Content available at: https://www.ipinnovative.com/open-access-journals

IP Indian Journal of Library Science and Information Technology

Journal homepage: https://www.ijlsit.org/

Review Article

Data migration and retrospective conversion using marcedit: A study of Koha

Gulafsha [1,*], Anhad Narad [1], Pooja Mishra [2], Atasi Chakravarty [3], Rupak Chakravarty [1]

[1]*Dept. of Library and Information Science, Panjab University, Chandigarh, India*
[2]*Bhubaneswar Engineering College, Khurda, Odisha, India*
[3]*Amity University Mohali, Punjab, India*

ARTICLE INFO

Article history:
Received 29-10-2022
Accepted 20-12-2022
Available online 23-06-2023

Keywords:
Koha Library Automation software
ReCON
Data migration
Communications Technologies (ICT)

ABSTRACT

The researcher introduces the work by explaining the background of the study. It goes ahead to state the statement of the problem (which in this case is the Data Migration and Retrospective Conversion using MarcEdit: a study of Koha objectives of the study, scope and limitation of the study and significance and scheme of the study.
The purpose of this study is the researcher prepare the data ready for Koha import. excel to koha software for enhancing library services and make housekeeping operations of library easy and quick. The researcher has an overview of the critical appreciation and tries to give the statement of knowledge in the field of study as well as try to identify the possible solutions of the study questions; investigator also gives review of previous relevant literature. Retrospective conversion helps catalogue in a library to convert traditional form into a machine-readable form. The researcher used tables when illustrating the data collected. Quantitative and qualitative data was then analysed and estimation by presented in tabular form using excel worksheets. The tools that are used to migrate data from excel to Koha that is MarcEdit 7.5.
Research output defines as the quantity and quality of finished research works. A conceptual model for data migration using Koha was also designed which could be tested for future scope of the study taking in more organization. In recent time ReCON considered as a powerful tool for any type of library to change existing library management system to another system for providing easy and relatively reliable to access the resources. The library management system of different kinds with powerful features are available in the market; which can also be used in academic library in order to promote library activities and services for the benefit of academic fraternity and research output. The researcher also states and gives a summarized discussion, a finding, suggestions for further research.

1. Introduction

Library Automation according to the adoption of emerging and new technologies, libraries and other information institutions are undergoing inevitable and significant shifts that extend beyond their traditional function as keepers of recorded knowledge. These technologies include, but are not limited to, incorporating fresh methods of information transmission, storage, and retrieval. Organizational structures, procedures, and tasks are all impacted simultaneously by these modifications, that are primarily made without ever using technology. To match the information-seeking behaviour and increasing user expectations, service delivery mechanisms are also shifting in line with these trends. One area in which these technology-driven changes and impacts are being seen

* *Corresponding author.*
E-mail address: chotu.dibiya@gmail.com (G.).

https://doi.org/10.18231/j.ijlsit.2023.002

in libraries is automation. It entails using technology to perform tasks more rapidly and efficiently while saving time and energy of personnel. [1–5]

The word automation has been derived from the Greek word “Automose” which means something which has the power of spontaneous motion or self-movement. The main objective of library automation is to liberate librarians and other library staff members so they can contribute more significantly to the dissemination of knowledge and information. (*Library Automation DLIS007*, n.d.)

As per the Encyclopaedia of Library and Information Science, "automation is the technology concerned with the design and development of process and system that minimize the necessity of human intervention in operation". (Kent, 1977) Automation, when used in a library context, refers to the computerization or mechanization of all library activities.

2. Koha

Koha is the first free and open-source integrated library system (ILS). Development is sponsored by libraries of varying types and sizes, volunteers, and support companies from around the world. (*Koha 21.11 Manual (En) — Koha Manual 21.11 Documentation,* 2021)

3. MarcEdit

In 1999, MarcEdit started out as a simple MARC=>plain text translation utility. Today, MarcEdit has become one of the more complete metadata edit suites available to librarians. While the name MarcEdit has become a bit of a misnomer over the years (as MarcEdit now edits so much more data than simply MARC records), the program has and will continue to evolve to meet the past, present, and future metadata needs of the library community.

