Open Source Software: Boon to Digital Libraries

Sangeeta N. Dhamdhere¹ and Ramdas Lihitkar²

¹Modern College of Arts, Science and Commerce, Ganesh-Khind, Pune, Maharashtra, India
²Government College of Science, Nagpur, Maharashtra, India

Correspondence should be addressed to Sangeeta N. Dhamdhere, modernlibrary.sangeeta@gmail.com

Publication Date: 22 June 2013


Abstract ABCD – Automation of Libraries and Centers of Documentations is a relatively new Free Open Source Software available to all libraries. It caters almost all present needs of three dimensional libraries of all sizes. This software is boon to all types of libraries those are still under automation and cannot afford to purchase commercial library software. It gives excellent content management solution and Interactive library website as part of the ABCD integrated library automation suite. It has excellent indexing and retrieval features with integrated meta-search and content management system to manage online, offline digital as well as physical resources. Librarians can create their library website without knowledge of HTML programming. Many libraries in the South, esp. Latin America, and Africa are using it very effectively. This paper gives short introduction to the ABCD software and its digital library features along with illustrations.

Keywords Open Source Software, FOSS, Digital Library Software, Library Website, Content Management Software

1. Introduction

Library 2.0 is using web 2.0 technologies in the libraries. Open Access Initiatives and Electronic Publishing have made impact on the libraries and their services, facilities. Being free and of very high quality, open source software is often more powerful than many commercially available products. In library industry too now many initiatives like ABCD, Koha, Newgenlib, Evergreen, Greenstone, Dspace, Drupal, Plone, Eprint, Jhoomla etc. have been taken to provide tools for integrated library management, digital library and content management.

Library professionals all over the world are using various web communication tools like Web OPAC, Website, Web blogs, Library portal, Digital libraries, Social networking sites to cater the evolving needs of library patrons. They have to depend on different tools like automation software, digital library software, content management software, and library websites e.g. Koha, Newgenlib, Greenstone, Dspace, Jhoomla are more popular in India.
Unlike other library softwares, ABCD software has some special features like it gives radical openness for database structures. One can create any type of database they want and gives full text capability for digital resources. Libraries can create their interactive website in very easy way using template and linked with social networking sites, RSS feed, etc. So the libraries will find it all in one solution for their library free of cost. Western country libraries started using this software.

2. Background Information

ABCD is the acronym for a software suite for the automation of libraries and documentation centers. In French this is “Automation des Bibliotheques et Centres de Documentation” in English we can say this as Automation of libraries and Centers of Documentation. This software package provides automation functions not only for the classic libraries but also other information providers such as documentation centers, archives, museums, NGOs, etc. It is developed by BIREME (WHO, Brazil) in collaboration with the Flemish Interuniversity Council, Belgium, and uses ISIS database technology of UNESCO. This software provides flexibility and versatility on top of storage efficiency and retrieval power. Any bibliographic structure can be created and managed by this software along with non-bibliographic structures, e.g. for factual information (events, organizations) with mainly textual content.

3. Aims and Objectives

The main aim of ABCD is providing an integrated library management tool covering all main functions in a library starting from acquisition, bibliographic database management, user’s management, loans transactions, and serial control up to online end-user searching on local and external bibliographic databases and library portal/website. It allows bibliographic records copying from the main leading libraries like Library of Congress, Oxford University (UK), Yale University (USA), Boston University (USA), University of Toronto (CAN), University of Chile (CHI), Australian National University (AUS) etc. through the Z39.50 protocol, which helps libraries to maintain international standards in bibliographical information such as MARC, CEPAL and AGRIS. It also gives facility to configure any additional Z39.50 servers. Librarians normally don’t want to go in detailed programming. Therefore it is designed as a tool for librarians rather than ICT technicians. Only they have to use the ‘Formatting Language’ of ISIS which allows them to manipulate all data in their databases in a high-granularity way in order to keep full control of it without programming. It is menu driven software and gives help instruction at every step.

