the results were compared with the statistical journals.

Table 4 shows the frequency distribution of journals based on the line value. The results exhibit power law with majority of journals, that is, 83 having line value of 1, which means that only one editor is common to a pair of journals. The maximum numbers of editors common to a journal are only 6 in case of Journal of the American Society for Information Science and Technology and Information Processing.
Editorial Board of Digital Library Journals: a social network analysis approach

The coverage percentage of editors who are members of more than one journal’s EB is 7.22. Gary E Gorman has the highest degree, as he is the EB member of eight journals. There are two editors who have a degree score of 4, twelve editors with the degree score of 3, and 90 editors who are members of 2 journals (Table 6).

There are 1585 scholars on the EB of the given number of journals and the average number of places occupied by an editor is 1.09. The average

**Table 3** Centrality score of top ten journals

<table>
<thead>
<tr>
<th>Journals</th>
<th>Degree</th>
<th>Journals</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>15</td>
<td>IPM</td>
<td>259.76</td>
</tr>
<tr>
<td>LM</td>
<td>15</td>
<td>LR</td>
<td>162.94</td>
</tr>
<tr>
<td>OIR</td>
<td>13</td>
<td>LM</td>
<td>162.38</td>
</tr>
<tr>
<td>IPM</td>
<td>13</td>
<td>IR</td>
<td>147.61</td>
</tr>
<tr>
<td>IR</td>
<td>11</td>
<td>LISR</td>
<td>98.77</td>
</tr>
<tr>
<td>SL</td>
<td>11</td>
<td>SL</td>
<td>93.37</td>
</tr>
<tr>
<td>LHT</td>
<td>10</td>
<td>EL</td>
<td>90.71</td>
</tr>
<tr>
<td>ID</td>
<td>9</td>
<td>MTA</td>
<td>88.83</td>
</tr>
<tr>
<td>CB</td>
<td>8</td>
<td>OIR</td>
<td>88.14</td>
</tr>
<tr>
<td>LCATS</td>
<td>8</td>
<td>Jodi</td>
<td>86.45</td>
</tr>
</tbody>
</table>

**Table 4** Line values frequency

<table>
<thead>
<tr>
<th>Line value</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

6 years with a coverage percentage of 8.67–10.90. These line values are very low compared to the study by Baccini and his colleagues. They report that one pair of journals has the line value as high as 83, that is, 83 editors are common to the EBs of a pair of journals.

**Editorial board analysis**

The second phase of the present study involves evaluating the EBs of these journals of DL. There are, in all, 105 editors who are on more than one EB. The results are almost similar to a study by Malin and Carley (2007). They report that the number of editors who are members of more than one editorial board varied from 96–136 in Management (Table 5). *Journal of Digital Information* and *International Journal on Digital Libraries* have 5 editors in common. These line values are very low compared to the study by Baccini and his colleagues. They report that one pair of journals has the line value as high as 83, that is, 83 editors are common to the EBs of a pair of journals.

**Table 5** Line values of journals

<table>
<thead>
<tr>
<th>Line value &gt;2</th>
<th>Journal pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>IPM JASIST</td>
</tr>
<tr>
<td>5</td>
<td>JODI JODL</td>
</tr>
<tr>
<td>4</td>
<td>JODI JASIST</td>
</tr>
<tr>
<td>4</td>
<td>JODL IPM</td>
</tr>
<tr>
<td>4</td>
<td>MRSQ JERML</td>
</tr>
<tr>
<td>3</td>
<td>OCLC LHT</td>
</tr>
<tr>
<td>3</td>
<td>LCATS SL</td>
</tr>
<tr>
<td>3</td>
<td>LCATS CB</td>
</tr>
<tr>
<td>3</td>
<td>IR IPM</td>
</tr>
<tr>
<td>3</td>
<td>JD IPM</td>
</tr>
<tr>
<td>3</td>
<td>JODL JASIST</td>
</tr>
<tr>
<td>3</td>
<td>IR JASIST</td>
</tr>
<tr>
<td>3</td>
<td>LISR OIR</td>
</tr>
<tr>
<td>3</td>
<td>EL OIR</td>
</tr>
</tbody>
</table>

**Table 6** Number of journals common to an editor

<table>
<thead>
<tr>
<th>Degree of editors</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>1</td>
<td>1348</td>
</tr>
</tbody>
</table>
number of places available per journal is 28.3. Baccini reports that the average number of seats per journal is 35.5, while the average number of seats per editor is 1.3, both the values are higher compared to ours.

Editor’s network: components
The network formed 11 components. Of these, the giant component constituted 84% of the editors. It shows that the editor network is very well-connected (Figure 3). There are some editors, who are part of the EB of more than 2 journals and these are the ones who have kept the network connected.

The results are similar to the findings made from earlier studies on co-authorship. The giant component of the author network of journals and conferences constituted 94% of authors indicating that DL is a well-connected world. Similar results were found in the DL community of LinkedIn site (94.1%). However, we observed that the DL community of journals, that is, D-Lib and International Journal on Digital Libraries (JODL) and an online database, that is, Citeseer exhibited different results forming a fragmented world. Study by Liu, Bollen, Nelson, and Sompel (2005) on the DL community of the JCDL conference report that the largest component has only 38% of all authors, indicating that there may be limited international collaboration.

Editors’ centrality
Among editors, Gary F Gorman has the highest degree centrality (Table 7), followed by Thomas Nisonger and Gary Marchionini. Table 8 shows ranking of editors based on all the three measures of centrality.

Gary Gorman is top ranked in all the three centrality measures indicating that he is the most powerful and central member of the network. Also, it can be observed that Gary Gorman, Edward Fox, and Gary Marchionini are the three editors who are the core members of the network, thus, playing a very important role in shaping policies of the journals, and

<table>
<thead>
<tr>
<th>Authors</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gary E Gorman</td>
<td>8</td>
</tr>
<tr>
<td>Thomas E Nisonger</td>
<td>4</td>
</tr>
<tr>
<td>Gary Marchionini</td>
<td>4</td>
</tr>
<tr>
<td>Alan Gilchrist</td>
<td>3</td>
</tr>
<tr>
<td>Amanda Spink</td>
<td>3</td>
</tr>
<tr>
<td>Andrew Dillon</td>
<td>3</td>
</tr>
<tr>
<td>Charles Oppenheim</td>
<td>3</td>
</tr>
<tr>
<td>Derek Law</td>
<td>3</td>
</tr>
<tr>
<td>Edward A Fox</td>
<td>3</td>
</tr>
<tr>
<td>Jennifer Rowley</td>
<td>3</td>
</tr>
<tr>
<td>Jim Jansen</td>
<td>3</td>
</tr>
<tr>
<td>Mike Thelwall</td>
<td>3</td>
</tr>
<tr>
<td>Niels Ole Pors</td>
<td>3</td>
</tr>
<tr>
<td>Ricardo Baeza-Yates</td>
<td>3</td>
</tr>
<tr>
<td>Christine L Borgman</td>
<td>3</td>
</tr>
</tbody>
</table>
consequently, the DL communities. Our earlier studies also show Edward Fox as the star of the DL world.

Geographical concentration of editors
To study the geographical distribution of these editors, an attribute file was created wherein the attribute of these editors was the country to which they belong. For the present study, we could get the country data of only 95% of the editors (ND stands for no data, which constitutes 5% of the total).

There are, in all, 68 countries to which the editors are affiliated. As can be observed from Table 9, 682 editors, with known institutional locations, come from US institutions—no less than 47% of the total (Figure 4). This is followed by UK, constituting 11%. There are 51 countries each of which constitutes less that 1%.

In case of the geographical distribution of editors, we could find similarity with the earlier studies. In LinkedIn, the DL group is centered around the US, constituting 70% of the members (Sharma and Urs 2008). Liu, Bollen, Nelson and Sompel (2005) report that approximately 72% of the authors are affiliated with US institutions. These studies exhibit similar patterns in geographical concentration of DL community with the US as a dominating country.

Conclusion
Editors are the invisible leaders in structuring and shaping the growth of any domain. The intellectual lineage of an evolving domain is normally diverse; it also changes course in terms of the centrifugal forces that shape it. The present study is an attempt to understand two important issues—diversified convergence or the structure
of the multi-disciplinarity of the DL domain and the dynamics of the editor network and its characteristics. Our study shows that journals come from a wide range of subjects, the majority being on Information Science and Library Science as well as Education and Educational Research. Not surprisingly, all the journals are categorized under Computer Science. The other roots from which this new domain is formed are subjects such as law, health care, history, and geography, to name a few. This shows the character of the diversified convergence of the DL domain.

We have also attempted to study the journal network and compared results with the study carried out by Baccini and his colleagues (2009) on statistical journals. The statistics domain is a very well-connected network wherein the network has only four isolates. The value line wherein the value of the ties of the network is the number of editors common to a pair of journals, varies from 1 to as high as 83; whereas in DL values ranged from 1–6. The difference in the results of the two studies could be because of the number of journals considered for the study, which is comparatively less. The important reason could be the nature of the two disciplines. DL is a highly multi-disciplinary domain as discussed earlier. Centrality of journal shows that even though *Library Review* and *Library Management* are the top ranked journals in terms of degree, but *Information Processing Management* is at the top in terms of betweenness. *Journal of the American Society of Information Science and Technology* and *Information Processing Management* have the most common number of editors. This again is an evidence of the dominance of Information Science in shaping the domain.

Evaluation of the structure of EBs shows that Gary E Gorman has the maximum centrality score, as he is the member of 8 journals. Edward Fox and Gary Marchionini also play an important role with high centrality scores. Fox

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>682</td>
<td>46.94</td>
</tr>
<tr>
<td>UK</td>
<td>155</td>
<td>10.67</td>
</tr>
<tr>
<td>ND</td>
<td>74</td>
<td>5.09</td>
</tr>
<tr>
<td>Germany</td>
<td>48</td>
<td>3.30</td>
</tr>
<tr>
<td>Canada</td>
<td>47</td>
<td>3.23</td>
</tr>
<tr>
<td>China</td>
<td>54</td>
<td>3.72</td>
</tr>
<tr>
<td>Japan</td>
<td>32</td>
<td>2.20</td>
</tr>
<tr>
<td>Australia</td>
<td>27</td>
<td>1.86</td>
</tr>
<tr>
<td>Italy</td>
<td>24</td>
<td>1.65</td>
</tr>
<tr>
<td>France</td>
<td>23</td>
<td>1.58</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>19</td>
<td>1.31</td>
</tr>
<tr>
<td>Singapore</td>
<td>18</td>
<td>1.24</td>
</tr>
<tr>
<td>Denmark</td>
<td>17</td>
<td>1.17</td>
</tr>
<tr>
<td>Spain</td>
<td>17</td>
<td>1.17</td>
</tr>
<tr>
<td>Sweden</td>
<td>16</td>
<td>1.10</td>
</tr>
<tr>
<td>India</td>
<td>15</td>
<td>1.03</td>
</tr>
</tbody>
</table>
and Marchionini, having figured in most of the previous studies of DL, appear to be leading the field. It also shows that the network is highly connected with giant component comprising 84.1\% of editors. Geographical distribution of these editors confirms the dominance of the US. Our study confirms that DL is converging from different disciplines but predominantly Information Sciences and Library Sciences with Computer Sciences as the common thread. The EB analysis is a useful tool for studying similar phenomenon.

References


DELOS. 2007. Digital libraries for the digital librarian: making the journey from traditional to digital libraries. Italy: Joint DELOS—NSDL Summer School


Gladwell M. 2003. The Tipping Point: how little things can make big difference. UK: Little Brown


Kuhn T. 1970. The Structure of Scientific Revolutions. Chicago: University of Chicago


Oracle of Bacon. Details available at http://oracleofbacon.org


Perceptions of South Asian LIS community towards open source software adoption in libraries

Muhammad Rafiq
PhD Scholar, Department of Library and Information Science, University of the Punjab, Lahore, Pakistan

Kanwal Ameen
Chairperson, Department of Library and Information Science, University of the Punjab, Lahore, Pakistan

Abstract
This study is intended to investigate the perceptions of South Asian LIS community towards open source software (OSS) adoption in libraries by adopting a quantitative research design. An online questionnaire-based survey was conducted to reach the geographically dispersed population in the vast region of South Asia. A Likert-type scale with 20 items was designed to collect data. A total of 190 responses were gathered from four countries, namely, Afghanistan, Bangladesh, India, and Pakistan. The collected data were analyzed using Statistical Package for Social Sciences (SPSS). Descriptive statistics, independent sample t-Test, and ANOVA were used to draw conclusions. Attention was given to two independent variables: organization type (public/private sector), and library type (academic/public/special). The overall results suggested that respondents had positive perceptions towards OSS adoption in libraries. However, the results of the t-Test and ANOVA indicated that ‘organization type’ and ‘library type’ were insignificant factors, as there were no significant differences of perceptions found among the respondents on the basis of ‘organization type’ or ‘library type’.

This study is the first of its kind from a South Asian perspective. Its findings are useful for the LIS community, library system developers, software developers, technology administrators, and library administrators.
Introduction

Open source is an alternative model to develop and implement technological applications. Open Source Software (OSS) has dominated the infrastructure of Internet and Web services, and has a long history of supporting technology infrastructure (Altman 2001). OSS has a source code that is available under a license permitting its use, right to change and improve the software, and to redistribute it in modified or unmodified form. Software developers from all over the world have been contributing towards OSS’ development. As a result, a huge OSS community has emerged, which has significant impact on various disciplines, including the library and information management profession since 1999. In that year, three key principles were created in a meeting of 80 senior American academic library managers to establish a foundation for future developments of library services. One of the key principles was to ‘create OSS for access, dissemination, and management of information’ (Chawner 2004).

OSS is an attractive proposition to libraries around the world, especially those in developing countries that cannot afford to purchase costly, commercial software. Hence, its use is spreading swiftly. So is the pace of literature emerging on the subject. In this context, it is deemed necessary to explore the perceptions of the LIS community towards OSS. The findings of this study reveal the behavioural landscape of the South Asian LIS community towards OSS adoption by libraries, which is an important factor in its adoption on a wider scale. It is believed that the LIS community, library system developers, software developers, and library administrators will find the findings useful.

Literature review

Existing literature comprises significant studies on open source and its applications. The literature on software engineering, education (Rooij 2007), and many other disciplines has produced significant works on OSS. Andersson and Karlsson (2005) conclude that ‘two camps have been established: those who advocate the possibilities of introducing OSS into companies, and those who do not’.

Literature on library and information science also presents a significant amount of scholastic discussion on this issue. A number of authors have made different claims about suitability of OSS to libraries, and have discussed the benefits, drawbacks, and related issues. Librarianship and OSS share many common principles and cultural values (Morgan 2002; Raymond 1999; Bretthauer 2001; Poynder 2001; Altman 2001; Jaffe, Lee, and Careaga 2007; Frumkin 2002), such as sharing information (Altman 2001; Morgan 2002; Raymond 1999) and promoting open standards (Altman 2001). Bretthauer (2001) considered OSS as an opportunity for those libraries with a ‘tendency to push innovations’.

