

Designing and Migration of Library Database

Dr. Sathiyapriya K¹, Bhavadharinie S², Atchaya R³, Kripaa H⁴, Sowmya V⁵

Dept. of Computer Science and Engineering, PSG College of Technology, Coimbatore, India

E-mail: spk.cse@psgtech.ac.in, bhavadharinie@gmail.com, atchaya920@gmail.com, kripaah26@gmail.com, sowmyaviswanathan008@gmail.com

Abstract

Library Management System is the finest tool for automating libraries in India. Data migration from such proprietary applications is difficult. In the majority of the current libraries, the tabular bibliographic data is exported to spreadsheet forms like CSV and XLSX. It is necessary to transform bibliographic information from spreadsheet format to the industry-standard MARC 21 format. MARC tags are applied to bibliographic information in spreadsheets using MARC Edit software. The .xlsx file is converted to .txt (tab delimited) which is then converted to .mrk format. Next, the .mrk file is changed to .mrc file. Processes for data mapping are necessary prior to this move. Later, the .mrc file must be loaded into a MySQL-compatible open source database. To achieve that, the migrated material needs to be properly validated. Therefore, it is crucial to use accurate data mapping when converting data from XLSX format to MARC 21 format. Following the importation into a database, users can obtain data for cataloguing and visualisation for easy understanding. Projects using open-source software are fundamentally about cooperation and transparency, and many eminent Indian institutions have switched to these open source platforms.

Keywords: MARC, Open Source Systems

1. Introduction

For faculty, students, and researchers, the GRD Memorial Library has a sizable collection of books, periodicals, and online tools. There are currently 240 printed journals, over 212 national journals, 9 international journals, 1075 e-journals, 165769 lending books, 23254 reference books, 18101 book bank, 13053 non-book materials, 33350 back volumes, 351

standards, 5117 department library collection, 12675 CDs/DVDs, and access to over 9000 web-based journals. Users can reach a user-friendly internet portal at the computerised library. A specialised digital library has 60 computers with high-speed internet access. The staff and library system are intensely focused on providing a relaxing, stimulating atmosphere to support learning. With the aid of Open-Source Software (OSS), libraries with necessary IT infrastructure can systematise the library services. Libraries can immediately adopt OSS and begin the automation processes from scratch even without automation software. Data export from vendor-locked older systems is not feasible, which presents a challenge for libraries that use proprietary software and wish to switch to OSS. Additional problems with data export are brought on by unstructured data from outdated robotic systems. Therefore, exporting the data into various library systems is made simple by converting the data into a common format like MARC 21.

The MARC 21 formats serve as standards for the encoding and transmission of bibliographic and related data in a machine-readable manner. The record structure, content designation, and data substance make up a MARC record. A MARC 21 format is a collection of codes and content designators used to encode machine-readable data. Bibliographic, holdings, authority, classification, and community data are five distinct types of data that have standard formats. After consulting with a variety of user groups, the Library of Congress continues to use the MARC 21 formats. The MARC 21 versions use the Information Interchange format. The forms also include other relevant ANSI standards. The needs of the printed collection were taken into account when developing earlier generations of library administration software. Both print and digital resources must be represented in today's integrated library administration software. To effectively implement the migration process to OSS without any data loss, library professionals must be familiar with the data structure, mapping, and different migration phases. The finest method for library automation in India is provided by open source library management systems. It is possible for libraries to exchange bibliographic resources more effectively by converting their datasets to a standard format like MARC 21. It is a protocol for sharing bibliographic data between digital library systems.

The scope of this research is to overcome the drawbacks of proprietary software by using Open Source software in Library Management Systems. Some benefits of the system are:

- i. Easy access to information for library staff due to effective search and update modules.
- ii. Interpretation of information becomes easier due to visualization.
- iii. Reduction in cost due to the use of open source system.
- iv. Further developments can be made by adding new features like insert.

2. Related Works

Various articles and journals have been reviewed for better understanding of data migration, why Open Source Software are preferred over proprietary software, different OSS available for library management, and MARC21 conversion and case studies about the database migration.

Vimal Kumar and Majeed Abdul [1] described a practical approach for data migration from Legacy systems To Koha. Detailed information about various stages involved in the migration which includes planning, data export, data mapping and importing were discussed.