3.1. Technical Features

- The software is fully web-based, so can be used and managed from any current web-browser. All main functions of the library management are integrated using the same interface and databases.
- Z39.50 facility: records can be downloaded from external library catalogs / server.
• Full MARC 21 compatibility as defined by Library of Congress OPAC with simple Google-like and advanced search like Boolean combinations, truncation, field-limitation.
• Gives access to both physical and electronic documents (local as well as on the internet) and managed with same interface.
• Librarians can define, edit any new database structure and also copy database structure from existing ISIS-applications e.g. MARC, CEPAL, UNIMARC, Dublin Core, etc.
• It is available in many languages like English, French, Spanish, and Portuguese while more translations are on their way (as the software offers an easy translation function in its own interface).
• Import and export data in ISO-2709 format and text-formats
• Library portal facility: contents and bibliographic resources (both local and external) can be added easily without HTML-programming.
• Excellent serials management fully implementing the ISSN standard and union catalog function.
• Statistical reports generation with graphical presentation of any defined set of variables in any of the databases.
• Freedom of database structure: ISIS accepts any structure and includes structure-definition tools, and so does ABCD. One can create any type of database and define records and run that database through this software. One can create dead stock database, CDs database, employee database, pictures database, etc. [3].

4. Installation and Download Tips

The software is very easy to install. It contains one package of all the necessary software parts (including Apache web-server and PHP, each with their own configuration). One can download the software for Windows and Linux platforms. One has to only unzip this folder in writeable space (C drive) of the computer. Demo and manuals for new version is available on ABCD software site [2].

4.1. ABCD Modules

ABCD software has following modules. Each module has separate login and password. All modules are menu-driven and easy to operate.
1. **ABCD Central Module**: It is a cataloguing module and data entry module.

2. **ABCD Advanced Loan Module (EmpWeb)** caters the advanced loan management system.

3. **ABCD OPAC** allows us to retrieve all defined catalogs, databases, and websites with simple, intermediate or advanced interfaces. Results can be displayed in any ISIS-format, including hyperlinks to the original documents for digital library functions, select, print, sent-to-email, bookmark in the major social networks and export to XML.

4. **ABCD Site**: A CMS (content management system) is provided for the creation of the ABCD Site using existing template. It enables to create, manage and publish the portal. Link with social networking site, RSS feeds. The site organizes information in a structure that integrates and interconnects databases created or from outside, specialist directories, events, sister institutions, libraries, full texts collection, catalogues of bibliographical resources available on the LAN, WAN or internet. It provides Online Public Access Catalogue or search tool for all types of resources.

![Virtual Health Library](image.png)

*Figure 2: Example of Library Website- A Virtual Health Library*

5. **ABCD Serials Control System (SeCS)** is an advanced management tool for serials or journals printed and online of all kinds in a single- or multiple libraries setup. There are following main options given:

- **Management of Titles**: In this we can add a new serial title and its details (in the full ISSN standard). We can export titles and can send the catalogue of records in union catalogue of serials. One can search the title through a serials index.

- **Management of Titles Plus**: allows creating and editing records with local administrative information on the serials. One can maintain issue records in this. Cardex functions are also fully present.

- **Management of Templates**: Allows creating and editing masks or templates of serial publication schemes (e.g. 4 issues per volume etc.)

- **Utilities**: allows user management, library management, statistical reports of serials by library and database. One can get statistical reports of serials holdings (print, online, free, paid, complimentary subscriptions, exchanged journal issues) [3].
4.2. ABCD as Digital Library Tool

According to the Digital Library Foundation (DLF) “Digital libraries are organizations that provide the resources, including the specialized staff to relate the structure, including the specialized staff to relate the structure, offer intellectual access to the interpret, distribute and preserve the integrity of and ensure the persistence over time of collection by a defined community or set of communities”.