OSS is reported to be a low-cost solution for technological applications that offers libraries cheap alternatives to expensive commercialized solutions (Altman 2001; Arkles 2002; Cervon 2003; Muir 2005; Wan 2007). According to Lerner and Tirole (2002), participation in OS projects can result in tangible rewards such as promotions and employment opportunities. This claim has been asserted by the findings of a survey by Ghosh, Glott, Krieger, et al. (2002), during which it was revealed that about one-third of respondents’ (developers) reported access to better job opportunities motivated them to participate in the OS activity. Dorman (2004) has opined that the use of OSS may change staffing patterns in libraries’ through the provision of inducting new staff with relevant software expertise in the library.

Tennant (2000) assumed that large libraries are more likely to have the staff with skills necessary for OSS installation and maintenance. Furthermore, the level of technical knowledge required to install and maintain OSS is an
obstacle to its widespread adoption (Bretthauer 2001; Poynder 2001). However, Bisson (2007) claimed that ‘OSS is solving problems for libraries of all types and is helping to get work done for people of all technical abilities’. Cervon (2003) asserted that OSS enhances the computing flexibility of an organization.

Nevertheless, Altman (2001) and Cervon (2003) considered OSS a less user-friendly option with lower functionality than its commercial equivalents. A major issue associated with OSS is the level of support available (Chawner 2004; Ho 2007). Another issue associated with OSS is documentation. Documentation requires excellent technical writing skills—a trait not necessarily possessed by every developer. According to Murray (2002), poor-quality documentation is a drawback of OSS. Many open source projects face significant challenges to generate and maintain high quality documentation (OSS watch—Documentation issues in open source 2007; Fundamental Issues with Open Source Software Development 2007).

Three studies were identified during intensive literature search on perceptions about OSS. An exploratory study (Rooij 2007) investigated the perceptions among academic decision-makers about technology and OSS’ benefits and risks vis-à-vis commercial software applications. The study further explored relations between outsourcing campus-wide deployment and maintenance of open source. The study concluded that respondents perceived OSS as a solution to the functionality gap and provides control over ‘one’s own destiny’. Moreover, the study established that respondents have negative perceptions about the value of commercial software in terms of its meeting the desired needs and functions in relation to its cost. The respondents expressed that the issues of security, technical support, and longevity can be outweighed by enhanced knowledge base of the institution’s internal IT staff. Moreover, the study revealed that respondents showed interest in the outsourcing concept. However, a significant amount of skepticism exists because of their previous experiences with external consultants’ regarding implementation cost overruns.

Another study was conducted by Forrester Consulting (2007) to know how open source software is being used in North America and Europe in order to understand its role in IT, and examine the barriers and benefits to enterprise customers. The study found the leading attributes of attraction were OSS’ usage without restrictions and prevention of software vendor lock in. On the other hand, the biggest concern was to find ‘technical support’. The survey revealed respondents’ perceptions as: OSS provides significant economical and technological benefits, including cost savings, improving overall efficiency of IT and quality of products and processes, more innovation, increased competition among service offerings, and more efficient use of resources across the industry.

A recent study was conducted by Rafiq (2009) to explore the LIS community’s perceptions towards OSS adoption in libraries. The study received 370 responses from 48 countries. The study concluded that respondents had positive perceptions towards OSS. However, the results also indicated that ‘organization type’ and ‘library type’ were insignificant factors, but ‘country type’ was a significant factor. This is because respondents from developed countries have significantly different perceptions as compared to their counterparts from developing countries.

Nevertheless, the extensive literature search confirmed that no prior survey has been conducted to know the perceptions of the South Asian LIS community towards OSS adoption in libraries. Hence, this is the first study that specifically explores the South Asian region.
Research design and methodology

The study adopted a quantitative research design. Online questionnaire-based survey was considered appropriate to reach the geographically widely distributed population. Online survey is considered an effective way to gather information quickly when the population/sample is widely distributed geographically (Sue and Ritter 2007). Each questionnaire had two sections. The first section gathered demographic information, including country name, parent organization type, and library type. The second section consisted of five-point Likert-type scale to measure perceptions against 20 items. The literature revealed a number of statements by different researchers related to benefits, drawbacks, issues, and so on associated with OSS. These statements were used in the second section of the questionnaire. Respondents were requested to choose their level of agreement or disagreement about the given statements. The responses to the items were recorded as: strongly agree = 5, agree = 4, no opinion = 3, disagree = 2, and strongly disagree = 1. The ‘No opinion’ option was added to cover those respondents who might have no answer to any item due to their lack of knowledge on the subject. Particular emphasis was given to two variables, namely, organization type (public/private sector), and library type (academic/public/special). The coefficient alpha (0.77) reliability from the results suggests that the instrument is stable enough to determine South Asian LIS community’s perception towards OSS.

The subjects of the study were all LIS professionals. To reach the population, the questionnaire was distributed by email to those LIS professionals known to the researcher. This method was coupled with the issuance of a ‘call for participation’ by emailing discussion groups. To facilitate participation, the questionnaire was also uploaded on the Survey Monkey website. The first call for participation was issued in July-August 2007 on three email discussion groups of Pakistan, namely, Pakistan Library Cooperation Groups (LIBCOOP), Pakistan Library Automation Group (PLAG), and Librarian Welfare Group. These email discussion groups were the representative e-groups of librarianship in the country. In September 2007, another call for participation was issued on different international email discussion groups and listservs, including Indian discussion groups. The researcher also issued reminders. The questionnaire elicited 190 responses from four South Asian countries. Unfortunately, there were only two responses from Bangladesh and one from Afghanistan.

Research questions

The study contained the following research questions.

1. What are the perceptions of South Asian LIS professionals towards OSS adoption in libraries?
2. Is there significant difference between the perceptions of South Asian LIS professionals serving in public sector and private sector institutions?
3. Is there significant difference among the perceptions of South Asian LIS professionals serving in academic libraries, public libraries, and special libraries?

Data analysis and discussions

Data analysis was accomplished using SPSS for Windows. The respondents belonged to four South Asian countries. The major groups were from India (107, 56.3%) and Pakistan (80, 42.1%). There were almost equal number of respondents from both public and private sector institutions. Majority of the respondents belonged to academic libraries (see Table 1).
Table 1 Respondents’ characteristics

<table>
<thead>
<tr>
<th>Sector</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>96</td>
<td>50.8</td>
</tr>
<tr>
<td>Private</td>
<td>93</td>
<td>49.2</td>
</tr>
<tr>
<td>Library type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic</td>
<td>119</td>
<td>63.6</td>
</tr>
<tr>
<td>Public</td>
<td>5</td>
<td>2.7</td>
</tr>
<tr>
<td>Special</td>
<td>63</td>
<td>33.7</td>
</tr>
</tbody>
</table>

Analysis of descriptive statistics of the responses (see Table 2 below) revealed that the respondents agreed with majority of the statements. However, the respondents selected 'no opinion' to four statements, namely, 'Staff of libraries in my country is competent enough to implement OSS' (Mean = 2.93), 'OSS are less user-friendly than commercial software' (Mean = 2.99), 'OSS provide lower functionality than commercial equivalents' (Mean = 2.93), and 'OSS include poor quality documentation' (Mean = 2.87). The last three statements are negative and present major issues associated with OSS. The 'no opinion' response revealed that respondents considered OSS to vary in terms of level of documentation, maturity, and functionality. Some OSS products entail significant level of documentation, functionality, and user-

Table 2 Respondents’ level of agreement with the statements (N=190)

<table>
<thead>
<tr>
<th>Statements</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS philosophy matches with libraries’ mission and objectives in a broader sense</td>
<td>4.08</td>
<td>.819</td>
</tr>
<tr>
<td>OSS is a good choice for libraries to adopt</td>
<td>4.19</td>
<td>.783</td>
</tr>
<tr>
<td>OSS is an economical solution for technological application in libraries</td>
<td>4.28</td>
<td>.771</td>
</tr>
<tr>
<td>Staff of libraries in my country is competent enough to implement OSS</td>
<td>2.93</td>
<td>1.171</td>
</tr>
<tr>
<td>OSS development includes sharing of knowledge and skills</td>
<td>4.17</td>
<td>.781</td>
</tr>
<tr>
<td>OSS development includes learning and developing new skills</td>
<td>4.25</td>
<td>.769</td>
</tr>
<tr>
<td>OSS model improves job opportunities</td>
<td>3.55</td>
<td>.973</td>
</tr>
<tr>
<td>OSS development/adoptions is to participate in a new form of cooperation</td>
<td>4.02</td>
<td>.793</td>
</tr>
<tr>
<td>OSS is a low-cost solution for libraries</td>
<td>4.23</td>
<td>.796</td>
</tr>
<tr>
<td>OSS provides flexibility to customize according to local needs of libraries</td>
<td>3.98</td>
<td>.897</td>
</tr>
<tr>
<td>OSS movement gives librarians an opportunity to become more active in determining the future development of the software they use, rather than letting vendors keep control</td>
<td>4.11</td>
<td>.896</td>
</tr>
<tr>
<td>Low startup cost associated with using OSS is a main attraction to use</td>
<td>3.79</td>
<td>.979</td>
</tr>
</tbody>
</table>

Contd...
Statements Mean SD

OSS approach provides more flexibility and a better match to libraries’ requirements 3.79 .870

OSS causes an increase in staff expertise through involvement in new developments 3.84 .924

With no vendor responsible for the software, support for OSS applications can vary, and often depends on the user/developer community’s commitment to the project 3.88 .830

Level of technical knowledge needed to install and maintain OSS can also be a barrier to its use 4.00 .943

OSS includes poor quality documentation 2.87 1.057

OSS is less user-friendly than commercial software 2.99 1.145

OSS provides lower functionality than its commercial equivalents 2.93 1.081

Large libraries are more likely to have staff with the necessary skills and experience to implement OSS 3.72 .987

Note: Strongly agree = 5, Agree = 4, No opinion = 3, Disagree = 2, Strongly disagree = 1

### Table 3 Results of t-Test regarding perceptions of the respondents by organization type (Public/Private)

<table>
<thead>
<tr>
<th>Statements</th>
<th>Public Mean</th>
<th>Private Mean</th>
<th>T</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Source philosophy matches with libraries’ mission and objectives in a broader sense</td>
<td>4.07</td>
<td>4.12</td>
<td>-.386</td>
<td>.700</td>
</tr>
<tr>
<td>OSS is a good choice for libraries to adopt</td>
<td>4.21</td>
<td>4.18</td>
<td>.223</td>
<td>.824</td>
</tr>
<tr>
<td>OSS is an economical solution for technological application in libraries</td>
<td>4.29</td>
<td>4.26</td>
<td>.299</td>
<td>.765</td>
</tr>
<tr>
<td>Staff of libraries in my country is competent enough to implement OSS</td>
<td>2.86</td>
<td>2.98</td>
<td>-.668</td>
<td>.505</td>
</tr>
<tr>
<td>OSS development includes sharing of knowledge and skills</td>
<td>4.16</td>
<td>4.19</td>
<td>-.327</td>
<td>.744</td>
</tr>
<tr>
<td>OSS development includes learning and developing new skills</td>
<td>4.20</td>
<td>4.31</td>
<td>-1.015</td>
<td>.311</td>
</tr>
<tr>
<td>OSS model improve job opportunities</td>
<td>3.49</td>
<td>3.62</td>
<td>-.945</td>
<td>.346</td>
</tr>
<tr>
<td>OSS development/adoption facilitates new forms of cooperation</td>
<td>4.08</td>
<td>3.94</td>
<td>1.280</td>
<td>.202</td>
</tr>
<tr>
<td>OSS is a low-cost solutions for libraries</td>
<td>4.27</td>
<td>4.19</td>
<td>.665</td>
<td>.507</td>
</tr>
</tbody>
</table>

Contd...


<table>
<thead>
<tr>
<th>Statements</th>
<th>Mean Public</th>
<th>Mean Private</th>
<th>T</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSS provides flexibility to customize according to local needs of libraries</td>
<td>3.95</td>
<td>4.01</td>
<td>-.479</td>
<td>.632</td>
</tr>
<tr>
<td>OSS movement gives librarians an opportunity to become more active in determining the future development of the software they use, rather than letting vendors keep control</td>
<td>4.11</td>
<td>4.10</td>
<td>.136</td>
<td>.892</td>
</tr>
<tr>
<td>Low start-up cost associated with using OSS is a main attraction to use</td>
<td>3.76</td>
<td>3.84</td>
<td>-.548</td>
<td>.584</td>
</tr>
<tr>
<td>OSS approach provides more flexibility and a better match to libraries' requirements</td>
<td>3.83</td>
<td>3.76</td>
<td>.551</td>
<td>.582</td>
</tr>
<tr>
<td>OSS causes an increase in staff expertise through involvement in new developments</td>
<td>3.99</td>
<td>3.69</td>
<td>2.253</td>
<td>.026*</td>
</tr>
<tr>
<td>With no vendor responsible for the software, support for OSS applications can vary, and often depends on the user/developer community’s commitment to the project</td>
<td>3.86</td>
<td>3.90</td>
<td>-.319</td>
<td>.750</td>
</tr>
<tr>
<td>Level of technical knowledge needed to install and maintain OSS can also be a barrier to its use</td>
<td>3.96</td>
<td>4.04</td>
<td>-.615</td>
<td>.540</td>
</tr>
<tr>
<td>OSS includes poor quality documentation</td>
<td>2.93</td>
<td>2.83</td>
<td>.645</td>
<td>.520</td>
</tr>
<tr>
<td>OSS is less user-friendly than commercial software</td>
<td>2.96</td>
<td>3.03</td>
<td>-.442</td>
<td>.659</td>
</tr>
<tr>
<td>OSS provides lower functionality than its commercial equivalents</td>
<td>2.85</td>
<td>3.01</td>
<td>-.995</td>
<td>.321</td>
</tr>
<tr>
<td>Large libraries are more likely to have staff with the necessary skills and experience to implement OSS</td>
<td>3.79</td>
<td>3.66</td>
<td>.944</td>
<td>.347</td>
</tr>
</tbody>
</table>

Note: The mean difference is significant at the .05 level.

friendliness, while others do not. This fact might have hindered the respondents to adopt a clear line of acceptance or rejection about the issues of documentation, maturity, and functionality.

The statistical analysis of independent sample t-test (Levine’s test for the equality of variance) shows that there is no significant difference of perceptions among the respondents by organization type (public sector / private sector), except for one statement: ‘OSS cause an increase in staff expertise through involvement in new developments’ (Table 3). The analysis of mean scores shows that the respondents from private sector organizations have a significant difference of opinion in comparison with those from public sector organizations. Respondents from public sector organizations agreed more to the statement.
Analysis on the basis of library type was not executed as the representation from public libraries was very low (five respondents only). Hence, it was not appropriate to execute ANOVA on the basis of library type.