MarcEdit presently includes a very rich feature set targeted at making metadata translation and editing easier for both the beginner and advanced users. These features include:

Marc Editing, Automatic Normalization support, Metadata Translation, MarcEditor, Link data Framework, Bibframework Workflows, Clustering, RDA Helper, Delimited Text, Translator, Export as Tab Delimited, Harvest OAI Data, XML Translations, XML Profiler

Characterset conversion, Z39.50/SRU Client, ILS and OCLC Integrations, MARC SQL Explorer, MARCValidator, XML Editor, Hex Editor, Native software for MacOS and Windows. [6–9]

3.1. Many more features

Like the ability to split, join, sort, etc. data in various formats. (https://marcedit.reeset.net/archives/author/reeset, 2013)

3.2. Data migration

Library automation is one of the primary activities of all kinds of libraries. For library automation, libraries are using some Integrated Library Management System (ILMS) software that may be proprietary or open-source. The question of data migration emerges when library administrators think of moving from an existing ILMS to a new ILMS (Saikia, 2019).[6-

Data Migration and Retrospective CONversion (ReCON) has always been a challenging yet essential task. Data migration is the process of moving data from one system to another involving a change in storage and database or application (Naeem, 2019). A strategic data migration plan should include consideration of these critical factors:

1. *Knowing the data* — before migration, source data needs to undergo a complete audit. Unexpected issues can surface if this step is ignored.
2. *Clean-up* — once you identify any issues with your source data, they must be resolved. This may require additional software tools and third-party resources because of the scale of the work.
3. *Maintenance and protection* — Data undergoes degradation after some time, making it unreliable. This means there must be controls in place to maintain data quality.
4. *Governance* — Tracking and reporting on data quality is important because it enables a better understanding of data integrity. The processes and tools used to produce this information should be highly usable and automate functions where possible. (*Data Migration: Strategy and Best Practices,* 2021)

Moving data from one place to another is as important as keeping useful previous data safe or transferred then it’s crucial to do the retrospective conversion. Retrospective Conversion is an important essential activity to be undertaken in libraries because already existing records are important for a library as well as its patrons too. [10–14]

4. Retrospective Conversion (ReCON)

The ability to migrate data is a key aspect of any change from one library management system to another the existing library data is an important asset and cannot be generated again and again as it is costly in terms of resources such as money, manpower, and time. (Data migration from one library management system to another: A case study in India, 2005).

In that case, retrospective conversion helps in data migration. Retrospective conversion or ReCON in library and information centre means "changing already existing catalogue from existing *traditional form to a machine-readable form.* (Dabas, 2004)

According to the ALA Glossary of Library & Information Science, ReCON has been defined as the process of converting the database of library holdings from non-machine-readable form to machine-readable form that is not converted during the day-to-day process. *(ALA Glossary of Library and Information Science, Fourth Edition | ALA Store,* 2013)

1. Retrospective Conversion (ReCON) is the conversion of a library's existing bibliographic records from manual to machine-readable format according to specified policies and standards" B. Jane, C. Joseph.
2. ReCON is the process of conversion of a printed catalogue of library holdings into machine-readable form, as quickly and as inexpensively as possible, in a pre-defined format using accepted standards for making it searchable through a machine. (Understanding MARC Bibliographic: Parts 1 to 6, 2022)
3. Harrod's Librarian's Glossary defines retrospective conversion (information retrieval) as a partial or complete conversion of an existing catalogue into machine-readable form as opposed to converting records created currently (Raymond John Prytherch, 2016).