Because of open access moment, web technology, online publishing and electronic publishing trend, extensive use of ICT in e-learning education system traditional libraries should convert to digital as the patrons demands are also changing with time. Digital libraries provide efficiency in information delivery, more economic and time saving.

According to Clifford Lynch (1995) “digital library is a system of providing a community to user with coherent access to a large organized repository of information and knowledge.”

ABCD software can operate as a digital library tool also with few limitations. Following are technical characteristics which are relevant for digital library applications:

Full-text indexing: the 'classic' ISIS since its early days - one could say from the 'prehistory' of computer science' as we are talking the 1970's! - had capabilities to extract individual words, except for non-meaningful words defined in a 'stop words'-list, from the fields of records in the database and to index these for searching into an 'Inverted File', which contains not only their record and repeated-field identifiers but also their position in the fields (allowing proximity searching e.g.) [1].

Since ABCD is based on ISIS for the data-storage and -retrieval on the one hand, and PHP for the web-interface creation on the other hand, some nice PHP-tools are embedded. E.g. FckEditor is a PHP-library offering a full HTML-editor shown below which can be embedded into a cataloging form to create full documents. Text from Word-documents also can be 'copy- pasted' into a field of an ISIS-record in ABCD by using this tool. The field will be presented as a web (HTML-) document but keeping the word-indexing technique for retrieval [4].

![Figure 3: Example of How to Add Text Using FCK Editor in the Field](image)

The use of non-Latin scripts: ABCD can use non-Latin scripts (e.g. Amharic, Chinese, and Greek) as part of the web-browser capabilities. ABCD team did successful tests on ABCD dealing with Amharic records (in the Ethiopian implementations) and records in Indian languages (in the workshops) too, whereas also an Amharic and Hindi version of ABCD is being prepared [4].
Further on in this article we will suggest some ideas for more fundamental extensions of the software for digital library purposes, including PDF word indexing, UNICODE and dropping storage limits for larger documents.

In view of the above described possibilities of ABCD, the software can deal with – but within the 1Mb per record limit – a multitude of non-classical library or collection types: collections of theses in higher education institutes, collections of URL’s and web-sites and scientific abstracts databases.

4.3. Theses

In the example of theses given below, most of these will be stored as PDF’s. ABCD now able to implement PFD-text extractor unit – index PDF’s, the meta-data of these documents (in fact much better than in Greenstone where the pure XML-storage without a real database-structure of the metadata means that larger collections of metadata will slow down the system considerably).

In the illustration here, taken from the sample CEPAL-database which was already part of the original first-release of ABCD, we show part of a record (right-truncated as image) not only showing images related to the record (covers of the books) but some hyperlinks, of which the first link is a PDF-link:
4.3.1. A Collection of Web-sites

The example of a web-page collection poses less problems, because web-pages mostly are not larger than the current ISIS-limits and HTML (or XHTML) is pure-text information, so can be fully contained inside an ABCD-record. ABCD (or in fact ISIS) can then fully index all words in the records as its normal operational work for retrieval.

4.3.2. Institutional Repositories

We can create separated database for different electronic materials like photos, sound, videos, patents, pdf, word files and other scanned material of the institute and make them available on the OPAC or site. The facility to protect document is one better feature in ABCD. One can protect the document with password.

4.3.3. Scientific Abstracts

Science community is more focused on abstracts of articles while searching the information as well written abstracts represents the core value and results of scientific research in themselves. Therefore now a day’s many KPOs are busy in preparing abstract databases for publishers. If full text indexed, they provide very handy access to immense scientific knowledge stored in databases. These abstracts can always be stored and indexed by ISIS. ISIS has no difficulties in dealing with the typical larger numbers of such abstracts; we e.g. ran tests on the full TEEAL database converted to ABCD (see the illustration below) and BIREME has a large experience in doing millions of such records with ISIS-technology [4].