**Conclusion**

OSS is rapidly gaining attention of the professional LIS community. OSS provides alternative, cheap, and innovative technological solution to libraries. For this reason, it can be an ideal alternative to the costly proprietary software for libraries. The overall findings of the study suggest that respondents have positive perceptions towards OSS adoption in libraries. However, the results of the t-Test indicated that organization type was an insignificant factor. Nonetheless, OSS adoption in libraries is still at infancy in these countries. That the community has favourable perceptions to OSS, yet the adoption is still very slow is an interesting observation. Hence, the findings necessitate the need for further enquiry to unfold the factors that are hindering the adoption of OSS in libraries on a wider scale.

**References**


Perceptions of South Asian LIS community towards open source software adoption in libraries


OSS watch—Documentation issues in open source. 2007. [Available at: http://www.oss-watch.ac.uk/resources/documentation.xml].


Agent-based web intelligent tutoring system

Rohit Chawla  
KRM School of Engineering, New Delhi  
rc.j2ee@gmail.com

Priyanka Chawla  
Amity University, Noida  
priyankamatrix@gmail.com

Abstract

Web Intelligence is a direction for scientific research that explores practical applications of Artificial Intelligence to the next generation of web-empowered systems. In this paper, we present an agent-based web intelligent tutoring system. The decision-making process conducted in our intelligent system is guided by knowledge source networks, based on synergistic integration of knowledge acquired from distributed sources in order to obtain new knowledge or complement insufficient knowledge. We have provided its possible application to logistics support in networked organizations in the form of a case study of routing plans design as a part of the logistics support task.
Introduction

Web-based learning systems are becoming increasingly popular due to their obvious advantages over traditional paper-based textbooks. Web courseware is easily accessible and offers greater flexibility, that is, students can control their own pace of study. Unlike printed textbooks, web-based tutoring systems can incorporate multimedia such as audio and video to explain a point better. However, since many current web-based tutoring systems are static HTML web pages, they suffer from two major shortcomings—they are neither interactive nor adaptive. Web intelligence is a direction for scientific research that explores practical applications of Artificial Intelligence to the next generation of web-empowered systems (Smimov, Chilov, and Levashova 2003). For instance, Yao and Yao (2003) argue that a system should be robust enough to deal with various types of users. In the context of web-based tutoring systems, Liu, Zheng, and Yang (2001) developed an intelligent system for assisting a user in solving a problem.

Logistics support in a networked organization presents numerous challenges due to a variety of different policies, procedures, and practices of the members of the organization, for example, diversity of goals and strategies, logistics flexibility, resource limitations, different stockage levels, interoperability concerns, and so on. Intensive knowledge exchange in such organizations leads to a need for acquisition, integration, and transfer of the right knowledge from distributed sources located in a network-centric environment. This knowledge has to be delivered in the right context to the right person, in the right time for the right purpose. These activities called knowledge logistics are required for global awareness, dynamic planning, and global information exchange in the network-centric environment. Transportation systems play an important role in distributed companies such as networked organizations. An intelligent transportation system may significantly enhance the transportation abilities. Therefore, the approach presented in the paper is illustrated through an example of designing efficient routing plans—as one of the major tasks of logistics support—under given constraints and preferences.

Background knowledge

In this section, we briefly review intelligent tutoring systems and the knowledge source networks.

Intelligent tutoring systems

Computers have been used in education for over three decades now. Traditional computer-assisted instruction (CAI) presents instructional materials in a rigid tree structure to guide the student from one content page to another, depending on the answers he/she is seeking. This approach is restrictive, in that, it does not consider the diversity of students’ knowledge states and their particular needs (Wang and Vassileva 2003). Moreover, CAI systems are not adaptive and are unable to provide the individualized attention that a human instructor can provide (Brusilovsky 1999).

An intelligent tutoring system (ITS) is a computer-based programme that presents educational materials in a flexible and personalized way (Brusilovsky 1999; Johnson 2001). These systems can be used in the normal educational process, in distant learning courses, either operating on stand-alone computers or as applications that deliver knowledge over the Internet. ITS must be able to achieve the following three main tasks.

(i) Accurately diagnose a student’s knowledge level using principles, rather than pre-programmed responses;
(ii) Decide what to do next and adapt instruction accordingly; and
(iii) Provide feedback.

This kind of diagnosis and adaptation, which is usually accomplished using Artificial Intelligence
techniques, is what distinguishes an ITS from CAI. Empirical studies have shown that individual one-on-one tutoring is the most effective mode of teaching and learning, and ITSs, uniquely, offer a technology to implement computer-assisted one-on-one tutoring.

Knowledge logistics

A networked organization environment can be characterized as an open system where users, units, and other resources come and go continually; and entities provide services to each other under various contract forms. At the same time, the business environment of the networked organization is highly dynamic and requires continuous engineering. As a result, a new direction called continuous business engineering (CBE) has emerged. CBE is aimed at integrating forward development at the business strategy level and engineering process at the software level. It addresses the joint evolution of business strategies and software technologies like open services.

The knowledge source network is driven by a will to expose business logic beyond one company. Through open services, companies can prepare existing business processes, publish them as services, search for and subscribe to other services, and exchange information and knowledge throughout and beyond their boundaries, thereby forming a collaborative environment of the networked organization. Some of the benefits of such a collaborative environment include:

- Decentralization: Business services can be completely decentralized and distributed over the Internet and accessed by a wide variety of communication devices.
- Dynamic interoperable: New business partnerships can be constructed dynamically and even automatically.
- Flexibility: A highly dynamic enterprise consortium has greater possibilities to use new market opportunities.

Architecture of the system

Knowledge logistics refers to dealing with knowledge contained in distributed and heterogeneous knowledge sources. As a result, the approach is oriented towards an ontological model providing a common method of knowledge representation. This approach proposes ontology-driven methodology to configure the knowledge source network.

Due to the problem’s distributed nature, this approach assumes an agent-based architecture. Multi-agent system architecture, based on Foundation for Intelligent Physical Agents (FIPA) Reference Model (Johnson 2001) as an infrastructure for definition of agent properties and functions, was chosen to form the technological basis for the system, implementing the approach since it provides standards for heterogeneous interacting agents and agent-based systems, and specifies ontological and negotiation protocols to support interoperability in specific application areas. FIPA-based technological kernel agents used in the system are—wrapper (interaction with knowledge sources); facilitator (“yellow pages” directory service for the agents); mediator (task execution control); and user agent (interaction with users). The following problem-oriented agents, specific for knowledge logistics, and scenarios for their collaboration were developed translation agent (terms translation between different vocabularies); knowledge fusion agent (knowledge fusion integration operation performance); configuration agent (efficient configuring of the knowledge source network); ontology management agent (ontology operations performance); expert assistant agent (interaction with experts); and monitoring agent (verifications of knowledge sources). A community of agents is represented in Figure 1, according to the above described principles and functions of the system.
Knowledge source network functionality

The main activity of this architecture is a remote procedure calling, that is, clients invoke pre-prepared tasks that facilitate some activity. Software objects can be accessed over the Internet using standard protocols to perform functions or make decisions.

Knowledge sources for creation of efficient routing plans

In the presented case study, a fictitious Indian region is considered. The aim of using this scenario is to provide for a rich environment, focusing on new aspects of coalition problems and new technologies, demonstrating the ability of distributed systems for intelligent support to supply services in an increasingly dynamic environment. For the considered task of creation of mobile hospitals, the following knowledge sources can be considered.

- Supplies related information (required quantities of materials, required times of delivery);
- Available suppliers (constraints on suppliers’ capabilities, capacities, and locations);
- Available providers of transportation services (constraints on available types, routes, and time of delivery); and
- Geography and weather of the region (constraints on types, routes, and time of delivery).

The problem of automatic knowledge seeking is a subject of future research. For the case study, a list of knowledge sources containing information for the user request processing had been prepared by a team of experts.
Information about suppliers is stored in three alternative knowledge sources represented by remote databases. These knowledge sources have different characteristics (access time, price, and reliability), and they are selected in a course of the user request processing with regard to the chosen preferences of the request processing. The route system is presented by a weighed graph of routes assigned transportation time, costs, and route type (highway, road, transportation by a plane, and so on). This information is stored in an external database accessed remotely. In order to provide up-to-date routing plans, the system monitors the current situation in the region. For this purpose, an emulated news website has been implemented that contains information about weather and events in the considered region. A specially designed wrapper reads news and finds the cited areas that are not currently available for transportation. Besides, it reads weather conditions and accordingly corrects transportation time and costs for appropriate routes. This list of sources is not fixed. The scalable architecture of the system allows seamless attaching of new sources in order to get new features and to account for more factors for the tasks being solved.

**Logistics support case study**

The presented example illustrates the creation of a routing plan for the same conditions, but with different user preferences such as the following.

- minimize time;
- minimize costs;
- minimize both time and costs;
- minimize costs, then time;
- minimize costs.

In Figure 3, results for different choices are presented and compared. Cities of the region are indicated.

**Conclusion**

Web intelligence explores the practical application of artificial intelligence to the next generation of web-empowered systems. The
paper describes an approach to knowledge logistics, combining both intelligent agents and Web service technologies. To illustrate the proposed approach, a task of creation of routing plans was undertaken, since it is one of the key tasks of logistics support in networked organizations. The scalable architecture of the approach enables its extension with regard to the number of knowledge information sources, and thereby to factors taken into account during complex problem solving. Implementation of the approach as a Web-service makes the system accessible from nearly any Web-service compatible application running on any platform.

References


Figure 3 Routing Plans for different criteria (time & costs minimization preferences)
Agent-based web intelligent tutoring system


A survey of digital preservation strategies

Feng Luan  
Department of Computer and Information Science, Norwegian University of Science and Technology, N-7031, Trondheim, Norway  
luan@idi.ntnu.no

Mads Nygård  
Department of Computer and Information Science, Norwegian University of Science and Technology, N-7031, Trondheim, Norway  
mads@idi.ntnu.no

Thomas Mestl  
Research and Innovation, Det Norske Veritas, N-1322, Høvik, Norway  
thomas.mestl@dnv.com

Abstract  
Digital information is created easily by a wide range of electronic devices. But, it is surprisingly difficult to preserve digital information over a long period, because both technology as well as relevant organizational context change over time. Without careful protection, there is a real risk that these data can neither be retrieved from old storage media, nor that its content be rendered from old file formats or understood. In the future, our society may indeed be confronted with a digital ‘black hole’. From the 1990s, scientists and curators started to design long-term digital preservation systems and strategies that shall enable the survival of digital information. This survey article gives an overview of the current main preservation strategies. We further discuss and compare those strategies. Finally, we design two procedures for making information accessible and manipulable over time. At the end of the article, some emerging challenges for preservation strategies are addressed briefly.
Introduction

Digital preservation is the exploration of systematic approaches to ingest, archive, and disseminate digitized information. The development of digital preservation is heavily influenced by the use of electronic devices such as personal computers, electronic game devices, mobiles, digital cameras, digital recorders, and digital TVs. In the last decade, increased dependency of our modern society on digital information has led scientists to worry about digital preservation issues.

Earlier research work on digital preservation was carried out by the Commission on Preservation and Access and the Research Libraries Group (RLG) in 1994. This task force summarized the basic requirements of preservation and issued a report (Waters and Garrett 1996). Between 1996 and 2002, many libraries, archives, and data centres started to create digital repositories for their digitized information. The Consultative Committee for Space Data System (CCSDS) proposed an infrastructure model—the Open Archive Information System (OAIS) (The Consultative Committee for Space Data Systems 2002). Later, in 2003, the OAIS became an ISO 14721:2003 standard. Besides defining the infrastructure, efforts were more focused on preservation policies that would define a trustworthy preservation system. For example, Dale and Ambacher (2007) defined the trusted digital repositories audit and certification (TRAC); while Dobratz, Hanger, Huth, et al. (2006) defined the nestor criteria.

Preservation metadata is another central element in digital preservation. In terms of the OAIS, Lavoie and Dale (2002) compiled a metadata dictionary, which is an aggregation of the CURL Exemplars in Digital Archive project (CEDARS), the National Library of Australia (NLA), the Networked European Deposit Library (NEDLIB), and the Online Computer Library Center, Inc. (OCLC). The National Library of New Zealand also proposed a metadata dictionary (the National Library of New Zealand 2003) for automatic metadata extraction. The PREMIS (PREservation Metadata: Implementations Strategies) working group from 2005 started to define a preservation metadata standard. The latest version is the PREMIS 2.0. In total, there are 13 preservation systems that have deployed PREMIS 2.0.

This article focuses on the very core of a preservation system, that is how to ensure that digital information can survive for a long period and the existing strategies. The readers will hopefully get a quick review about:

- popular preservation systems around the world
- various preservation strategies
- selecting the most appropriate preservation strategy
- possible challenges for the preservation strategies

The article starts with an overview of related work on digital preservation, including the OAIS standard, preservation systems, and research work groups, in Section 2. Threats to digital information are summarized in Section 3, and various preservation strategies are introduced in Section 4, followed by a further discussion and analysis of the strategies in Section 5. Finally, we summarize future research works in Section 6.

Related work

The OAIS standard

The CCSDS published a report, the Reference Model for an Open Archival Information System (OAIS) (The Consultative Committee for Space Data Systems 2002). The OAIS became an ISO Archiving Standard in 2003. Many preservation systems or repositories adapted the OAIS standard as their infrastructural model. Figure 1 gives an overview of the OAIS standard.

1 http://www.loc.gov/standards/premis/premis-registry.php
The OAIS standard defines three roles for a person: producer, management, and consumer. A producer produces digital information. As long as the producer wants to submit digital information to a preservation system, the producer should bundle the digital information with metadata into a submission information package (SIP) based on the requirements of the preservation system. After receiving the SIP, the preservation system will divide the SIP to an AIP (Archival Information Package) and related descriptive information. The AIP will be maintained in the long-term archival storage. The description information will be stored in a database of data management so that any consumer can browse through them.

When a consumer wants to access an AIP, the preservation system retrieves the AIP from the archival storage and transfers it to the access module. The access module then transforms the AIP to the DIP (Dissemination Information Package), and sends the DIP off to the consumer.

A manager with proper management rights is responsible for any necessary maintenance task in the preservation system. The manager should design the charter of the preservation system, such as the access scope for producers and consumers, and the agreements with producers and consumers. In addition, the manager should design the preservation plan so that all AIPs can be accessed and manipulated in future.

**Preservation systems**

Most of the well-known preservation systems are open-source systems that can be used free of charge by libraries and digital repositories. Some systems also offer free discussion groups.

**LOCKSS**

LOCKSS (Lots of Copies Keep Stuff Safe) is an open-source system under the auspices of the Stanford University. LOCKSS is a distributed, peer-to-peer preservation system that is able to manage multiple copies at remote data repositories. The system was released in 2004 and had been tested at more than 50 libraries worldwide.

**Eprints**

Eprints is a set of open-source software applications for building open access services. Eprints was developed by the University of Southampton in UK. The current version is Eprint 3, offering improvements in the architecture, automatic extraction of metadata, access control, flexible work flows, format support, and thumbnails.