Booth, R. (2018) has given the following Database Readiness Factors:

1. Catalogue records must be carefully converted from manual to machine-readable formats;
2. Collections must be prepared for conversion through effective and ongoing weeding and inventory programs;
3. Once converted, collections must be properly maintained as titles are added, withdrawn, transferred, and re-catalogued; and,
4. Standards – for bibliographic, item, and patron records as well – must be adhered to.

This process is a prerequisite for library computerization. Libraries cannot begin online transactions of other functions such as acquisition, cataloguing, circulation, OPAC and serial control, etc. unless and until the retrospective conversion is not fully done.

5. Need and Benefits of Retrospective Conversion

Exponential rise in published information and the limited library funds has increased the dependency on resource sharing for which MARC (Machine-Readable Cataloguing) provides a standard format. Machine-readable bibliographic data is exchanged across different library systems using MARC. The records produced during the retrospective conversion must adhere to the MARC standards in order for the records to comply with the international standards, simplifying the transfer of data between systems.

The foundation for all library automation software current and future automation efforts is the development of a top-notch MARC compliant library database. Vendors for the library may change over time, hardware may become outdated as new software replaces the old, but a well-built, well-maintained database will always be a valuable asset for the library. The quality of the corresponding databases and collection usage is improved via MARC-based ReCON.[15]

The conversion of the database of library holdings from non-machine-readable form to machine-readable form is a prerequisite to implementing an automated system. These records provide the means of generating statistics and other information that is needed to improve the existing services and introduction of new ones (Dabas, 2004). ReCON enables libraries to switch over to a new MARC based ILMS to create a local database, improve and expedite the user services, improve internal library procedures, create an integrated file, eliminate the cost of maintaining parallel systems, provide flexibility in system changes, maximize return on investment involved in library automation, data protection, create union catalogue, exchange and share bibliographic data & resources, facilitate collection rationalization and contribute in national databases and strengthen Universal Bibliographic Control (UBC). Adherence to well-established and accepted standards of description for bibliographic information in a machine-readable database is critical because:

1. Without standards, files cannot easily be transferred from one automated system to another, and,
2. It is essential for libraries wishing to participate in resource-sharing arrangements with other libraries, that will require such adherence as a condition of participation. (Cohn, et. Al.,1997)

5.1. Pre-conversion Issues

1. *Record Format:* To ensure that the first conversion is the last conversion of the library or information centre, the bibliographical information of the documents must be converted into the standard format.
2. *Fields in the Database:* It is time to decide what fields are to be included in the database and whatnot. Current practices for recording the bibliographical information may need modification when added to the machine-readable form. The decision should be taken after considering the users' present requirements and future needs.
3. *Source for Bibliographical Information:* Will the accession register, shelf list, or public catalogue be the source to record the bibliographical information for conversion? The choice should be based on that source that has accurate and complete information to meet the requirements and quality standards.

4. *Priorities of Conversion:* Determine priorities of the areas of conversion so that more important areas of the holdings may be converted first in a machine-readable format.
5. *Establish Standards for the Recon Project*

6. Approaches of Retrospective Conversion (ReCON) Process

There are several ways for recon to convert the catalogue into machine-readable form (MARC). The best method for a library or information centre depends upon the available resources at hand. The library may decide on - partial or full conversion, Capturing data by machine - optical character recognition (OCR), voice data input; By an outside recon vendor/commercial vendor or a combination of in-house and vendor recon. There would be a need to find the option, which offers the lowest cost, the highest quality, and the one with the least negative impact on the library.

6.1. Key to successful RECON project

1. Careful planning: Close examination of each method of the conversion, taking into account the library's budget, time, and manpower constraints;
2. Weeding of titles and copies that have marginal value. Since it is expensive, it makes no economic sense to convert materials that are of little or no worth.