A sample record of TEEAL in ABCD:

\[\text{Figure 6: Example Shows How to Create Abstract Database}\]

4.4. A PHP-Extension for PDF-Indexing in ABCD

ABCD team has successfully worked on a PHP-based extractor of words from PDF-documents. The idea is to store the extracted words of the PDF into the record and apply the traditional ISIS-indexing on it for retrieval and presentation of the document, with a (hyper-) link to the original PDF-formatted document.

This extension of files based on an approach similar to the one already used in ABCD for loading images into records: in fact not the images themselves but the URL’s to the image files, kept within the realm of the web-server (ABCD mostly uses Apache) for access rights reasons, are stored:

First an icon is presented next to a field for uploading a file.

When clicking on this upload-icon, a dialog is presented to identify the file to be uploaded (in our case it would be a PDF):
Now, unlike with a picture (of loan system users in the example above), where simply the file has to be stored into the given location (and ABCD creates the hyperlink automatically based on this location and the file-name), in the case of the PDF-extractor, the PDF-to-text converted will be launched and actually extract all words into a text file. This text file will act as the input file for a command by which ISIS (the mx-tool of the CISIS-utilities package) can create a field of a record into the database, taking and indexing all words of that input-file [4].

The mx-utility is a full ISIS-engine running from the command-line which can take a multitude of parameters, some of them identifying the input to act on (e.g. a database, an ISO-file, the console, or in our case: a sequential file): mx seq/1m=filename where the /1m optional parameter indicates the use of the much-higher-than-standard 1 megabyte option and the filename would be the output of the PDF extractor control. ABCD avails a whole set of such command-line ‘CISIS’ tools for advanced database management. Everything not available from the ABCD-interface can be found here. Mx, like ISIS, can also run a process (with the ‘proc=-’ parameter) to add the output of the processing of the input (in our case: the words from the PDF) into a field of the actual record [2, 4].

Standard ISIS-capabilities to full-text index this field then will be applied and all words of the PDF are then available for searching. The record will be displayed as text format (like Greenstone does) and the PDF itself will be available by a hyper-link (again as does Greenstone) [4].

4.5. Integration into the ABCD Meta-Search Portal

ABCD has, as one of the suite-parts, a full-fledged CMS to create a portal, integrating not only links to external information resources and messages for the users in the same web-page with the links to the OPAC’s of the databases available, but also providing a simple search box in which a Google-like search statement can be sent out to interrogate all defined databases. These can be both local ISIS- databases (e.g. the catalog to the books on the shelves, but also the digital library collections!) and remote databases in ISIS-format.

ABCD search performs one by one databases and results of these searches presented separately (so not a real ‘federated search’). So users could search in one effort both the physical library collections and the digital libraries, since all these are just ABCD-databases combined in the ABCD-Site. This might be a real added-value solution as compared to using e.g. Greenstone next to the classic library catalog software [4].
4.6. The ABCD Site CMS as a Solution for Libraries in the South

In the South - on the WWW, the ABCD site Content Management System (CMS) is a powerful tool for libraries to create library websites and is basically meant for libraries only, as part of the ABCD library automation software suite. The ABCD site CMS, as part of the free ABCD software for library automation and digital libraries [2], supports the design of a library website for any library, irrespective of whatever software used. Many libraries in the South, esp. Latin America are using it very effectively. It does not require programming skills, only adding information in the CMS worksheets provided. Application of Web 2.0 tools and access to dynamic electronic resources is rather easy to incorporate into the ABCD Site.

The ABCD-Site (Content Management System) is an integral tool which enables libraries to create, manage and publish their site. It is a customized partial version of the Virtual Health Library (VHL)-Site, produced and maintained by BIREME (a WHO-organization in Brazil).

A Virtual Library is a model for the management of information and knowledge. It is ideal example of cooperation and convergence between institutions, systems, networks, users and resources in the local, national, regional and international information network favoring open and universal access.

As discussed before libraries can link the various databases, directories, events of institutions, sister institutes, libraries, catalogues of various resources available online paid and free and full text databases and give search to all above also links library with social networking sites and make it interactive.