**DSpace**

DSpace is an open-source software system developed by the Massachusetts Institute of Technology Libraries and Hewlett-Packard. The objective of DSpace is to provide a repository for research data sets and educational materials. DSpace can preserve not only digital journals and digitized documents, but also 3D digital objects, research data sets, and films.

**e-Depot**

e-Depot (Oltmans and Wijngaarden 2004) is a long-term preservation system developed by the
National Library of Netherlands. The core part of e-Depot is the DIAS\(^5\) that was developed by IBM. The objective of e-Depot is to maintain the integrity of stored digital objects so that these objects are accessible.

**FEDORA**

FEDORA\(^6\) (Flexible Extensible Digital Object Repository Architecture) is constructed under an open-source digital preservation infrastructure. FEDORA began in 1997, headed by the Cornell University and University of Virginia. In the latest version, FEDORA provides not only the basic functions of preservation systems, but also a model using semantic techniques, that is, the Resource Description Framework (RDF) maintaining relations between digital objects.

**iRODS**

iRODS\(^7\) (Integrated Rule Oriented Data System) is an open-source data grid software system. Developed by the Data Intensive Cyber Environments (DICE) research group and collaborators, it is the successor of the Storage Resource Broker (SRB). A peculiar feature of iRODS is its ability to represent the data preservation policies in a set of rules. Thereby, iRODS can interpret the rules and execute a sequence of pre-defined actions based on a given situation.

**Recent research work groups**

Current research groups that focus on the long-term preservation of digital information are as follows:

**PADI**

The National Library of Australia’s Preservation Access to Digital Information\(^8\) (PADI) is a gateway to digital preservation issues. PADI provides rich resources about every aspect of a preservation system such as archiving, preservation strategies, data documentation and metadata, intellectual property rights management, format and media, management, digitization, approaches at national libraries, and digital preservation tools.

**InterPARES**

InterPARES\(^9\) (International Research on Permanent Authentic Records in Electronic Systems), starting in 1999, has currently entered the third phase. In total, 15 teams from different countries are included in InterPARES. The objective of InterPARES is to develop the necessary knowledge for survival of digital record over a long-term period. InterPARES mainly works on the theoretical foundation for a preservation system, including aspects such as access, creation, maintenance, and security.

**DCC**

DCC\(^10\) (Digital Curator Centre) is a discussion centre for digital curators. It is funded by the Joint Information Systems Committee (JISC). DCC provides rich knowledge resources about digital preservation, such as a curation reference manual, curation lifecycle model, policies and legal reports, case studies, tools and applications, standards, publications, and a curation journal.

**CAMiLEON**

CAMiLEON\(^11\) (Creative Archiving at Michigan & Leeds: Emulating the Old on the New) was funded by the NSF/JISC. CAMiLEON explores various ways to keep the original functionality and ‘look and feel’ of digital objects. The emulation and migration on access, which will

---

7. https://www.irods.org/  
10. http://www.dcc.ac.uk/  
11. http://www2.si.umich.edu/CAMILEON/
be introduced in Section 4, are key outputs of CAMILEON.

**PLANETS**

PLANETS12 (Preservation and Long-Term Access through Networked Services) is funded by the European Union under the Sixth Framework Programme of 2006. The PLANETS objective is to improve decision-making in long-term preservation, so that the valued digital objects can be accessed. It has investigated several digital preservation challenges, such as a preservation plan, a set of characterizations of digital objects, preservation actions, an interoperability framework, and a test bed.

**CASPAR**

CASPAR13 (Cultural, Artistic, and Scientific knowledge for Preservation, Access, and Retrieval) is another project funded under the Sixth Framework Programme. CASPAR, implemented based on the OAIS guidelines, focuses on the following:

- what metadata are needed to describe the representation information and other relevant information
- how to integrate digital intellectual property rights for the preserved digital information
- how to integrate authentication and accreditation into the long-term preservation mechanism

**SHAMAN**

SHAMAN14 (Sustaining Heritage Access through Multivalent ArchiviNg) is financed by the European Union within the Seventh Framework Programme. The goal of SHAMAN is to develop new approaches for digital preservation. Those approaches shall not only guarantee preservation of digital content, but will also keep track of the digital content’s integrity, authenticity, semantics, and usage context.

12 http://www.planets-project.eu/
13 http://www.casparpreserves.eu/
14 http://shaman-ip.eu/shaman/

**Threats to preservation of digital information**

Traditional information carriers such as paper and stones can preserve information over long periods, for example, several decades or even hundreds of years. In contrast, our experience with digital information being generated in the last half century has shown us that digital information is confronted with considerably more threats than traditional information. The reasons are mainly due to its dependence on both storage media and interpretation software. Table 1 gives a short description of the main threats.

**Disasters/accidents**

Disasters (for example, flooding, earthquakes or terrorist attacks) and accidents (for example fire) could damage the necessary electronic or storage equipment leading to (partial) loss of digital information.

**Storage media fault**

Storage media are the core components of any preservation system as they hold the digital information in terms of physical characteristics, for example, magnetism and optics. The quality of those characteristics degrades as time passes. Thereby, faults would appear in a storage medium. Such faults are usually correctable unless the number of faults reaches a critical point, when (parts of) the information might be lost.

**Hardware/storage media obsolescence**

Hardware refers to any component of a computer system. The rapid evolution of hardware and storage media can make it difficult to access information on old storage media. For example, floppy disks were the standard storage media from mid-1970s to late 1990s, but 5¼-inch floppy disc readers are no longer available.
now. It will be only a matter of time before optical disks are replaced by memory sticks or online storage.

**Software/format obsolescence**
Software provides a way to organize and manipulate digital information according to pre-defined format structure(s), whereas formats must rely on software for correct interpretation of the sequence of 0s and 1s (bits). Software and file formats largely depend on each other. Any change in one, must be accompanied by an adaption of the other. The current rapid development in computing technology such as functionalities, increased performance, and new and better graphical interface, can result in a situation where old file formats can no longer be interpreted correctly by the newer software, and vice versa.

**Malicious attacks**
A preservation system that stores valuable information for an organization represents a potential target for malicious attacks such as a disgruntled employee and institutional espionage.

<table>
<thead>
<tr>
<th>Threat</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disasters/accidents</td>
<td>A natural/man-made disaster or accident destroys the preservation system.</td>
</tr>
<tr>
<td>Storage media fault</td>
<td>A storage medium has faults that make it unreadable.</td>
</tr>
<tr>
<td>Hardware/storage media obsolescence</td>
<td>Hardware or storage medium becomes too old, making it difficult to find a replacement part or get technical support.</td>
</tr>
<tr>
<td>Software/format obsolescence</td>
<td>If software/format is too old, the preservation system will replace it.</td>
</tr>
<tr>
<td>Malicious attacks</td>
<td>Malicious attacks on a preservation system can result in information modification or even loss of information.</td>
</tr>
<tr>
<td>Lack of context</td>
<td>A preservation system stores too little context information for correct interpretation/understanding of the preserved information.</td>
</tr>
<tr>
<td>Lack of authenticity</td>
<td>Since very few evidences are stored, the preservation system cannot prove authenticity of the preserved digital information.</td>
</tr>
<tr>
<td>Financial problems</td>
<td>Too little funding could threaten the necessary operation of a preservation system.</td>
</tr>
</tbody>
</table>

Lack of context
The context provides necessary background information for correct interpretation, understanding, and use of digital information. This usually includes a profile of digital information, information about the creation intention, and relations to other digital information. Lesser the context provided with the digital information, more difficult it will be for a future user to understand the data.

Lack of authenticity
Authenticity provides information on whether the preserved information is same as (or similar to) the original, and whether any modification on the information is done by authorized actions/personnel. Authenticity information includes annotated text, screenshots, annotation to screenshots, video clips, digital signatures,
access control, and so on. If the evidence is incomplete or lost, trustworthiness of the information is in doubt.

**Financial problems**

Digital preservation requires a high volume of funding, for example, hardware/software update, operational costs, personnel, information acquisition, insurance, unavailability, and cost of losing document (Crespo and Garcia-Molina 2001). Therefore, any successful long-term preservation system depends on steady financial support.

**Preservation strategies**

As digital information is highly dependent on computer technology, its preservation will necessarily have to be different from that of traditional, or printed information. It is not enough to just preserve the digital storage media (equivalent to paper), but interpretation and manipulation software (equivalent to the language and understanding capability of the human reader), as well. Analysis of a generic read-and-write process for digital information may provide some clues for potential preservation strategies.

Consider the situation depicted in Figure 2 (below) where a text entry (for example, ‘World Digital Libraries: an international journal’) shall be modified. The text data are stored in terms of the ASCII code. In the first phase, the text data are read from the storage media, and the bit stream is sent off via the storage driver into a software application (left side of figure). In the second phase, the software translates the bit stream into a human readable and understandable rendering (right side of the figure). After having modified the journal entry, this process is reversed.

This simple generic example highlights two fundamental aspects in digital preservation: storage accessibility (Phase 1) and bits manipulability (Phase 2). The storage accessibility requirement shall ensure the correct reading and writing to and from the data storage, whereas the bits manipulability requirement shall guarantee that the data can be made operational.

With this perspective in mind, four strategies could be envisioned that would maintain storage accessibility, and six strategies for maintaining bits manipulability. In the next section, we will shortly describe these preservation strategies.

**Storage accessibility maintenance**

Storage accessibility requires that digital information can be completely retrieved from — or written back to — a storage medium. Bit error or hardware faults can threaten storage accessibility. Baker, Shah, Rosenthal, et al. (2005) have analysed the reliability of storage media, and they argue that storage reliability can be increased by using better storage media, auditing storage media more frequently, automatic storage media repair, providing hot space drives, increasing the number of replicas, and increasing independence of the replicas. Accessibility could, therefore, be
Auditing

Auditing is an inspection procedure that looks for faults in a storage medium. Manual auditing is not feasible for large-scale preservation systems, where the volume of information may be too large. In addition, it may be difficult for administrators or readers to understand all the audit information. Once a fault is found, the preservation system should repair the storage as soon as possible. Otherwise, the information might be lost. For example, if all copies of digital information have various faults at the same time, the preservation system may not be able to recover them anymore. Thereby, auditing needs not only an application to find faults, but also an application to repair the faults as soon as possible.

Refreshment

Refreshment is a procedure where an old storage medium is replaced by a new one by copying.
the bits from the old medium to the new. This term differs from the OAIS’ definition that simply refers to copying between storage media within the same type and does not permit any alteration to the storage mapping infrastructure. However, we deem that refreshment can be performed between any two storage media, and that the storage mapping infrastructure can also be modified. Thus, refreshment includes replacement of a storage medium and updating the entire storage system. Refreshment is considered relatively simple, but it is not easy to determine the time when the refreshment should be executed. Most preservation systems perform refreshment every 3–4 years to avoid expiry of warranty of the storage media. However, at that time, most of the storage media are still in a perfect working condition, and hence, a lot of money is wasted in the process.

Replication

Replication refers to an approach where the preserved information has several copies using different formats that might be saved at different places. This definition also differs from that of the OAIS. The difference lies in that the OAIS requires package information, content information, and preservation description information cannot be changed, whilst we deem that digital information should use different formats. In this way, it results in the enhancement of reliability of digital information. Similar to refreshment, replication is a simple, but expensive approach, as it requires more advanced technology and larger storage space for replicas. Most preservation systems use the RAID (Redundant Array Independent Disks) technique, which offers duplication of data and a better I/O performance. The preservation systems also allow usage of several formats. For instance, the National Library of Norway uses the PDF format for their publications, XML format for full-text search, TIFF format for preservation, and JPEG2000 format for dissemination on the Internet.

Federation

Federation refers to a distributed network environment, where each independent preservation system shares its digital information with other preservation systems. If the digital information is not readable, the preservation system can recover this information from other preservation systems. Federation is a promising approach because it not only overcomes the threats to accessibility, but also reduces the dangers related to financial problems. Recent preservation systems such as LOCKSS and DSpace prefer building a federation environment. However, there are still some problems related to this approach, such as synchronization issues among the systems, and access right and confidentiality issues.

Bits manipulability maintenance

Bits manipulability requires that users can successfully manipulate the bit stream in a manner understandable by humans. The key in this requirement is the technical environment, which supports software (for example, format interpreter) that can be run. Software or format obsolescence is the typical threat, to tackle which six practical strategies can be proposed.

Computer museum

The computer museum strategy requires that the whole computer system, including hardware and software, should be preserved, so that the preserved digital information can be read and manipulated in the original environment. For example, the National Library of Norway preserves old audio devices since they often get old audio media from other organizations or individuals. This approach is only suitable in the short-term, because:

- Every hardware component has a limited lifetime.
- Currently, most hardware and software are proprietary. It is impossible for a
preservation system to produce the old hardware and software by itself.

- Hardware and software vendors might not support all of their previous products.
- In some situations, it is more expensive to maintain old hardware/software than to buy a new one.

Emulation

Emulation uses a software application to imitate the function of an old hardware component. In this way, the preservation system is able to execute the old software without the real hardware component. When using emulation, the relevant software and file formats can be handled at the same time. This makes emulation a useful strategy for complex digital information, such as databases and computer games. However, Granger (2000) has argued that emulation is not suitable for long-term preservation because:

- The preservation task becomes more complex, since the preservation target switches to the whole computer system rather than just digital information.
- Information management becomes harder because the preserved information is dispersed over various emulation applications.
- It is difficult to develop an emulation application. Thus, the preservation system has to rely on an external entity.

Encapsulation

In encapsulation, a new software application is developed that is able to manipulate the preserved digital information. Encapsulation requires that the archive package must preserve the digital information and relevant format specifications. The main preservation task of encapsulation is to develop a new software application. Encapsulation is not feasible in the long term according to (Waugh, Wilkinson, Hills, et al. 2000). They wrote about three major challenges:

- There may be no application that can automatically generate the archive package for the encapsulation. The preservation system should automatically perform this function.
- The storage overhead of encapsulating the format specification with digital information might be a problem.
- Some formats are proprietary and their specifications may be unpublished. A part of the encapsulation approach might include a transformation from unpublished, proprietary formats to published, open formats.

In addition, we may point out another challenge, that is, the formats in the preservation system are too many to create the manipulation applications as time passes. The cost for developing such applications might surpass the preservation system's ability. This situation might be mitigated by converting the original format to some standard format.

Universal virtual computer

The universal virtual computer (UVC) proposed by Lorie (2001; 2002) is an innovative approach. UVC is similar to the JAVA Virtual Machine. Digital information and the relevant and original software applications are compiled with a set of special instructions of UVC. For each computer generation, the digital information can be manipulated by the original application in UVC. The National Library of Netherlands has used such an UVC for JPEG and PDF files manipulation, but more pilot projects are needed to test this approach for other formats. Moreover, Lorie mentioned that UVC may be hard to optimize.