7. Review of Related Literature

Purpose this case study aims to demonstrate that in-house integrated library systems migration can be accomplished by a dedicated team of librarians without advanced tools or prior experience with data migration or systems integration. Design/methodology/approach this migration was accomplished by academic librarians using freely available tools: Open Office Calc, MarcEdit, and the Koha Integrated Library System. Findings the data migration pathway presented here was developed and successfully used to transfer over 48,000 records in less than two months. Practical implications this case study presents an original process that is particularly effective for smaller libraries. Originality/value while similar case studies exist, most employ expensive third-party contractors for data migration or rely heavily on institutional IT departments (Todd, 2018).

In 2015, Rutgers became only the second accredited law school in the United States to select the open-source integrated library system, Koha. The merger of two unique catalogues at Rutgers Law School has presented unique challenges concerning migration mapping, data recall for large records, and relevancy ranking, all of which affect search results and usability of the Online Public Access Catalogue. System migrations always result in some data being lost or incorrectly transferred. The hope is to minimize just how much data is compromised while fixing errors that might not have come to light but for the migration. (Mitchell, 2017)

7.1. Research gap and research problem

There exists a lack of study demonstrating the practical use of MarcEdit for ReCON and data migration facilitating spreadsheet data import to Koha ILS. Many libraries are looking for adopting open source ILMS like Koha but they are unable to do so due to lack of procedural clarity regarding use of MarcEdit for ReCON to convert non-MARC bibliographic data into MARC and subsequently migrating the converted data into the ILMS.

7.2. Rationale and significance of the study

It's always a challenge to plan smooth migrations. Without basic software knowledge address common conceptual and technical issues encountered in migrations. An important planning and implementation tool that will help prevent headaches and frustration in this Using step-by-step instructions and checklists in the MarcEdit software saves time. The study demonstrates how process and prepare the data ready for Koha import. Using free software compared to in purchase software or IT staff will overhead expenses to the library. MarcEdit software is a free and useful tool for migrating the library's foremost data with saving expenses and manpower.

8. Objectives (Ob) of the Study

Every research has its aims and objectives to achieve. In every research, there must be some specific aims and objectives for which the study is to be conducted. Without definite aims and objectives, research has no value and may be considered meaningless. Hence, the study focused on the following basic objectives-

1. To examines the effects of batch import during data migration in the MARCedit software
2. To facilitate the functionalities and features of MarcEdit tools
3. To examine issues while doing Data migration (DM and ReCon with a smooth transition
4. To complete the process of importing bibliographic data into Koha
5. To covert the spreadsheet file into marc format

8.1. Scope of the study

1. It is essential to have a good understanding of each aspect of the data migration process for achieving the desired results. Decision made on one part of the project. Thus, once it is decided to adopt the data migration and retrospective conversion using MarcEdit in Koha. It is necessary to understand the scope of the project.

2. Deciding what areas of the collections will be converted;
3. Prioritizing the order of data conversion;
4. By whom and how the conversion is to be completed (Bharat, 2004)

8.2. Research design

The project will describe the process of data migration and retrospective conversion in the library automation world. While doing practical it will explore the challenges of retrospective conversion and data migration and some matters that are important for an ideal step.

This study drowns upon the practical experience of MarcEdit software also discusses the suggested tools of MarcEdit.

The study also describes importance of MARC21 fields in the migration process and their effective implementation in academic libraries by using the Koha demo. Koha's demo shows the advancements in open source, free integrated library management system (LMS) for cataloguing, circulation, flexible reporting, and automated library services especially in academic communities have gained extreme importance.

The project will record a self-attempting structured procedure that will capture by the student under guidance.

Moreover, a sequence of tasks will require completing the whole process. The major procedures will start as importing sample data into MarcEdit software, creating the spreadsheet, and uploading bibliographic entries in marc21 format.

The next part will demonstrate a step-by-step trial-and-error method involving several iterations detecting root cause, implementing corrective actions, and finally resolving the problem of data migration and recording them too.

The study will also be useful to interested individuals and Library personnel who should have adequate library science educational qualifications to do the DL work.