The possibilities of the ABCD Site CMS are discussed below:

The ABCD-Site currently has interfaces in 4 languages and it is possible to work in any number or combinations of these languages. The websites themselves can be in any language. The content of each interface is managed separately with the use of a code and a password.

In ABCD-Site administration system there are two types of users one Administrator who has full access to all areas of ABCD whole site and another is Content manager who is restricted to only the component areas of site i.e. contents.
4.7. Possibilities of the ABCD Site CMS

The ABCD Site CMS allows librarians to build their own website themselves without any knowledge of web-technology or HTML-coding. A whole range of elements can be presented in a rather well-structured way. The own resources (catalogs, databases) will be accessible directly from the website with an integrated search facility (‘metasearch’) but surrounded by other resources and information from both the Web 1.0 era (links to external information) and the Web 2.0 (interaction with users). The principal elements of such a web-site are discussed here [4].

4.7.1. General Tools These are related to features available in any ABCD content management window, whether in STRUCTURE, COMPONENTS or ACCESS. Each service box helps us to design the website as per our need, e.g. to add items according to the type of contents existing that section, modify content, delete content and save the changes made in that section and records them into the interface files.

4.7.2. Item-Lists List the items according to the type of content existing in the section. The information is shown in a hierarchical way, determining how it is displayed in the ABCD site. Up and down arrows enable us to change the sequence, while left-right arrows allow positioning the item correctly in the hierarchy of the page.

4.7.3. Specific Tools These vary according to the type available for each component of the system. To illustrate the multitude of possible elements we list them here with a very brief description [5].

The Structure Area Management:

- ABCD Logo: adds URL of the logo and URL for the responsible organization providing ABCD. Local images and logos will add a lot to a ‘localized’ look-and-feeling of the site.
- Identification: shows the software name-ABCD that will be displayed in the banner and at the top of the administration system page. One can edit it.
- Contact: enables to add E-mail to contact the person(s) responsible for managing and administrating this ABCD software. More than one address can be registered, which in the public interface will look in the form of combo box.
- Version: makes available public interfaces in other languages. English, French, Spanish, Portuguese language interfaces are available currently in the administration module. One can add other languages too like Hindi, Marathi, Philippines, up to 9 in total.
- Institution: enables to add institutions or organization or university name, image/logo and link for reference.
- Meta-information: It enables to verify and edit labels of research services available on the public interface.
- Responsible institution: Here host institutes can enter data for contacting the institution in charge of the Site, address and phone no, etc on the site footnote.
- Meta information of the site: is the space for description of ABCD, which allows ABCD to be indexed and retrieved by web browsers. It also allows to enter description of author, keyword and description, etc. figure shows how easy to add information in the site.
Texts: manages the list comprising all labels and image URLs that identify common actions and areas of the ABCD site, other than the search services. A lot of localization re language and terminology is possible here.

File manager: allows the uploading of local files to a dedicated area of the server into a directory which can be defined by the manager.

The administrator can filter the type of files admitted to upload, modifying the variable in the configuration file. It is recommended to design a clear strategy of filing archives in this area classifying them by type. It is recommended to prepare a balanced hierarchical structure if quantity of files is more of one class, so that there are not too many files in a folder at the same level.

Components: This part of the Site enables administration of the system via “columns” subdivided into components (or content types), with no limits as to the number of component inside each one of them in that:

- **Column 1** – links to other Libraries or related portals.
- **Column 2** – collection of ABCD information sources (local as well as non-local).
- **Column 3** – this column provides space for displaying highlights, news, etc.
A basic list of the 3 ABCD site columns is illustrated below. By clicking on any of these, more detailed settings can be edited while the CMS itself will translate all these settings into the proper HTML-coding [5].

![Columns Administration Window](image1.png)

**Figure 11: Columns Administration Window**

### 4.8. Sources Collection for the Meta-Search

Information Sources is probably the most important – therefore central – part of the Site: ABCD-based resources/databases can be included into the list, making them part of the ‘meta-search’ integrated approach (the search key will be searched in all listed and active resources with results displayed separately) [5].