Josiline Chigwada and Blessing Chiparausha [2] explored different experiences with migration of library management systems. Solutions to problems encountered and proper planning practices for feasibility before migration were focussed. Challenges like limited time, lack of expertise, data corruption and loss were also discussed.

Vandana Singh [3] proposed the experiences of migrating to an open-source integrated library system. The paper provided various information from different librarians who worked in libraries that migrated to open-source integrated library systems or are in the process of migrating. It also provides an overall view of various steps involved in the migration process and the difficulties faced by them in each step.

T.Raghunadha Reddy and K. Kumar [4] explored open source software and their impacts on library and information centre. The benefits of choosing Open-Source Software over Commercial software such as flexibility, speed, and openness, along with the difficulties involved in selection of appropriate library management software, were listed. The various OSS available for Library Management Software (like KOHA, ABCD, Evergreen) and their features were discussed.

Ram Kumar Matoria and P.K. Upadhyay [5] studied migration of data from one library management system to another. The reasons to switch from one library management system to another are lack of required features in existing version/software or the cost of maintenance of the existing software etc. The migration of data was done by a step by step approach. The issues considered while migrating data were nature of the library management software, availability of the input/export formats, data elements available, field/record separator, relationship among the data elements and date, currency etc.

3. Problem Definition and Preliminaries

3.1 Problem Definition

Today's majority of library services, which are based on information technology and materials available in electronic formats, generate digitalized databases. Libraries need cutting-edge retrieval tools and high-end integrated software to manage all types of resources and information. However, the majority of libraries are unable to use them due to the expensive cost of such software. Additionally, compared to proprietary software, open-source software is more adaptable and transparent. This kind of automation is required because of the huge volume of data in today's libraries. As a result, switching from proprietary to open-source software is inevitable.

There are numerous issues with the current library systems that use proprietary software. One such negative is the maintenance fee, and users occasionally have to pay for services they don't need, which raises the price. Additionally, libraries ought to rely on developers for maintenance, updates, and upgrades. Moreover, there is no common format for storing bibliographic information in proprietary software. Therefore, these library systems ought to be converted to open source ones.

3.2 Preliminaries

3.2.1 Functional Requirements

Data Pre-processing and Integrating - Processing of raw data by cleaning, normalizing, and integrating the dataset.

Data Mapping - Matching tags and indicators for each bibliographic elements (e.g., author, title, etc.) with MARC standard.

Conversion to MARC 21 - To convert .xlsx files to .mrk and then to .mrc format, using MARCEdit software.

Importing to Open Source - Importing MARC records into open source system using graphical interface and command line.

Test Run - Testing and validation to check whether the records are converted and imported successfully without bugs.

Data Retrieval - Retrieving data from the database, creation of search, and update field modules.

Data Visualization - Curating data into a form easier to understand, and highlighting the trends and outliers.

3.2.2 Non-Functional Requirements

Usability Requirements - The user (librarian) can use this system in a user-friendly manner with some prior knowledge.

Reliability Requirements - The system should perform consistently without any unexpected failure. These measures should be properly handled in the test case scenario.

Efficiency Requirements - The system should take a minimum time for loading the data or inserting data into database. It should be efficient with different environments irrespective of the number of records in the database.