9. Results and Discussion

Library automation is a process that needs proper planning, timely implementation, and periodical evaluation. An automated library can provide better library services to its users and can maintain the library more properly which a manual library can't do. The record-keeping activities and various report generation become very easy in an automated library system. But the success of any library automation program depends upon its proper planning and execution. Hence library professionals need to take the right initiatives in the right direction. For the successful implementation of an integrated library system, all the key factors must be in place, support from administration, staff competence, consideration of user requirements, availability of the infrastructure, and excellent managerial skill of the co-coordinator of the project. With the emergence of Cloud-based ILMS accessible over web browsers, libraries have also started considering shifting from On-Premise (On-Prem) solutions to cloud ILMS as they require their infrastructure and involve up-front cost and recurring maintenance. Retrospective conversion of bibliographic records in libraries plays a crucial role in enhancing the migration of library resources to an online environment and improving the visibility of the library.

9.1. First step of data migration

First, prepare an excel sheet that is shorted on the bases of title to accumulate all unique titles in the same place.

The above screenshot defines different fields as like

1. Column A: ISBN
2. Column B: Language
3. Column C: Class no
4. Column D: Author Mark
5. Column E: Personal Name
6. Column F: Title
7. Column G Substitute
8. Column H: Statement of Responsibility
9. Column I: Edition
10. Column J: Place
11. Column K: Publisher
12. Column L: Year
13. Column M: Physical Description
14. Column N: Series
15. Column O: Volume
16. Column P: General note
17. Column Q: Bibliography
18. Column R: Summary
19. Column S: Sub Author
20. Column T: Location
21. Column U: Type
22. Column V: Price
23. Column W: Accession No

Convert your excel data into MARC format. The process is given below. See that the sheet given in the excel sheet contains all the bibliographical details about the document.

Now, open MarcEdit 7.5 and follow the procedure

Choose Delimited Text Translator and Select the Excel File. On the next page, choose excel or a text file that contains your original data. Then Select the delimiter, generally, it is 'comma'.

1. Step: - Click on the yellow symbol folder to select the input file (spreadsheet) from PC.
2. Step: - Click on the yellow folder with the green arrow to select the location where want to save the MRK file.
3. Step: - Click to select sheet number from the spreadsheet.

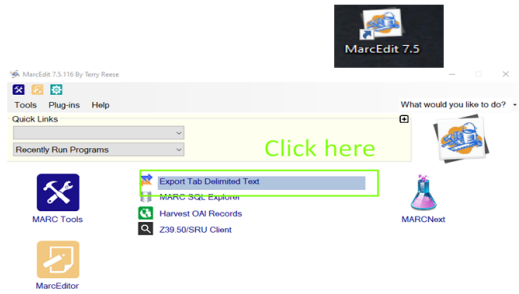

Fig. 1:

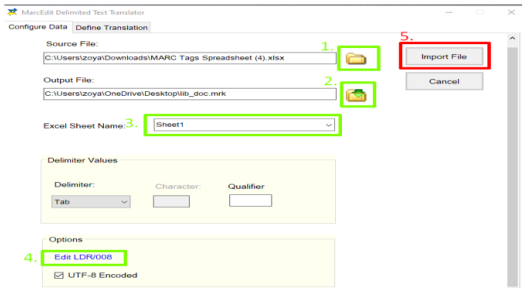

Fig. 2:

4. Step: - Click on EditLDR/008 and select book.
5. Step: - Click on import file button.

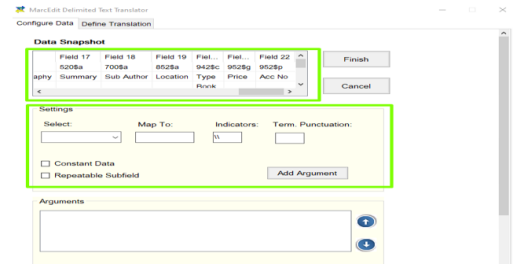

Fig. 3:

Now it is the most important part is mapping the field manually as per the MARC tag.