![Information Sources](image2.png)

**Figure 12: Shows How to Add Information Sources in the List**

Its output on the site will be something like e.g.: (example taken from the Virtual Health Library)
4.8.1. **Portal Web Page** The ABCD Site can be used to contain links to second level pages (therefore acting as a portal), edited in XHTML (with a built-in easy HTML-editor). An example from the VHL is shown here:

![VHL Portal Showing External Link](image)

**Figure 14: VHL Portal Showing External Link**

4.8.2. **Communities** It enables to do registration to online communities like virtual communities, forums, etc.

4.8.3. **Portals** It enables to connect different links to other web portals related to ABCD.

4.8.4. **RSS** It enables to create links with news agencies that make the RSS feature available to users.

An example for a typical element in this communication area of ABCD site is shown here, as a result of copy/pasting some commonly available HTML-code for Twitter and Face book into the dedicated XHTML-editor field of the CMS (shown in right part of the illustration):
4.9. ACCESS Management

This last part refers to the management of users, rights, passwords and those responsible for the administration of ABCD. Here administrator can add, edit or delete the registration of users and passwords to limit the management of the Interface.

4.10. Dealing with the Graphical Design

Since ABCD uses CSS-techniques all the way, it is relatively easy to change the overall ‘style’ or ‘theme’ of the website, e.g. by adding background pictures, colors and other graphical elements.

4.11. Some Examples

Esp. in the context of the ‘Virtual Health Library’ in many Latin American countries one can find many websites created with this CMS, which became integrated into ABCD as the ABCD Site CMS. Although, as is intended, with many different layouts and contents, one can mostly still easily recognize the main basic 3-parts structure of such a website with the central ‘meta-search’ part somehow built in.

The first example is a full-blown example of BIREME’s VHL site itself.
The second example is a more typical one, created in a 2-hour session on the ABCD-Site in an ABCD-workshop (in this case a workshop in Maputo, Mozambique, December 2011 conducted by Egbert De Smet), where all the elements were created by the participants as a hands-on training on ABCD.

The ABCD Site Content Management System is suitable for all types of libraries in developed and developing countries. The example of Virtual Health Library and their allied libraries mentioned above illustrates its excellent functionality, presenting a wealth of information sources around its own ‘metasearch’ core for availing local information. As this tool is for libraries in any type of institutes and quite easy to manage and create, it can be used to publish and share all types of information and library services.

UNESCO is supported efforts to re-design, quite drastically, the storage and indexing engine of ISIS by replacing the ‘classic’ technology of ISIS by a new database (Berkeley DB, see http://www.oracle.com/technetwork/database/berkeleydb/overview/index.html) based on the same no-
SQL principles of ISIS where records are just numerical identifiers followed by an 'object' which could be either an ISO-2709 record, a BLOB, an XML-record etc.) and a new indexer: Lucene is widely used nowadays for full-text indexing and adds to the existing features of the ISIS-query language things like relevance ranking and search- keys without limits.

In new beta version of ABCD one can manage videos, pictures, music files, text files, pdf files and other records very easily and size of file is not a problem. One can adjust the record size as per their wish and as per the capacity of their server. So libraries can use it effectively and easily.

The team is now at final stage of using the rich Java- based libraries for extracting words from PDF- and other document-formats (see e.g. PDF Box at http://java-source.net/open-source/pdf-libraries/pdf-box) While Lucene has no problem indexing huge amounts of such text-files.

5. Conclusion

To cope with the all modern technologies coming up for the benefit of users and librarians, to cope up with the shrinking library budget, to cope up with online and offline resources, and to fulfill users demands, some open source softwares like ABCD are offering an excellent solution for all types of libraries along with other free open source softwares. This article covers basic information about ABCD software. Libraries can make use of this open source tool for library automation easily without any technological background, staying close to the librarian's views on library automation.

References