Batch migration

Batch migration aims to overcome the threats of software obsolescence and format obsolescence in that the preservation system periodically transforms digital information from one format
to another. The process seems simple, but requires a careful experiment and validation phase, where alternative solutions should be tested. Strodl, Becker, Neumayer, et al. (2007) used utility analysis to determine the preferred format. Several case studies at many European libraries show that utility analysis is a feasible approach for batch migration. However, the disadvantages of the process are:

- Digital information is not preserved in its original format, and consequently a degradation of the digital information might occur.
- The time to execute the batch migration will become a challenge. In the latest system update, it took the National Library of Norway more than three months to do refreshment. When doing batch migration, more time is needed. A dangerous scenario could emerge where the old batch migration is still not finished before a new batch migration has to be started.

Migration on access

Migration on access was proposed by Mellor, Wheatley, and Sergeant (2002), where the transformation from the original format to a current one is only executed when the preserved digital information is accessed. Rosenthal, Lipkis, Robertson, et al. (2005) believe that a characteristic feature of migration on access is to have the original file that maintains its authenticity and integrity. Moreover, migration cost can be considerably reduced. However,

- Migration on access will cause additional delay while accessing.
- Migration on access will increase the workload for the preservation system.
- Migration on access should be closely integrated with the dissemination process.
- There must exist a file format converter from the original format to the current format.

**Further discussion**

The 10 preservation strategies presented above are further analysed in this section. Their intended use will be discussed later on.

**How to use the strategies to maintain storage accessibility?**

Table 3 (below) illustrates the advantages and disadvantages of the aforementioned strategies for maintaining accessibility. The strategies cannot be ranked and should have the same

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Auditing</strong></td>
<td>Cheap</td>
</tr>
<tr>
<td></td>
<td>Finds faults in time</td>
</tr>
<tr>
<td></td>
<td>Postpones refreshment</td>
</tr>
<tr>
<td><strong>Refreshment</strong></td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>High reliability</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Replication</strong></td>
<td>Simple</td>
</tr>
<tr>
<td></td>
<td>High reliability</td>
</tr>
<tr>
<td><strong>Federation</strong></td>
<td>Independence</td>
</tr>
<tr>
<td></td>
<td>High reliability</td>
</tr>
</tbody>
</table>

---

Table 3 Advantages and disadvantages of strategies for maintaining accessibility
priority in a preservation system. For instance, the current preservation policy of the National Library of Norway is to use replication and refreshment, where XML files, PDF files, TIFF files, and JPEG2000 files are created for a digital information object. All these files are packaged and stored in a long-term storage system that consists of two tapes and one RAID storage system at two different places. For refreshment, they replace some tapes and hard disks every 3–4 years, when the warranty expires.

Replication and refreshment are widely used by other preservation systems. However, since no audit application is used, threats like latent, irrecoverable faults might exist. The best preservation policy might be to use all four strategies together. Figure 3 illustrates such an approach where four strategies are deployed simultaneously for a preservation system.

**How to use the strategies to maintain bits manipulability?**

The strategies for maintaining bits manipulability lead to other priorities for a preservation system. The administrator of the preservation system usually chooses one strategy as the main one. Other strategies are only used for some special digital information objects in a short period.

The required material or documents for each strategy are illustrated in Table 4. In total, all the strategies need six types of materials and documents:

- **Change in bits** means for a given strategy how many times the bits of the digital object have been changed over its lifetime. Possible answers are ‘no change’, ‘changed once’ or ‘changed many times’.
- **Hardware specification** means whether a strategy should have hardware specifications. Possible answers are ‘no’ or ‘yes’.
- **Hardware component** means whether a strategy should have previous hardware components. The possible answers are ‘no’ or ‘yes’.
- **Format specification** refers to the type of format specification the strategy should have. Possible answers are ‘the original format specifications’, ‘the latest format specifications’, ‘the mediatory format specifications’ or ‘N/A’.
- **Format converter** refers to the type of format converter the strategy should have. Possible answers are ‘from the original format to a mediatory format’, ‘from the format currently used to a new format’, ‘from the original format to a new format’ or ‘N/A’.
- **Format interpreter** refers to the type of format interpreter application the strategy should have. Possible answers are ‘the original interpreter’, ‘the current interpreter’ or ‘the new interpreter’.

Based on these criteria, two dimensions, that is, the possibility of losing data and the difficulty of implementation, are used to rank the strategies.
The possibility of losing data depends on change in bits, format converter, and format interpreter. Change in bits and format converter include the format transformation procedure. If a format is not compatible with another format, some parts of the digital information object may be lost during format transformation. Moreover, old formats have a higher possibility of incomprehension. This is because the technique may be so old that little is known about it. Even the specification might be difficult to understand. Format interpreter includes developing interpretation procedures for a preserved format specification. The interpreter procedure might not work well if the format’s specification is not well documented or it is too complex to understand. Thereby, some content of the digital information may not be rendered.

The difficulty of implementation will depend on hardware specification, hardware component, format specification, format converter, and format interpreter. To implement any strategy, the first difficulty is to preserve relevant material and documents. The less material and documents a strategy needs, easier is its implementation. The second difficulty may lie in understanding the relevant technique specifications, including hardware specification and format specification. In general, the older a specification is, the more difficult it might be to understand. The third and final difficulty is the implementation target. For instance, emulation needs emulation applications; batch migration and migration on access need format transformation applications;

<table>
<thead>
<tr>
<th>Change in bits</th>
<th>Computer museum</th>
<th>Emulation</th>
<th>Encapsulation</th>
<th>UVC</th>
<th>Batch migration</th>
<th>Migration on access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware specification</td>
<td>No change</td>
<td>No change</td>
<td>No change</td>
<td>Changes once</td>
<td>Changes many times</td>
<td>Changes once</td>
</tr>
<tr>
<td>Hardware component</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Format specification</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Format converter</td>
<td>N/A</td>
<td>N/A</td>
<td>The original format's specifications</td>
<td>The mediatory format's specifications</td>
<td>The latest format's specifications</td>
<td>The original format's specifications</td>
</tr>
<tr>
<td>Format interpreter</td>
<td>The original interpreter</td>
<td>The original interpreter</td>
<td>The new interpreter</td>
<td>The new interpreter</td>
<td>The current interpreter</td>
<td>The current interpreter</td>
</tr>
</tbody>
</table>

Figure 4 Ranking the strategies for maintaining manipulability

(see Figure 4). The possibility of losing data depends on change in bits, format converter, and format interpreter. Change in bits and format converter include the format transformation procedure. If a format is not compatible with another format, some parts of the digital information object may be lost during format transformation. Moreover, old formats have a higher possibility of incomprehension. This is because the technique may be so old that little is known about it. Even the specification might be difficult to understand. Format interpreter includes developing interpretation procedures for a preserved format specification. The interpreter procedure might not work well if the format’s specification is not well documented or it is too complex to understand. Thereby, some content of the digital information may not be rendered.

The difficulty of implementation will depend on hardware specification, hardware component, format specification, format converter, and format interpreter. To implement any strategy, the first difficulty is to preserve relevant material and documents. The less material and documents a strategy needs, easier is its implementation. The second difficulty may lie in understanding the relevant technique specifications, including hardware specification and format specification. In general, the older a specification is, the more difficult it might be to understand. The third and final difficulty is the implementation target. For instance, emulation needs emulation applications; batch migration and migration on access need format transformation applications;
while encapsulation and UVC need format interpretation applications. From practical experience, it can be concluded that emulation application is the hardest to implement. Format transformation applications might also be difficult, since developers need to know the relation between the two formats. It seems that the easiest implementation is the format interpretation application.

Figure 5 further illustrates the above-mentioned preservation strategies with the help of a selection diagram. The selection depends on the preservation period, that is, one, several, and many computer generations. If the digital information is to be preserved for just one computer generation, then the computer museum might be the ideal choice as the necessary hardware and software may still be easily obtainable from the market. However, this strategy does not scale up for a longer preservation period, as compatible hardware and software alternatives become increasingly rare with time.

For preserving over many computer generations, batch migration to the latest computer technology seems to be the most viable solution. Currently, batch migration is the most widely used approach in long-term preservation. Researchers believe that migration could avoid loss of data. Wheatley (2001) summarized the batch migration activities for digitized materials of BBC. The Digital Preservation TestBed (2001) published a report on the practices of migration. Caplan (2007) introduced the migration solution in the Florida Digital Archive. Becker, Kulovits, Guttenbrunner, *et al.* (2009) described a systematic evaluation procedure to select the best format for batch migration. Distributed technology is also used for batch migration. For example, PANIC (Hunter and Choudhury 2006) and CRIB (Ferreira, Baptista, and Ramalho 2007) are two web migration projects, which use web

---

15 www.gdfr.info/
16 www.nationalarchives.gov.uk/PRONOM/Default.aspx

**Figure 5** Selection of preservation strategies for bits manipulability maintenance
may be chosen, as it does not require any format specification. If the preservation system has a format specification, but does not choose to preserve it, UVC might be chosen since the format will be converted to UVC instructions. If the digital information can be actively manipulated, the preservation system should use encapsulation.

Concluding remarks
Preservation systems are a set of complex information systems, containing an ingest function, an archive function, a dissemination function, a search function, a user interface function, and a management function. From the management perspective, we have identified three main challenges:

Very large volume of data
Gantz, Chute, Manfrediz, et al. (2008) released a survey report on the state of the digital world. They estimated that in 2007, the total size of the digital world was 281 Exabyte\(^1\), and that the annual growth in the digital world is approximately 60%. In 2011, the amount of digital information should be roughly equal to 1800 Exabyte. Moreover, they believe that the volume of digital objects might exceed the storage capacity. They have found that in 2008, the volume of information created exceeded the capability of storage media. If this is correct, then it can be expected that twice as much information will be generated in 2011.

Besides possible shortage in storage capacity, storage performance may be another challenge. In 2007, the National Library of Norway spent nearly three months to transfer all of its digital information objects to the new storage system. We tested the read/write performance of normal storage media, for example, a RAID storage system, a hard disk of desktops, and a memory stick. The RAID storage system had the highest performance because of its read-and-write speed of about 110 MB/second. The hard disk had read-and-write speeds of about 50 MB/second. The memory stick, with read-and-write speeds of about 10 MB/second and 1.5 MB/second, respectively, was the slowest. Thus, if all the 957 Terabyte of data in the National Library of Norway (in 2007) were to be transferred, 116 days, 232 days, and 7743 days would be respectively required for the RAID storage system, the hard disk, and the memory stick. Therefore, as the amount of digital information rises exponentially, there is real concern that the time required for the next system replacement will far exceed the support time, that is, the previous copying process of all data is still not finished before the next has to be initiated.

Lack of a comprehensive preservation strategy
Storage access and bits manipulability are two basic requirements. In order to ensure that the digital information objects are useful, two more requirements, that is, content understandability and object trustworthiness are needed. The content understandability requirement means that readers should obtain relevant context information to understand the meaning of the content of the digital information object. The object trustworthiness requirement means that readers should be able to consider the preserved digital information objects as accurate and authentic in terms of the preserved evidences.

Among all the above-mentioned preservation strategies, emulation seems to be the most suitable and easiest strategy that could satisfy these four requirements. However, this strategy is difficult to implement.

Regarding the other preservation strategies, especially batch migration and migration on access, most efforts are employed at the levels of accessibility and manipulability. Little research is directed towards understandability and trustworthiness issues, thereby indicating

\(^1\) 1 Exabyte = 1024\(^4\) Gigabyte = 1024\(^4\) Byte
that a more comprehensive strategy may be needed.

**What metadata should be preserved?**

Metadata are data about data. They provide supplementary information about a digital information object. Currently, there are numerous metadata standards for information management, such as the Dublin Core (The Dublin Core Metadata Initiative 2008), the Moreq 2 (Moreq 2008), the ISO 23081 (ISO 23081 2009), the PREMIS (the PREMIS Editorial Committee 2008), and the OAIS metadata model (Lavoie and Dale 2002). Even though researchers agree that metadata are essential for a successful preservation strategy, however, till date, there are only a few metadata standards proposed. We believe that metadata research for a long-term preservation system should also cover the following aspects:

- What metadata are necessary for preservation strategies?
- What metadata can improve authenticity of the preserved digital objects?
- What metadata should be recorded after a preservation strategy is carried out?
- How can the history of a preserved digital object be visualized?

**Conclusion**

Ingest, archive, disseminate, search, interface, and management are the important functions for any preservation system. This article focused on the management issues. We described and assessed 10 different preservation strategies. Auditing, refreshment, replication, and federation are often used together to keep bit integrity, that is, storage accessibility maintenance. Computer museum, emulation, encapsulation, UVC, batch migration, and migration on access are often used to offer bits manipulability. The objective of the preservation system is to determine which of those six strategies should be selected.

**Acknowledgement**

The research work for this article was funded by the Norwegian Research Council and our industry partners as part of the LongRec¹⁸ project.

---

**References**


¹⁸ www.longrec.com
A survey of digital preservation strategies


The Consultative Committee for Space Data Systems. 2002. \textit{The reference model for an Open Archival Information System (OAIS)}. Published by the CCSDS Secretariat Program Integration Division (Code M-3) National Aeronautics and Space Administration Washington, DC 20546, USA, 148 pp.


Abstract
This article describes how the new e-model evolved a new role for the NKC service group, how this service group contributes to the e-content lifecycle management process and how the new model ensures that NKC’s global e-library services are available to all associates regardless of location. It also provides details on how the service group empowers end users to locate needed resources and produces a positive impact on the company’s success.
Background

Novartis was formed in 1997 with the merger of two large Swiss firms, Sandoz and Ciba. Both firms were based in Basel, Switzerland and had large US operations in New Jersey. Answering the challenge of a competitive environment, Novartis transformed itself into a lean global company. Expertise was retained in centres of excellence, while duplicate functions and services were eliminated. Today, Novartis, with its nearly 100,000 employees in 140 countries, offers a wide range of healthcare products through the Pharmaceutical, Vaccines and Diagnostics, Sandoz, and Consumer Health Divisions. The Novartis Knowledge Center (NKC) is the only library and information service hub in the global Novartis community, consisting of five physical locations: Switzerland; New Jersey, Massachusetts, California in USA; and the UK. Novartis users can access e-library services 24x7 from anywhere in the world. The collection consists of electronic journals (3000+), e-books (4000+), and databases (nearly 200) delivered via the library portal.

The focus of this discussion will be on the NKC service group. The NKC information services portfolio includes Information Consultancy (traditionally called Expert Searching), Learning / Training, Document Service, and a Help Desk. During the past decade, these services have transformed to meet the challenges of the global business expansion and information technology innovations.

Evolution of user-enabled global e-library business model

Prior to 2002, NKC operated using a traditional library model. The expert searching, document services, learning/training, and help desk support were oriented towards the physical library collection.

- A user would come in to request an “expert search”, training or an article. The professional staff would answer them as requests were received.
- Search results were delivered as standalone data without special formatting or consideration for the user’s perspective or intended use of the material.
- Training sessions were limited to monthly orientations for new employees and vendor product training. Customized trainings for departments were rare.
- 85% of the collections were printed materials. The local online catalog (OPAC) included only printed material and was the only automated access system.
- Document service requests were filled by photocopying the printed materials in-house and delivered via “inter-office mail”.