4. Proposed Work

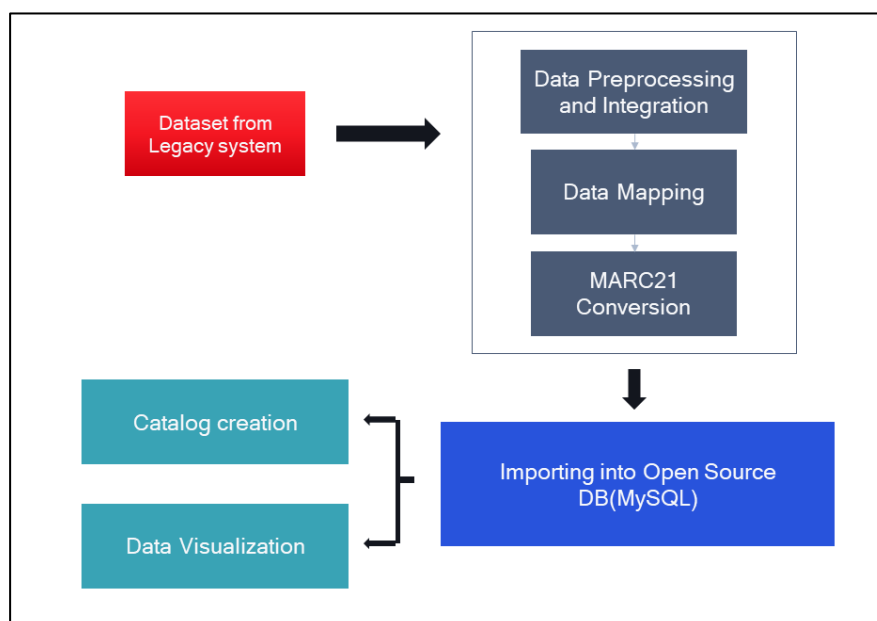


Figure 1. Overall Architecture Diagram

Metadata in xlsx is a collection of library dataset and analysis of the form of data, the attributes and the records. During data cleaning, processing of raw data by integrating, cleaning, and normalizing the dataset are performed. Matching tags for each bibliographic elements (e.g., author, title, etc.) with MARC standard i.e., Data Mapping, is executed. By using MARCEdit software, .xlsx files are converted to .mrk and then to .mrc format. Then importing MARC records into open source system (MySQL) is carried out using graphical interface and command line. Finally, testing is done to check whether the records are imported successfully without bugs. After successful import, the data can be used for the creation of digital catalogue to ease user access. Later, data visualization is done to obtain a clear idea of what the information means, by giving it a visual context through graphs.

5. Results and Discussion

5.1 Experimental Settings

There are totally 8 lending book list. The attributes of the datasets are Accession Number, Book Title, Author, Published Year, Publisher, Call Number, Material Status, Material Type, ISBN, Floor Number, Price, Rack Name, Language, Library, AcHead Name, Volume and Material Group. The raw data contains many empty values, special characters at

end or beginning and unnecessary white spaces. After pre-processing, the missing values are replaced with 'UNKNOWN' or '0' depending on their datatype. White spaces are removed using strip function. The data is then integrated to a single .xlsx file with 177854 records.

The following are the steps in data pre-processing and integration: First, the files are read in .xlsx format. Then, a data frame is created. After that, special characters are removed and missing values are filled. Next, it is converted to the .xlsx file. Finally, all the files are integrated. During data mapping, the respective row of the .xlsx file is filled with appropriate marc tags for fields and sub-fields as shown in Table 1.

Table 1. MARC tags and indicators

FIELD NAMES	FIELD TAGS	SUB FIELDS	INDICATOR
SNO	082	a	0\
ACCNNO	082	b	0\
TITLE	245	a	10
AUTHOR_NAME	100	a	1\
EDITION	250	a	\
PUB_YEAR	260	a	\
PUBLISHER	260	b	\
CALL_NO	952	o	\
MATERIAL_STATUS	300	a	\
MATERIAL_TYPE	300	b	\
ISBN	020	a	\
FLOOR NO	952	c	\
PRICE	952	g	\
RACK_NAME	952	y	\
LANGUAGE	041	a	\
LIBRARY	952	a	\

AC_HEAD_NAME	952	b	\\
VOLUME	773	g	0\
MATERIAL_GROUP	300	c	\\

After data mapping, MARC21 conversion takes place. In MarcEdit, Delimited Text Translator is selected. The .xlsx file is chosen to convert, and the required .mrk file is named. In the translator, each field is mapped to its corresponding tags and indicators as in table 1. After adding all arguments, the fields with common field tag is selected and joined. Upon completion, the data is converted to .mrk format. The steps in the conversion of .mrc format are as follows: In MarcEdit, MarcEditor is opened, and then the .mrk file is opened. File -> Compile File is given to MARC. The resultant .mrc file is produced. Next, the .mrc file is validated with MARC rules to check whether the data has been converted properly. MARC records validated in MarcEdit can be used for this.