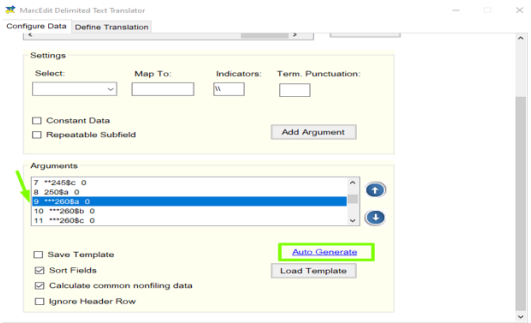

Fig. 4:

1. Otherwise, by clicking the Auto generates button it will automatically record data.
2. The arrow shows joint fields, forward of joints fields have * star symbols. (For e.g.: - 260a, 260b, 260c) these are joint fields.
3. After auto-generates, all data click on the Finish button then one dialogue box will appear to click OK.
4. Done, now MRK file looks like this.

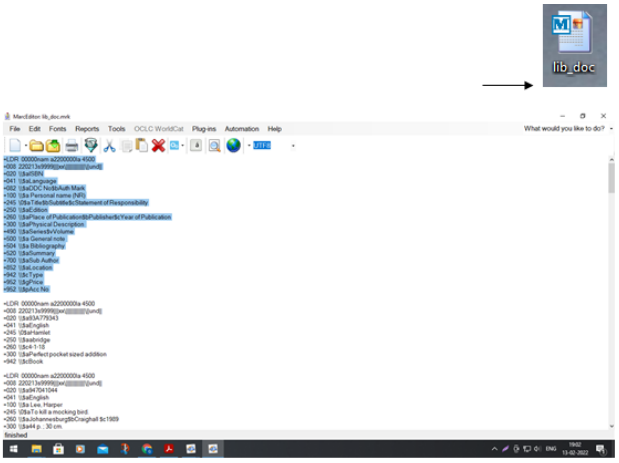

Fig. 5:

After opening of MRK file delete the first entries

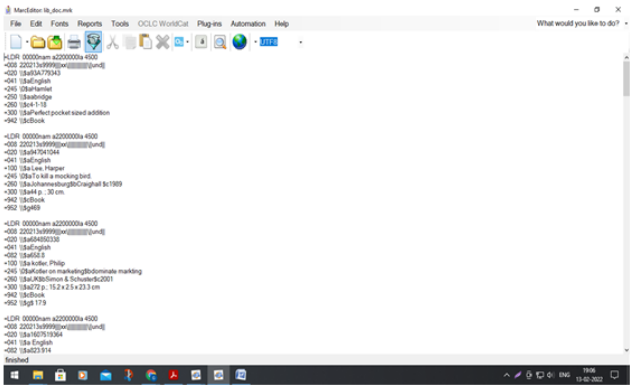

Fig. 6:

1. After completion, this is the final MRK file.
2. To convert. MRK file to. MRC click on Compile File into Marc.

9.2. Second step: Data migration in koha

9.2.1. Stage MARC records for import

Get there: More > Tools > Catalog > Stage MARC records for import.

This tool can be used to import both bibliographic and authority records that are saved in MARC format.

Importing records into Koha includes two steps. The first is to stagerecords for import.