The Novartis environment began to change rapidly in 2002. The corporation moved to a global organization. The implementation of advanced technology enabled NKC to transition from the traditional model to a hybrid (physical and electronic) model between 2002–2005.

In 2005, NKC committed to building a collection with 60% of the journals offered electronically and providing pharmaceutical industry news and databases to the end user’s desktop via a library Home Page. E-collections were still limited to local/country access. An immediate result was that fewer employees used the physical space and the printed material circulation was reduced sharply.

Another impact was that as more e-resources became available, the learning sessions were changing to be geared towards e-resource

![Figure 1](image-url)
Global e-library services in Novartis Knowledge Center

introduction and improvement of searching skills. The “expert search” requests became more sophisticated due to the availability of more electronic resources on the end users’ desk top.

The third impact was on the Document Delivery process, which became partially automated via Automated Literature Ordering system (ALOS). At this time, it was limited to Basel, Switzerland, the location of the headquarters.

In 2006, the hybrid model was migrated to a globally accessible e-model reaching all associates in different time zones. NKC set a goal to maximize the company’s investment and contribute actively to positive business results.

NKC downsized physical space and collections. The major content purchases became electronic, globally licensed resources to support associates worldwide 24x7. The e-library interface was restructured from a resource list approach to a search engine embedded in the collection of networked electronic information resources with full text linking.

**NKC user-enabled e-library business model**

As the Knowledge Center increased its global scope, NKC developed a new business model with three function groups bound in close interdependencies – Information Acquisition, acquiring all global e-resources; E-library, deploying e-resources to the end user through the NKC portal; and Information Delivery, providing the end user access to the content through consultation, learning programmes, and document services.

**NKC business model**

The e-content collections and technologies require a significant investment. To meet the emerging challenges and to ensure the best return on investment, the Information Delivery service group described in this model, plays a key role as the primary interface between the user community and the two other functions in terms of enabling users of the electronic library, facilitating the access to the targeted resources, and involvement in e-content cycle management.

**NKC Information Delivery**

A web-based user needs survey was conducted in 2008. The analysis of the data collected from the consultancy, learning programmes, and document services revealed a clearer picture of the customer’s business needs, types of requests, locations, gaps in the collection, and underserved populations. The analysis also revealed how well NKC was aligned with Novartis’ business priorities. With a solid baseline in place, a framework was created to measure the global e-library’s success. Success factors, determined

![Figure 2 NKC user-enabled e-library business model](image-url)
by business relevance and efficiency, measured what value NKC was generating for the end user.

**NKC global e-library service excellence**

With a baseline and success factors in place, we are now in a good position to improve and deliver on NKC e-library service excellence. We propose to:

1. **Optimize e-content utilization via the e-content lifecycle work stream**

   This is a collaborative process between all three NKC functions (Information Acquisitions, E-library Technology, and Information Delivery). To achieve the best possible terms of effectiveness and efficiency, the Information Delivery function is actively engaged in each step of this cycle.

   With a solid knowledge of resources and user requirements, this process enables the service group to interact with the users by being proactive, collecting feedback on what the users want and need, and how they want them delivered. Response is based on knowledge of user needs as opposed to being based on our assumptions. This approach provides a more targeted service delivery with higher quality and relevance.

   As the Knowledge Center improves the usability and findability of the resources in the NKC portal, the Information Delivery is creating a larger user base and increasing the utilization of our global licenses.

   **NKC** learning and strategic marketing programmes are penetrating the underserved population to add to awareness of NKC services and are promoting growth and use of the resources (see more in learning section).

2. **Maximize the business impact by combining e-content and information staff expertise by connecting to the company’s top priorities and the key customer groups**

   Building the customer relationship and the engagement with business objectives are key to bringing users to the e-library.

   As the e-library technology infrastructure advances and more e-content resources are made available, more focus is placed on the service requests for the top business priorities. These can be handled based on staff expertise regardless of geographical locations. For example, requests from Oncology and Vaccine divisions can be fulfilled by the information consultants in Massachusetts or California, while drug safety and regulatory requests can be handled by information consultants in New Jersey.

3. **Empower users through E-learning programmes**

   One of the survey results demonstrated the need for more self-paced learning programmes. To address these needs, the service group has collected e-tutorials from various databases and made them available under an “L” (learning) button next to the database name (see the screenshot below).
Since many of NKC customers work in global groups, NKC e-library services now offer more webinars instead of on-site training. To address the question of "too many resources", the service groups have been increasing the number of customized (distance) learning programmes to similar business functions in different geographical locations. To ensure the effectiveness of learning programmes, online surveys are sent out after each learning session. The collected pilot survey analysis has confirmed that the learning programmes have increased users' searching skills and enabled them to better utilize the e-library in their job performance. Further, orientations have created an awareness of other NKC resources, leading users to indicate that they will attend more in-depth subject sessions.

**Figure 5** KMC e-learning programme

**Figure 6** The ALOS service architecture: NKC ALOS flow chart (This service was listed as the top 3 NKC most valuable service in 2008 user needs survey.)
4. Meet the challenges of a “do-it-yourself system”: global Automated Literature Ordering Services (gALOS)

In 2007, NKC migrated from ALOS to gALOS. It has been utilized by more than 50 countries. To meet the ever-increasing demand (17% growth from 2007 to 2008) with no increase in resources, the service group has instituted features, which provide instant desktop delivery for in-house e-resources, auto check for copyright compliance, and provide documents in the format required by health authorities.

5. Global Help Desk

The Global Help Desk is remotely staffed by Switzerland and New Jersey personnel, (the two major locations). All associates from all sites are able to submit their requests anytime, anywhere in the world. The average turnaround time for this service is 0.4 hours.

To address the needs of “remote” or “field-based” associates, NKC has instituted a designated global service page in the portal, offering the same services available to onsite associates.

To make all these services and application tools easy to find, all the services have been indexed and are searchable via the portal’s search engine.

A glance at NKC’s e-library service requester’s characteristics

The following chart provides the breakdown by Novartis division of the use of NKC information consultancy and learning and training programmes in 2008 via the e-library infrastructure.

Conclusions

Our primary goal is to position NKC as an integral part of the decision-making process of the company by providing relevant and effective knowledge in this digital age.

NKC Information Delivery services meet the information management challenges and serve as a primary interface between the user community and two other NKC functions in the new e-model. The team:

- Partners with the key business customer groups to maximize the business impact
- Contributes to NKC’s ROI with resource knowledge and user’s requirements
- Empowers the users in improving their “finding skills” for more efficient and effective information utilization
- Grows the user market and pursues opportunities to facilitate and better serve Novartis business objectives

Figure 7 Information service requests by division

Figure 8 Document service requests by division
What is next?
Over the past few years, the Knowledge Center has adapted to changing business conditions and moved from a physical, site-bound organization to one that prides itself on its global reach. This has helped us realize that to be a centre of excellence, we must align ourselves proactively with changing business priorities and needs and take advantage of technology innovations.

With the creation of our critical success factors, we can now measure how all of NKC is living up to its mission of achieving excellence in the provision of global service. We will continue with our needs assessment survey every 3–5 years and conduct ROI studies. But, most importantly, we will use our consultancy skills to listen and adjust our portfolio according to the voice of the customer.
In this paper, we have focused on open source software in the field of libraries, and have highlighted the features of several useful open source library software.

**Introduction**

In the current digital era, majority of library services are information technology-based, with the resources available in digital and electronic formats. To manage all kinds of resources and facilitate their retrieval, libraries require high-quality integrated software, including cutting-edge retrieval tools. However, the steep price of such software prevents most of the libraries from using them. To tackle this issue, and for the benefit of scholars and the library community, various NGOs/organizations/individuals have developed software that are distributed free of cost. Known as free/open source software, these are widely available on the Internet and can be downloaded, installed, and distributed for free. Few examples include DSpace, Koha, NewGenLib, Green Stone, ABCD, E-Prints, Fedora, and so on.

There were several reasons behind the perceptible need for such software in library and information resource management. These were:

- Workload—speedy disposal of tasks;
- Multifaceted tasks;
- Upgradation of library services and systems;
- Changed formats of information resources such as e-resources, online resources, digital documents, and so on;
- Changing pattern in users’ demand for information;
- Preservation of information;
- Round-the-clock remote access to resources;
- Keeping track for administration of institutional research work;
- Revolutionary changes occurring in the publishing industry; and
- Exchange of information across different platforms.

In developed countries, the first steps towards library automation had been taken several years ago, in response to changing information demand and emergence of new trends in information management. Different software applications were developed either on contract basis by commercial software firms or by in-house teams of software developers and library professionals. Gradually, with the development of complete software solutions on the basis of expertise gained by interactions with the libraries, software development firms started making inroads into the library software market. These developments gave a concrete shape to the library software market. The market was further benefited by affluent organizations when they started earmarking significant amounts to software in their library budget.
However, the situation is different in underdeveloped nations and developing countries like India, where cost sensitivity is higher. In these countries, people care more about the cost of software than, say for example, an average American does (Behlendorf 2008). Moreover, libraries and information centres of these regions are known to face financial problems in matters related to infrastructural development, software applications, manpower deployment, and so on. Inspired by development initiatives in libraries of the USA, UK, and other developed nations, governments, library professionals, and academicians of developing and underdeveloped countries realized that in the emerging scenario, they would have to work towards infrastructural development of libraries and information centres so as to hasten educational, research, and development activities. Particular emphasis was given on aspects like library automation and application of different categories of software for library and information management. Many organizations, research centres, universities, and commercial organizations were eager to adopt software applications in their libraries. However, the high costs of library software—irrespective of whether these were bought from the market or developed in-house—proved to be a hindrance for most libraries, barring those few who had sufficient library budget and capability to invest considerable amounts in proprietary software. Till date, a large number of libraries in schools, colleges, and other institutions in countries like India are yet to start even the automation process, mainly due to the ever-increasing costs of proprietary software solutions and the equally steep costs associated with in-house software development.

**Investment on software**

Understanding the needs of libraries, new private software firms entering the market for library software began to make automation software available at affordable prices. Keeping in mind the ever-changing requirement affected by the emergence of latest technological trends, they kept improving and upgrading their software on a regular basis. These software firms also developed application software for modern library systems to fulfill prevalent needs such as building digital libraries, institutional repositories, managing online resources, and so on. With these innovations in library software business, proprietary software firms grabbed the market. As a result, the costs of proprietary software started rising, ultimately reaching a stage when only affluent organizations could afford these software applications. With only a few software firms developing these software, competition was less, thereby fuelling the rising prices. Some firms manufacturing low-cost library software did appear in the market, but being new in the field they were unable to immediately grasp the intricacies in functionalities, and thus could not fulfill the needs of libraries in an appropriate manner.

Hence, it is obvious that using proprietary software entails a financial burden for libraries with low budget due to the high costs of software, AMC charges of vendors, cost of upgraded versions, and so on. Details of a few premier library proprietary software available in the Indian library software market, along with their financial implications, are described in Table 1 (below).

As shown in Table 1, investment in proprietary software is not a plausible option for most low-budget libraries and information centres, who give the utmost consideration to the following economic factors while planning the use of proprietary software packages:

- Costly software;
- High AMC charges;
- Expenditure on customer supports;
- Cost of upgraded versions;
- Cost of accompanying software; and
- Delay in customer support.

**Emergence of open source software**

We tend to treat anything offered for free with suspicion. The suspicion results from a fear
of hidden costs. Free/open source software (F/OSS) is no exception to this rule. However, the fact is that such software is genuinely free for use under the Open Source Licence. A free software movement was started in 1984 by Richard M Stallman based on his personal belief in ‘freedom’ of information, and his increasing disillusionment with software vendors who were not supplying source code along with executable code for hardware devices such as printers (Elliot and Scacchi 2008). Virtual communities formed over the Internet are the main contributors to the development of F/OSS. These communities include individuals, researchers, academicians, government bodies, corporates, academic institutions, and so on. Development of F/OSS is the result of the economic and social movements by these communities in the software field. With time, this movement has spread all across the world, and has taken an organized form with the creation of different consortium and forum in support of OSS initiatives or projects. Perens (2005) has categorized the various open source contributors as:

- Volunteers;
- Linux distribution companies;
- Companies with a single open source programme as their main product;
- Companies for whom open source software enables sales of hardware or solutions;
- Businesses belonging to the service industry;
- End-user businesses and their contractors;
- Government bodies; and
- Academicians and scientific researchers.

As mentioned above, open source software is the result of collaborative efforts by individuals or communities bound together by similar kinds of jobs and software requirement. Usually, the initial development occurs in-house under a contract development paradigm. Eventually, the software is released to the public, generally before it is considered a ‘finished product’. Thus, OSS is made available lot earlier than a retail product (Perens 2005). Further, OSS users can modify the software, adding extra functions as per their needs.

Numerous writers have defined the term ‘free/open source software’ differently, but all these

Table 1  Software available in Indian library software market, with financial implications

<table>
<thead>
<tr>
<th>Sr. no.</th>
<th>Name of software</th>
<th>Brief description</th>
<th>Cost in ₹ (approx.)</th>
<th>AMC (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LIBSYS 4</td>
<td>All housekeeping activities, Web OPAC</td>
<td>0.4 million</td>
<td>10–12% of the software cost</td>
</tr>
<tr>
<td>2</td>
<td>Alice for Windows</td>
<td>All housekeeping activities, Web OPAC</td>
<td>0.26 million</td>
<td>10–15% of the software cost</td>
</tr>
<tr>
<td>3</td>
<td>NetLib</td>
<td>All housekeeping activities, Web OPAC</td>
<td>0.1 million</td>
<td>10% of the software cost</td>
</tr>
<tr>
<td>4</td>
<td>Liberty 3</td>
<td>All housekeeping activities, Web OPAC, RSS Feed, Z39.50</td>
<td>0.45 million (for 3 users)</td>
<td>10–15% of the software cost</td>
</tr>
<tr>
<td>5</td>
<td>VTLS</td>
<td>All housekeeping activities, Web OPAC, RSS Feed, Z39.50, RFID Integration, URL Locator</td>
<td>2.9 million (for 8 users)</td>
<td>15–18% of the software cost</td>
</tr>
</tbody>
</table>

Source: User libraries of each of the above-mentioned software.
definitions tend to have the same meaning. OSS programmes are those whose licence permits users the freedom to run the programme for any purpose, study and modify the programme, and freely redistribute copies of the original or modified programme (without having to pay royalties to previous developers) (Wheeler 2007).

The definition of ‘free’ software, as mentioned in the GNU website, clears any confusion about the term. It states that ‘free’ means:

- The freedom to run a programme for any purpose;
- The freedom to study how the programme works and adapt it to one’s needs. Access to the source code is a precondition in this regard;
- The freedom to redistribute copies of the programme; and
- The freedom to improve the programme, and release the improvements to the public so that the whole community benefits. Access to the source code is the precondition in this regard.

Source: www.gnu.org/philosophy/free-sw.html

It should also be mentioned that the terms ‘free software’ and ‘open source software’ have slightly different meanings. However, as far as this discussion is concerned, we shall not be delving into details on these aspects.