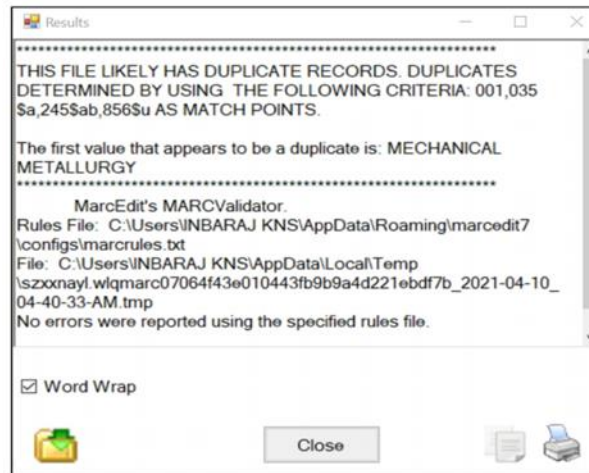


Figure 2. Validation result

After successful conversion and validation, importing into open source DB (Mysql) process is carried out. This can be done with the help of MarcEdit 'MARC SQL EXPLORER'.

The .mrc file is chosen and the connection to the MySQL database is done. (Database Server : localhost Username : root Password). This connection changes according to the server. Now the .mrc file is imported to the database as given in the database name.

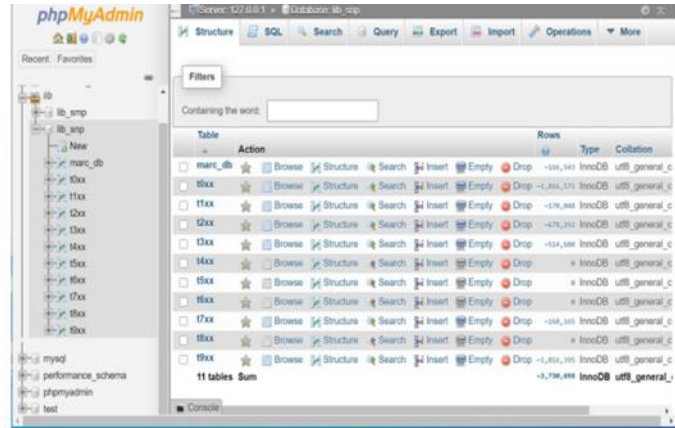


Figure 3. MySQL database

Next module is the Catalog creation, where search and update options are executed. In the search option, the input of either the title of the book or the author name is used to filter the records and display the details having the input text alone; whereas in the update option, the data to be updated can be found in two ways:

1. The data record to be updated can be searched with its title or author name.

2. The list of fields having missing records will be shown. The field to be updated can be selected. It shows all the records having the missing fields. Then the required record can be updated using the update module.

Final module is data visualization, where Python programming with matplotlib is used to visualize data as bar charts and pie charts. The data is retrieved from database and used to plot the charts. Visualization is done for 4 fields, namely:

1. Material_type
2. Material_status
3. Floor_No
4. Missing records

5.2 Performance under different test cases

Table 2. Summary of testing for data pre-processing

Test case no.	Inputs	Expected Output	Obtained Output	Test case Acceptance
1.	Dataset containing many files with missing column values and presence of special characters in it	A single file with all columns having values	A single file with all columns having values	Yes

Table 3. Summary of unit testing for data mapping

Test case no.	Inputs	Expected Output	Obtained Output	Test case Acceptance
1.	Preprocessed data file	File with tags for each bibliographic element matching with MARC standard	File with tags for each bibliographic element matching with MARC standard	Yes

Table 4. Summary of unit testing for conversion to MARC21

Test case no.	Inputs	Expected Output	Obtained Output	Test case Acceptance
1.	Data mapped file	.mrc format file	.mrc format file	Yes

Table 5. Summary of unit testing for importing to MySQL

Test case no.	Inputs	Expected Output	Obtained Output	Test case Acceptance

1.	.mrc file	Data records imported in MySQL	Data records imported in MySQL	Yes
----	-----------	--------------------------------	--------------------------------	-----