To understand the concept of free/open source software, one first needs to understand what is meant by ‘software development’. Software, or a software programme for any specific task is designed by writing the source code using a scripting language. This is known as ‘software programming’, which is usually the job of a computer programmer. The source code enables a computer to perform a specific task by translating the source code into an executable programme. As long as access to the source code is restricted, the programme cannot be modified. Moreover, such a programme can only be installed and used for the specific task it is designed for. Software developers must permit access to the source code for the sake of making changes, modifications, and any other improvement. Viewing and access to source codes of proprietary/commercial software are restricted under the licence of their use. On the other hand, open source licence provides freedom to access the source code of a programme. However, access to the source code is simply a basic characteristic of OSS, and is not enough to define such software. According to the website of Open Source Initiative, the following criteria must be complied with in the case of OSS:

- Free redistribution;
- Source code sharing;
- Derived works;
- Integrity of the author’s source code;
- No discrimination against any person or groups;
- No discrimination against fields of endeavour;
- Distribution of licence;
- Licence must not be specific to a product;
- Licence must not restrict other software; and
- Licence must be technology-neutral.

‘Open source’ refers only to the legal terms of the distribution licence (Dorman 2005). Thus, open source software refers to software that is distributed under the open source licence, where the term ‘open source’ refers to the legal terms of the licence. There are two basic types of open source software license, namely, GNU public licence (GPL) and lesser (library) general public licence (LGPL). Hein (2004) has described these licences as follows:

- GPL: It emphasizes that original work and derived works must remain free. GPL is ideal for developers who wish to prevent inclusion for a proprietary or commercial product.
- LGPL: In this licence, the internal code is GPL, but the software that interacts through published interfaces is not considered as derived work. It permits use of LGPL libraries in proprietary or commercial products.

Another important open source licence is MIT/BSD/Apache. Most of the other open source
licences, including IBM, Intel, Mozilla, Apple,
Nokia, Sun, Lucent, and a few others are based on
the above-mentioned basic licences.

**OSS in the field of library and
information management**

In comparison to other sectors, the emergence
of F/OSS in the field of library science and
information management is a more viable option
because cooperation and coordination are the
key issues in library science. Library professionals
have always focused on cooperation, resource
sharing, consortium, open access, open standards,
open archives initiatives, and so on in order to
help each other in collection development and
implementation of tools and technologies, among
others. This attitude and the prevalent economic
situation have facilitated the development of F/
OSS in the field of library science and information
management.

In the initial stage of development, due to
lack of awareness on technical aspects, right of
use, sustainability, and so on, there was a lot of
confusion among librarians and others about F/
OSS. For instance, several arguments have been
put forward that free software, by its very nature,
is not well supported, and can only be installed
and made ready to use by computer experts
(Dorman 2004). This misconception prevented
non-computer professionals and librarians to opt
for this type of software, who preferred the easy-
to-use commercial software that had user-friendly
interface.

But, now F/OSS is on its way towards
becoming the preeminent solution due to the
various emerging problems related to proprietary
software. Library professionals have started to
take advantage of F/OSS for their requirements
without any fear. Similarly, numerous libraries
and information centres are now successfully
using F/OSS for fulfilling their requirements.
Open Source Software for different uses in library
and information management are now being
made available by various organizations and other
institutions. The different categories of software
in the field of library and information resource
management in which open source software are
being developed are as under:

- Software for library automation (bibliographic
database);
- Software for digital library management;
- Software for repository management;
- Software for e-resource management;
- Software for bibliography management; and
- Software for thesaurus.

Not only can OSS in the above categories be
used easily without any technical expertise, it
can also be modified, upgraded, and customized
to fulfill any other requirement with the help of
a little knowledge of scripting languages. Today,
numerous projects are being carried out towards
the development of open source software in the
field of library and information management,
thus making it possible to choose the one best
suited for oneself. Some of these OSS are described
below:

**KOHA**

KOHA has the distinction of being the first open
source integrated library management system,
which includes all the main functions related
to library management. It is a web-based open
source software distributed under the general
public licence. The KOHA Development Team
offers to host the website for KOHA library
system on its server. KOHA also has the capacity
to manage digital libraries and online and offline
electronic resources.

**Features:**
- Comprises basic library functions such as
  acquisition, cataloguing, serial control,
circulation, Web OPAC, and so on;
- Simple and comprehensive acquisition
  options;
- Facilitates RSS feed of new acquisitions;
- Tailored catalogue module for special libraries;
- E-mail and/or text patron’s overdues and other
  notices;
- Simple, clear search interface for all users;
• Supports standards and protocols like MARC21, UNIMARC, Z39.50, and so on;
• Same tool for managing online and offline resources;
• Easy barcode printing;
• Allows multitasking and enables updates of circulation, cataloguing, and so on; and
• Supports Linux, Unix, Windows, and Mac OS platforms.

**ABCD**

ABCD, a popular library management software, is promoted and coordinated by BIREME, with support from VLIR. ABCD is a web-based software for the ISIS community that comprises the main library functions. The first version of ABCD (v1.0) was released on 5 December 2009.

**Features:**
• Published as free and open source software with accompanying tools for the developer community;
• Comprises basic library functions such as acquisition, cataloguing, data entry, circulation, Web OPAC, and so on;
• Integrated with basic statistics package for various types of report generation, statistical functions, and a powerful circulation system;
• Compatible with programming languages accepted by the GNU licences such as PHP, Java, JavaScript, Python, and so on. The current version of ABCD is written in PHP v.5 and IsisScript;
• Multilingual system in four languages, namely, Spanish, English, French, and Portuguese. Can also be developed in other languages;
• Uses MARC-21 and supports Dublin Core, METS, and Z39.50;
• Any number of databases can be created, while existing databases of WinISIS can also be copied;
• Compatible with CDS/ISIS database technology for bibliographic databases;
• Allows easy print generation, printing of barcodes, SDI services through user profile, and other online services; and
• Runs on both Windows and Linux platforms.

**NewGenLib**

NewGenLib is another integrated library management system distributed free under the GNU general public licence. This software is the result of collaborative efforts of Kesavan Institute of Information and Knowledge Management (KIICM), a professional charitable trust and Versus Solutions Pvt. Ltd (VSPL), a software development company. Before 2007, when it became an open source product, it was already in use in 122 libraries of India, Syria, Sudan, and Cambodia as a commercial product.

**Features:**
• Available under GNU GPL v3;
• Functional modules are completely web-based; uses Java Web Start™ technology;
• Compatible with MARC-21, MARC-XML, z39.50, SRU/W, and OAI-PMH;
• Uses mainly open source components;
• Scalable, manageable, and efficient;
• Uses z39.50 client for federated searching;
• Is an internationalized application (I18N);
• Unicode 3.0 complaint;
• Arabic version available;
• Easily extensible to support other languages;
• Data entry, storage, and retrieval in any (Unicode 3.0) language;
• RFID integration; and
• Supports both Windows and Linux.

**D-Space**

D-Space, a system for management of institutional repository, was developed jointly by MIT Libraries and Hewlett-Packard Labs in 2002. It is distributed under the BSD Open Source Licence, with a few restrictions. D-Space has become the choice of academic, commercial, and other organizations when it comes to maintaining institutional repositories due to its ability to efficiently manage different types of materials.
like books, research papers, theses, preprint, technical reports, images, videos, and so on. As per the Directory of Open Access Repositories, as on 6 December 2010, a total of 1,805 open access repositories had been registered worldwide, out of which 659 repositories (37%) are using D-Space.

**Features:**
- Written in Java, with PostgreSQL, Lucene, and Apache/Tomcat;
- The system is organized into communities, sub-communities, and collections;
- Supports all types of digital formats, including books, theses, data sets, computer programmes, bibliographic datasets, images, audio files, video files, learning objects, web pages, and so on;
- Allows submission of multiple formats of the same. For example, a TIFF file and a GIF file of the same image;
- Multi-user system for both maintenance and searching;
- Long-term physical storage and management of digital items in a secure, professionally managed repository, with important functions such as backup, mirroring, disaster recovery, and so on;
- All records have a persistent identifier;
- Access control over items in repository at collection and individual item levels;
- Supports standards such as Dublin Core, OAI PMH V2.0, UNICODE, and so on;
- Allows easy migration of items in the system across newer versions;
- Able to interoperate other systems in the organization; and
- Allows customization of subsystems as per requirement.

**E-Prints**
E-prints was the first freely available open source software (since 2000) for building high-quality and high-value institutional repositories. It was developed at the University of Southampton’s (UK) School of Electronics and Computer Science, and distributed under the GNU licence. It is one of the most widely used repository software, and enjoys the largest installed base among this category of software.

According to the Directory of Open Access Repositories, as on 6 December 2010, out of a total of 1,805 open access repositories that had been registered worldwide, 296 repositories (16%) were using E-Prints.

**Features:**
- Easiest and fastest way to set up repositories of research outputs of literature, scientific data, theses, and reports or multimedia artifacts;
- Easy to install, configure, and maintenance;
- Supports huge number of digital formats such as research papers, theses, patents, audio, video, images, scientific data, fine arts compositions, exhibition, and teaching resources;
- Facilitates browsing and viewing of records by any complex criteria;
- Basic and advanced search facilities;
- Efficient management of bibliography;
- Provides RSS feed for the entire repository;
- Maintains quality of metadata;
- Fast deposits from disk or directly from the web and quick import of data from other repositories;
- Warning on deposition of duplicate or similar records;
- Flexible plug-in architecture for developing extensions;
- Search output can be exported in different formats such as METS, Dublin Core, and other formats;
- Special facilities for registered users;
- Written in Perl, with MySQL and Apache;
- Multi-language support; and
- Is OAI compliant.

**Greenstone**
Greenstone is an OSS for building and distributing digital library collections. It helps in organizing and publishing information on the web as well as in CD-ROM. Greenstone has been developed
by the New Zealand Digital Library Project at Waikato University, and is distributed and promoted in cooperation with UNESCO and Human Info, an NGO. It is distributed under the GNU general public licence.

As per the Directory of Open Access Repositories, as on 6 December 2010, out of a total of 1,805 open access repositories registered worldwide, 24 (1%) were using Greenstone.

Features:
- Supports various operating systems such as all versions of Windows and Linux, Sun Solaris, and Mac OSX;
- Different interfaces for user choice for collection-building like command mode, web, and Java-based GUI interface;
- Multimedia and multilingual support;
- Content development in three alternate ways;
- Indexing of terms from full text of documents and metadata associated with the documents;
- Variety of search and browse options and customization there of, as per requirement;
- Structured metadata in XML using Dublin Core;
- Extraction of existing metadata, already associated with a document;
- Supports various type of file formats such as HTML, PDF, DOC, RTF, e-mail, plain text, PPT, image, video, and so on;
- Use of plug-in for converting the file format in to standard XML-based internal format for indexing purposes;
- Interoperability and OAI compliant;
- Tight administrative functions, allows password protection; and
- Customization of various features fulfilling specific user requirements.

**FEDORA (Digital Library)**

FEDORA stands for Flexible Extensible Digital Object Repository Architecture. It was initially developed by Sandra Payette and her team in 1997 as a research project funded by DARPA and NSF at Cornell University. After several architectural modifications, version 2.1 was released in 2005. The current version is 3.1. Presently, the project is funded by the Andrew W Mellon Foundation and the Gordon and Betty Moore Foundation, and is directed by Sandy Payette from Cornell and Thornton Staples from the University of Virginia.

Features:
- Distributed under the terms of the Apache licence;
- FEDORA supports ingest and export of digital objects in a variety of XML formats;
- The FEDORA digital object model allows tight management of metadata and digital content, irrespective of format;
- FEDORA server architecture is based on four main application programming interfaces (APIs): manage, access, search, and the Open Archival Initiative service;
- Digital objects are stored as XML-encoded files that conform to an extension of the METS schema;
- FEDORA digital object has a primary Dublin Core record;
- The OAI Protocol for metadata harvesting is a standard for sharing metadata across repositories;
- FEDORA repository system provides a search interface for both full text and field-specific queries across metadata fields;
- FEDORA repository system includes a batch utility as part of the management client that enables mass creation and loading of data objects;
- Comprises component management module and parameterized disseminators; and
- Access control and authentication based on IP address.

**Advantages of F/OSS**
- Free/Open Source Software are available absolutely free of cost to anyone who needs it. With just the basic IT infrastructure in place and after investing a small amount, libraries can go ahead in using these software;
Most of the accompanying software required for operating any F/OSS are also distributed free of cost under open source licences;

- Library professionals can access the source code to understand the functionalities, and modify the same if they want additional features;
- User communities and developers of most of the F/OSS are always ready to offer solutions to any technical problem related to the software. No dependency on commercial experts for customer supports;
- No delay in getting answer for queries;
- Manuals and relevant documentation for the F/OSS are also distributed by the developers of the software;
- Prevents the misuse of monopoly positions in the field of software;
- Gives full right to use the software in any way and redistribute it after modification;

- Most of the F/OSS support international standards with features of interoperability across different platforms;
- F/OSS can be tested for the requirement of any library and replaced with another in case it is found to be unsuitable to the specific context. As such, the most suitable software can be selected without investing money on replacing existing software; and
- The use of F/OSS in libraries also encourages library professionals in acquiring knowledge on software programming and generating awareness in the field of information and communication technology.

**Conclusion**

With the open source software movement gaining steam, the world is now enjoying the advantages of collaboration and cooperation in software development. Open source software are slowly but surely gaining in popularity among developers and user communities. Educational institutions, research organizations, government enterprises, NGOs, and libraries have all started using operating systems, web servers, library management systems, institutional repositories, course management systems, content management systems, and so on—all belonging to the category of open source software for various purposes. OSS has, in fact, become a trendsetter in the arena of software development and distribution.

Libraries with small budgets have always considered automation of housekeeping operations as a financial burden due to the high cost of software. However, development of open source software is an effective way to automate library operations without undertaking substantial financial investment. Librarians need to understand open source licence for promoting the use of OSS. This is the only way to face the challenges posed by commercial software in the market. It will also increase the autonomy and control of these professionals over software solutions. In conclusion, the advent of open source library software has ushered in a revolution in the field of library and information resource management, and has became a popular choice for most library and information science professionals because of its numerous benefits and useful features.

**References**

ABCD. (Details available at http://www.librarytechnology.org/ltg-displaytext.pl?RC=14406, last accessed on 6 December 2010).


General Public Licence. (Details available at www.gnu.org/licences/gpl.html, Last accessed on 3 February 2010).


Usage of Open Access Repository Software. (Details available at http://www.opendoar.org/, last accessed on 6 December 2010).


http://www.koha.org
Imperativeness of e-learning in a digital environment in an organization

R K Bhatt, Associate Professor and Head, Department of Library and Information Science, University of Delhi


Introduction

E-learning has opened up new avenues for learners to pursue education as per individual choice and pace by eliminating the geographical and traditional impediments. Worldwide, the opportunities to pursue educational objectives through e-learning are emerging at a very rapid rate and in many different forms. These are variously described as e-courses, web courses, virtual learning, online courses, digital courses, and so on. Learners are attracted by the convenience. Despite these benefits, the process has not yet matured to attain the level of confidence commanded by the conventional system of teaching and learning.

Digital technology enhances access processing and dissemination of information and knowledge. Institutions typically select treasures for digitization not because they are in high demand, but by digitizing they seek to make their valuables better known and more widely accessible. Thus, the relevant considerations when selecting materials for digitization are not primarily about preserving the original, but enabling its easy access. In the digital world, preservation is the act of safeguarding access to high quality, high value, well-protected, and fully integrated version of an original document.

Learning: conceptional framework

‘Learning’ denotes the acquisition of knowledge or skill by instruction or self-study. Psychologically, ‘learning’ is defined as a relatively permanent change, caused by experience, in an organism’s behaviour. Its study is central to psychology, interacting with the studies of concept formation and attainment, decision theory, perception, problem solving, reasoning, and development. It is one of the oldest in psychology, beginning with ancient Greek philosophical speculation about how humans acquire knowledge (Lynch 1994).

Libraries have always been used by instructors and learners to search and obtain learning resources. Even though libraries store valuable materials, till recently, their contributions to learning activities were not truly acknowledged. The most important reason was lack of public recognition of the importance of library materials for education. Secondly, majority of learners earned their certificates and/or degrees in physical classrooms. Even though visits to libraries were certainly beneficial to the learners’ knowledge accumulation, but such visits were not compulsory. The third reason was e-learning, virtual classrooms, and electronic collaboration. These comprise the delivery of content via Internet,
Intranet/Extranet (LAN/WAN), audio and videotapes, satellite broadcast, interactive TV, and CD-ROM. According to Stokes (1994) e-learning ‘is a means of becoming literate involving new mechanisms for communication: computer networks, content portals, search engines, electronic libraries, distance learning, and Web-enabled classrooms. E-learning is characterized by speed, technological transformation, and mediated human interactions’.

As per the Department of Education and Skills, UK (2002), e-learning include ‘a range of activities, from the effective use of electronic resources and learning technologies in the classroom, to a personal learning experience enabled through individual access at home or elsewhere.’

The definition of e-learning (or electronic learning) in Wikipedia reads: (it) is a term that encompasses all forms of technology-enhanced learning (TEL) or very specific type of TEL such as online or web-based learning.

E-learning and the e-learning library

While concept of e-learning is gaining popularity, the question ‘What is e-learning?’ ought to be answered. The following are the definitions given by different people and organizations. According to Kaplan-Leiserson (2001), e-learning ‘covers a wide set of applications and processes such as web-based learning, computer-based applications, processes and services, to which the electronic library is able to make contributions, such as:

- On-demand e-learning: providing learners with on-demand learning materials through electronic full-text and/or multimedia databases, the electronic document delivery service, VOD, and so on;
- Live online e-learning: providing synchronous programmes between a reference librarian and users through WAP technology;
- Knowledge-based packages: constructing electronic databases containing learning contents in a searchable environment;
- Simulation-based learning: providing learners with interactive contents to learn on a simulation playing field. This can be done, for example, through virtual reality design in a digital library.

A library supporting e-learning should identify usable learning resources and services for each of the above contributions, and then provide learning materials and seamless, integrated access to a range of resources across media and across boundaries of curatorial tradition (Johnston 2001).

Elements of e-learning and their use

To better utilize e-learning, we need to know the tools and techniques associated with it. These include:

E-mail

E-mail is the foundation for all forms of online learning and teaching. It is a highly appreciable tool of e-learning. For instance, e-mail-based discussion forums play an important role in the e-learning process.

Example: LIS-forum (moderated discussion forum)

http://www.ncsi.iisc.ernet.in/ncsi/services/lisforum.html is a good example depicting library and information professionals sharing their knowledge with other people through e-mail.

Real-time conferencing

It covers any form of online synchronous interaction. One of the simplest forms of real-time conference is online chatting. Here, participants exchange typed messages with using a common interface. Each message is preceded by the name of the sender to identify who said what. Instant messaging software provides interesting features for communication or sharing of knowledge. The use of file transfer protocol (FTP) gives it an edge by allowing transfer of files during an ongoing conference. Yahoo messenger, MSN messenger, and Google’s GTalk all have this feature.
Multi-user domains (MUDs) and object-oriented MUDs (MOOs)

These are an interesting category of real-time conferences (virtual conferences) specifically designed to facilitate group interaction. MUD/MOOS allow several people to share a virtual world, usually set up as ‘rooms’ containing objects, which can be viewed or manipulated. People can interact with each other by sending chat messages as well as by performing simulated actions. The main features are:

• Real-time collaboration;
• Interactive data sharing;
• Instantaneous communication;
• Multi-location; and
• Share virtually anything.

An excellent example is eZmeeting (http://www.ezmeeting.com). eZmeeting is a low-cost, no-maintenance document collaboration application. With eZmeeting, up to 32 people together, on any PC, can view, mark, and comment on any document in real time. The comments made, the changes implemented, and the fixes are displayed on all the participants’ computers as they happen (Elmer 1999).

Desktop video

It is the most advanced form of real-time conferencing. A desktop video system is basically a chat system that uses video images instead of text messages. The video images (including audio) are captured by a small digital camera that is connected to the PC. These cameras are relatively inexpensive and can be connected to any computer (including laptops). Using software—that comes with the camera or is obtained separately—it is possible to connect to a server running a video conferencing programme or make a direct link with another person using their IP address for a two-person video session (called a point-to-point connection).

Essentials for e-learning

Interactivity and participation

The most important factor in e-learning is to ensure a high degree of interactivity and participation. This entails designing and conducting learning activities in such a way that it results in engagement with the subject matter and fellow students. Coursework should focus on assignments and projects that are relevant and realistic in nature. There should be plenty of opportunities for input from the instructor and fellow students. What strongly affects the amount of student interaction and participation is the level of instructor involvement. If the instructor regularly posts messages in the discussion forum or provides comments to students via e-mail, it increases student involvement and participation in a course. So, a cardinal rule of good online teaching is that the instructor must participate to get students to do likewise.

Feedback

A primary task of the teacher is to provide feedback. In e-learning, teacher’s feedbacks or comments appear as e-mail messages. The files submitted by the students are corrected, and the students can download these files to view the comments. Ideally, individual feedback should be provided to each student, as well as group feedback. Group feedback can take the form of messages posted in a discussion forum or during a conference, which summarize/synthesize the individual responses on a given topic or activity.

Moderating and facilitating

E-learning requires sound moderation and facilitation skills. Moderation involves encouraging students to participate in discussion forums and conferences, while ensuring that some of them do not end up dominating the discussion, keeping discussions focused on the topic at hand, and summarizing/synthesizing the highlights...
of discussions. Facilitation entails providing information that will help students complete their assignments, suggesting ideas or strategies for them to pursue in their course work, and getting students to reflect on their responses and work.

Faculty collaboration
E-learning offers various opportunities for student interaction. It also provides numerous possibilities for collaboration among teachers and students. The following kinds of collaborations can be thought of:

- Teacher-teacher collaboration;
- Teacher-student collaboration; and
- Student-student collaboration.

There occurs no face-to-face interaction as in a traditional classroom scenario. Hence, nobody feels any hesitation to present his/her opinion (Sahu and Singhal 2002).

Technologies and tools for e-learning
E-learning set up in the context of training in the country must encompass the minimum requirements as given below (Rama Devi 2006):

Content development
Courseware must be prepared using simple authoring tools such as HTML or XML. The course content may include text-only materials such as theme papers, reports, conference/journal papers or PPT presentations or even live recordings of classroom teaching/lectures delivered in conferences by attending or through video-conferencing. The training divisions have to seriously engage in concrete action plans to record and preserve the class lectures and other output generated by guest faculty.

Content serving
The content created as per the earlier step must be either offered in CDs, loaded in standalone PCs for sharing through Intranet/Internet. Care should be taken to keep materials in such formats from occupying a lot of space. Collaborative sharing of information resources can be explored in this context.

Content access
Learners can access the courseware through standard web browsers and multimedia players. Many institutions and agencies are engaged in training in the form of short-term courses, workshops, tutorials, and so on. Most of these courses come up with a printed volume of study materials. The institutions organizing such courses and the funding agencies sponsoring the same should encourage preparing the course materials in electronic form for easy distribution and allow network access to these resources.

Databases
Database of different lectures and indexes can be prepared to enable faster search and access to select materials. It must also contain profiles of teachers/guest faculties and learners to evolve personalized learning. The database should be developed in regional languages with particular emphasis on locally relevant information. The following variables have to be taken into account while selecting technology for a given training application.

Challenges faced by e-learners
Development of e-learning has led to emergence of new problems related to copyright and intellectual property rights' implications of electronic text. Students, researchers, staff, employees, and other end-users affiliated with virtual universities or digital libraries should be allowed to print (on paper) excerpts of digitally available works, on the same conditions as are followed while photocopying print material. Security is the most pressing challenge when it comes to digital media. Piracy of database, viral invasions, and parallel satellite networking stress are some of the issues that digital libraries are confronted with on a regular basis (Jayaprakash and Venkatramana 2006).
The other major challenges are:

• There is no mechanism available to establish standards for online materials, instruction, design, and quality of interaction;
• Study materials are accessible only by specified students;
• Since course materials are instructionally designed, they hardly provide for individual variations and further revision;
• Dangers of increased learner isolation as students learn from the screen, and not through interaction with their peers and teachers;
• Crossing national boundaries leads to logistical and organizational problems for distance-learning institutions. For instance, the facilities available to and aptitude level of European students and developing countries are different);
• Operating overseas can also expose one to all the problems of international business, exchange-rate fluctuations, restriction on foreign exchange and the export of money from the country of operation to pay for services, sources from another jurisdiction, political turmoil, civil unrest and war, and so on;
• Information providers are more interested in profit than in quality services;
• Lack of organization of information on the Internet;
• Not all sites are updated regularly;
• Absence of monitoring mechanism to evaluate the course ware;
• Lack of awareness about the use of electronic equipment (Ram 2008);
• Lack of human interaction: It is difficult to judge how much a student has understood of a particular topic. In teacher-student face-to-face interaction, the teacher can gauge the level of understanding of each student and accordingly s/he can try to explain the problem with suitable examples;
• Lack of expertise: Not many vendors/experts are available in the country and even abroad. Overseas vendors charge a lot and are also reluctant to import techniques/technology;
• Access to Internet in developing countries including India may not be as easy or widespread as that in developed countries; and
• Lack of motivation: In a traditional classroom, the teacher and students interact spontaneously while discussing and understanding a topic, which creates motivation among the students towards learning. In e-learning, lack of motivation might lead to the topic at hand appearing dull.

Conclusion

E-learning in the digital network environment enables students, teachers and researchers to access any learning resource anytime and from any place. The networked digital environment has rapidly transformed the traditional means and manner of learning and teaching. Networked learning environment is an exciting idea, but there are a number of challenges before it can become a reality. These include finance, infrastructure, skills, and so on. Digital libraries help academicians and students by providing easy access to wide range of reference material, and also aims at encouraging better use of information resources available on the Internet in the digitized form. It is time to harness technology and deliver the goods for the learners’ community.

References


The UF Libraries Latin American Collection
The UF Libraries’ Latin American Collection contains one of the finest collections of Latin American materials in the USA. It consists of over 500,000 volumes, some 50,000 reels of microfilm (many unique and very scarce), renowned newspaper and government-document holdings, and a growing access to computer-based electronic information resources. Areas of collection focus include all disciplines, although literature, the humanities and the social sciences are best represented. All regions of Latin America are also well represented, with the Caribbean, Circum-Caribbean and Brazil having the deepest holdings. Materials on women’s issues are strong.

Other units of the UF Libraries also contain important resources and researchers are encouraged to utilize them as well. The UF Map Library houses approximately 500,000 maps and atlases, some 50,000 of which deal with Latin American topics.

Source: http://www.uflib.ufl.edu/lac..

News Release: Digital forensics and born-digital content in cultural heritage collections
by Matthew G. Kirschenbaum, Richard Ovenden, Gabriela Redwine
While the purview of digital forensics was once specialized to fields of law enforcement, computer security, and national defense, the increasing ubiquity of computers and electronic devices means that digital forensics is now used in a wide variety of cases and circumstances. Most records today are born digital, and libraries and other collecting institutions increasingly receive computer storage media as part of their acquisition of “papers” from writers, scholars, scientists, musicians, and public figures. This poses new challenges to librarians, archivists, and curators to accessing and preserving legacy formats, recovering data and ensuring authenticity. The methods and tools developed by forensics experts represent a novel approach to these demands. For example, the same forensics software that indexes a criminal suspect’s hard drive allows the archivist to prepare a comprehensive manifest of the electronic files a donor has turned over for accession.

Source: http://www.clir.org/pubs/reports/pub149/pub149.pdf

RAPTOR: Understanding your users’ use of e-resources
Understanding e-resource usage in an institution is of utmost importance to decide which resources they need to keep subscribing to, the JISC-funded RAPTOR (reporting, analysis,
and presentation toolkit for usage of online resources) project is building a software toolkit for the purpose. The reporting e-resource usage statistics will be done by analyzing log files from relevant software systems. Initially, supported software will be the Shibboleth Identity Provider, EZProxy, and OpenAthens LA, but the RAPTOR software is easily extensible so should be able to work with any such system. It can also be extended to able to interface with local identity information (e.g., through LDAP) to pull in additional information for enhanced reporting. The RAPTOR software is designed such a way to be easy for system administrators to install and configure, and its web interface is easy to use for both technical and non-technical staff. The RAPTOR project is due to complete and release its first full release in April 2011.

For more information on RAPTOR, visit http://iam.cf.ac.uk/RAPTOR.

Source: D-lib Magazine, January/February 2011

President Obama Signs Museum and Library Services Act of 2010
The President signed into law the Museum and Library Services Act of 2010 (S. 3984) on December 22, 2010. The new law reauthorizes
- The existing programs of the Institute of Museum and Library Services (IMLS) with some important changes.
- The updated language calls on IMLS to take an active role in research and data collection and to advise the President and Congress on museum, library, and information services.

- This Act also clearly recognizes how libraries and museums contribute to a competitive workforce and engaged citizenry.

The legislation advances the roles of libraries and museums in education, lifelong learning, preservation and workforce development.
IMLS has been particularly active in providing leadership to align the activities of libraries and museums around our national priority for an educated workforce.

Source: D-lib Magazine, January/February 2011

National Archives contributes to improved digital records preservation and access system
The National Archives, USA has contributed to the update of a ground-breaking system that supports long-term preservation of and access to electronic records. The ‘new and improved’ version of this ‘PRONOM’ system was developed in partnership with the National Archives of the United Kingdom and the Georgia Tech Research Institute.

PRONOM is a web-based public technical registry of more than 750 different digital file formats that enables digital archivists, records managers and the public to precisely identify and confirm digital file formats. This identification is the first step to ensuring long-term electronic file preservation by enabling the identification of those file formats that are in danger of becoming obsolete.

Source: D-lib Magazine, January/February 2011