Table 6. Summary of unit testing for catalog creation

Test case no.	Inputs	Expected Output	Obtained Output	Test case Acceptance
1.	For search of book details, attributes regarding the book like title or author are given	All book details with the given attributes are displayed	All book details with the given attributes are displayed	Yes
2.	For update of book details, attributes regarding the book like title or author are given	All book details with the given attributes are displayed for update and are updated	All book details with the given attributes are displayed for update and are updated	Yes
3.	For update of book details, fields with missing values are given	Fields are updated	Fields are updated	Yes

Table 7. Summary of unit testing for visualization

Test case no.	Inputs	Expected Output	Obtained Output	Test case Acceptance
1.	Material type data from database	A pie chart representing material types	A pie chart representing material types	Yes

2.	Material status data from database	A pie chart representing material status	A pie chart representing material status	Yes
3.	Floor details from database	A bar chart representing number of books in each floor	A bar chart representing number of books in each floor	Yes
4.	Missing value details from database	A bar chart representing missing values in each field	A bar chart representing missing values in each field	Yes

All the modules have passed the unit testing. But when they were all integrated together, certain issues had to be resolved. In executing the above modules together, the model had to be hosted on the local server and the following errors/issues were identified during this process.

Table 8. Summary of integration testing

Issue No.	Description	Solution to resolve issue	Completion of resolving issue
1.	Couldn't integrate files first	Preprocess the files and integrate	Yes
2.	Access error while trying to convert to MySQL	Granting permission in SQL configuration file	Yes
3.	Slower speed while importing into database	Update the software	Yes

TITLE	AUTHOR	ISBN	EDITION	PUBLISHER	PUB YEAR	STATUS	GROUP
ANALYSIS OF STRUCTURES: ANALYSIS, DESIGN AND DETAILS OF STRUCTURES - VOL. 1	VAZIRANI V N, RATWANI M M	9788174091406	0	KHANNA PUBLISHERS	1976	AVAILABLE	LENDING
ANALYSIS OF STRUCTURES: ANALYSIS, DESIGN AND DETAILS OF STRUCTURES - VOL. 2	VAZIRANI V N, RATWANI M M	9788174091406	0	KHANNA PUBLISHERS	1976	AVAILABLE	LENDING
	DOUGLAS A SKODOG			CENGAGE			

Figure 4a. Results of Search and Update modules

PUBLISHER: UNKNOWN

EDITION: 0

PUB YEAR: 1994

ISBN: 0

UPDATE

Figure 4b. Results of Search and Update modules

Figure 4a and 4b show the user interface of the search and update pages, where an user can give the author's name or the book's title to display the available books. Also, there is an update mode for the librarian to update book details. Figure 5 to Figure 9 show the different types of visualization modules that have been implemented.

1. Figure 5 shows the visualization module of a bar chart for displaying the number of books with unknown fields that has to be filled by the librarian.
2. Figure 6 shows the bar chart to find the number of books in different floors.
3. Figure 7 shows how many values are missing in each category like author, publisher, etc.
4. Figure 8 and 9 show how many books in different types like ebooks, books, etc. are available and how many books in different status like available, issued, etc. are available respectively.

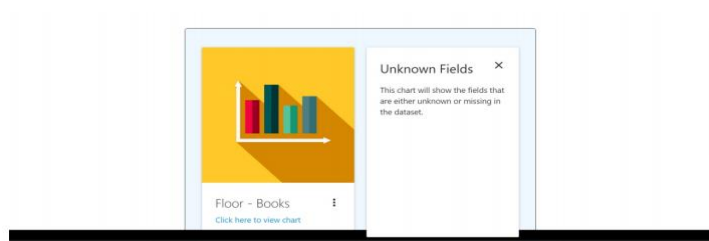


Figure 5. Visualization module

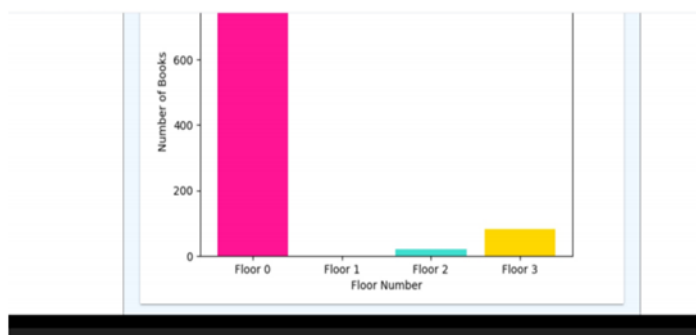


Figure 6. Floor_no. visualization

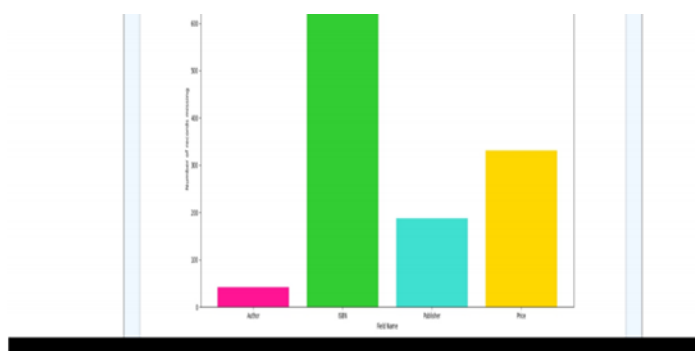


Figure 7. Missing values visualization

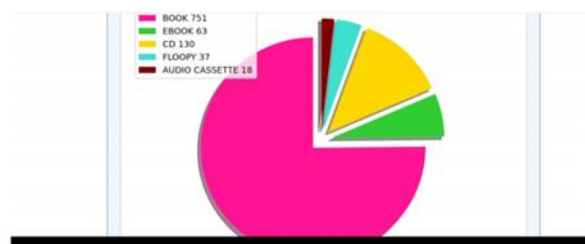


Figure 8. Material type visualization

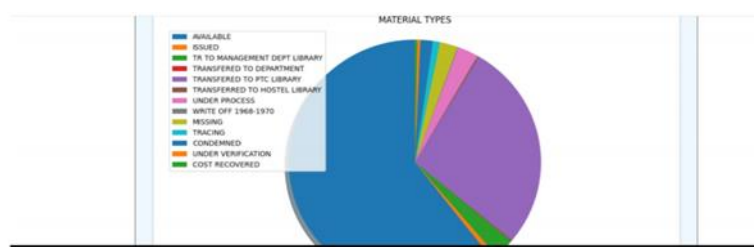


Figure 9. Material status visualization

6. Conclusion

Due to the numerous problems with the legacy system, many libraries would prefer to switch from proprietary software. Every Open Source Software initiative is built on the principles of transparency and cooperation. Increased benefits can be acquired from frequent updates and improved automations, thanks to open source library management systems. Additionally, open source software systems allow for internal system modifications as needed. Moreover, the material is diverse and dynamic to use, thanks to MARC and MARC tags. It is difficult to adopt Open Source Systems in place of legacy systems. An effective transfer requires proper dataset pre-processing and integration. The collection can be used with a variety of open source library systems that support the MARC21 standard. Furthermore, the data search, update, and visualisation offer the librarians a user-friendly setting for accessing and interpreting the data. Additional upgrades could be made, including the incorporation of

other library features like the insert module, a full OPAC, bespoke reporting, and documenting. Besides, visualisation modules can be created for other variables like publisher and pub year.

7. References

- [1] Vimal Kumar, and Majeed Abdul, “Data Migration From Legacy Systems To Koha: A Practical Approach”, *Library Philosophy and Practice*, 29 Apr., p. 19, 2019.
- [2] Josiline Chigwada, and Blessing Chiparausha, “Experiences with library management systems migration”, *SCECSAL XXII*, 2016, p. 8
- [3] Vandana Singh, “Experiences of Migrating to an OpenSource Integrated Library System”, *Information technology and libraries*, vol. 32, no. 1, pp. 36-53, 2013.
- [4] T. Raghunadha Reddy, and K. Kumar, “Open source software’s and their impact on library and information centre: An overview”, *International Journal of Library and Information Science*, vol. 5, pp. 90-96, 2013.
- [5] Ram Kumar Matoria, and P.K. Upadhyay, “Migration of data from one library management system to another: a case study in India”, *Electronic Library and Information System*, vol. 39, no.2, p. 8, 2005