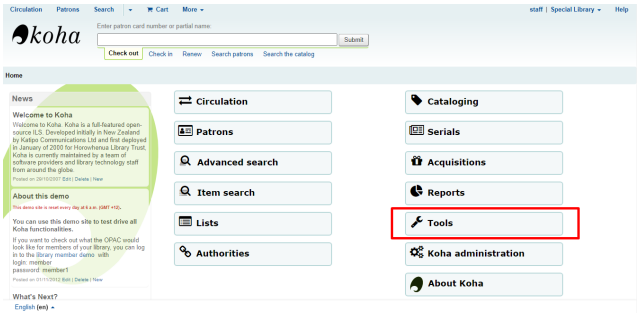

Fig. 7:

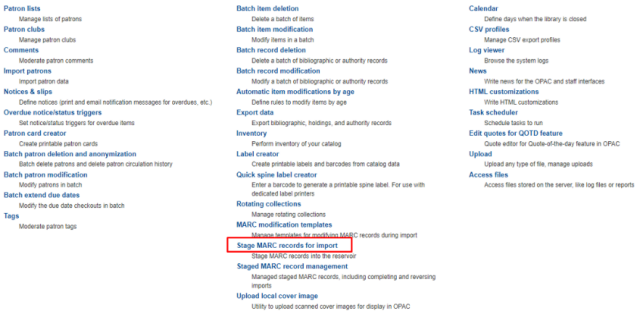

Fig. 8:

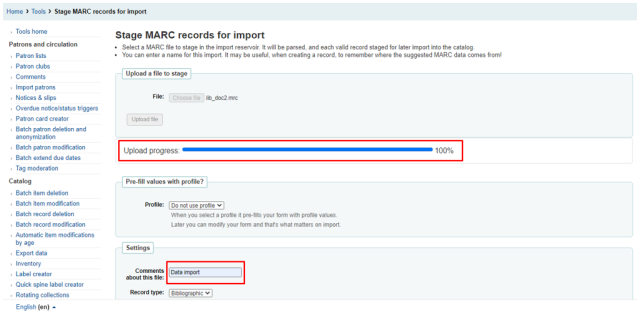

Fig. 9:

1. First, find the MARC file on the desktop then click > choose > upload file.
2. Comments about this file': enter comments to identify your upload when going to the 'Manage staged MARC records' tool.
3. Record type': tell Koha which type of file this is, bibliographic or authority,

Click 'Stage for import'

This will be presented with a confirmation of your MARC import

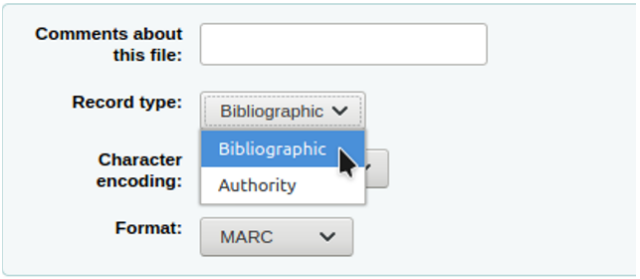

Fig. 10:

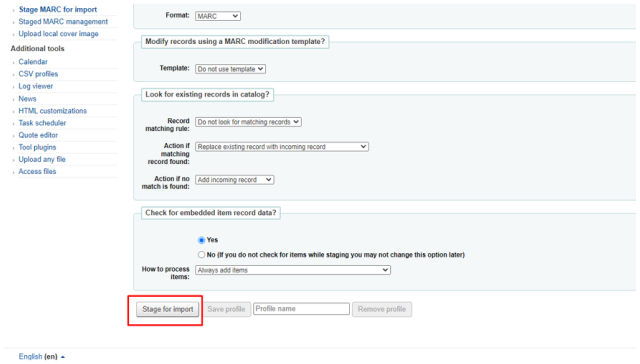

Fig. 11:

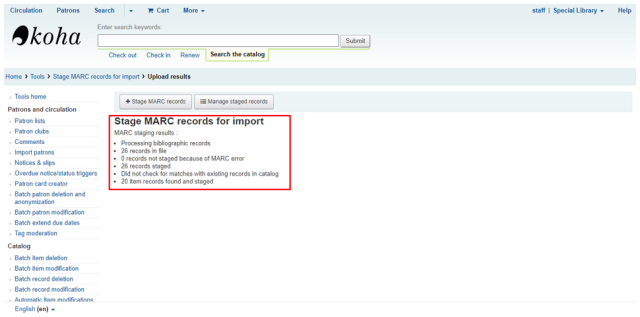

Fig. 12:

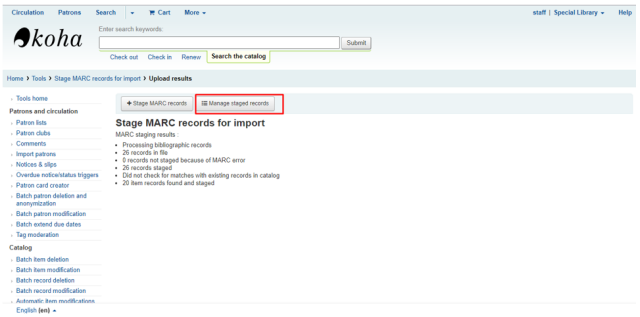

Fig. 13:

To complete the process, continue to the Manage staged MARC records tool by clicking on the 'Manage staged records' button.

9.2.2. Staged MARC record management

Get there: More > Tools > Catalog > Staged MARC record management

Once you have staged your records for import you can complete the import using this tool.

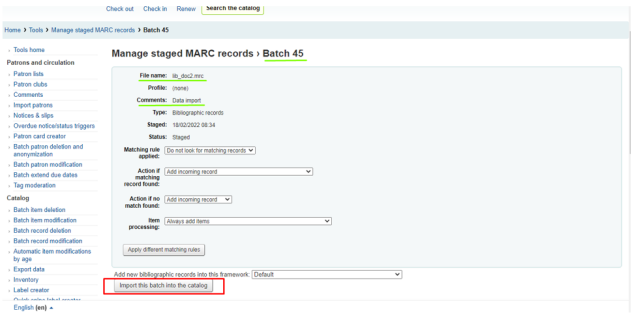

Fig. 14:

In this figure, all information is given (*see underlined part*) like batch number, file name, and a comment which was added in figure 13. Click on import batch into the catalogue.

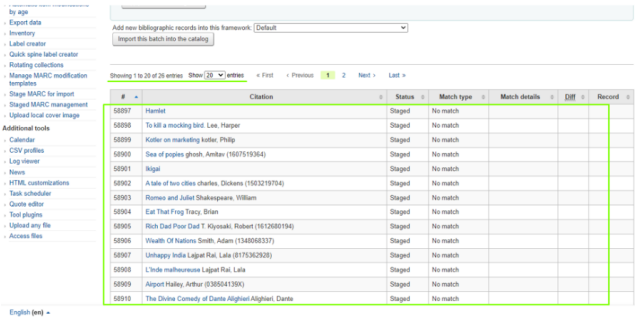

Fig. 15:

Once your import is complete a link to the new records will appear to the right of each title that was imported.

You can also undo your import by clicking the 'Undo import into catalog' button

For checking the successful task Click on search the catalogue and search any book from the imported entries, if it shows that same book it means the task is hundred percent done.

10. Findings and Suggestions

There are certain findings and suggestions based on those, future work can be carried out. These are: -

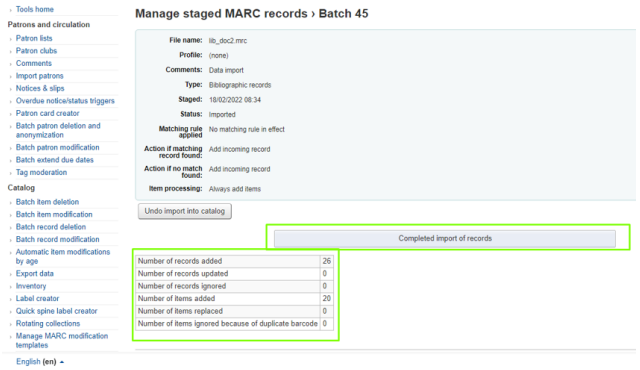

Fig. 16:

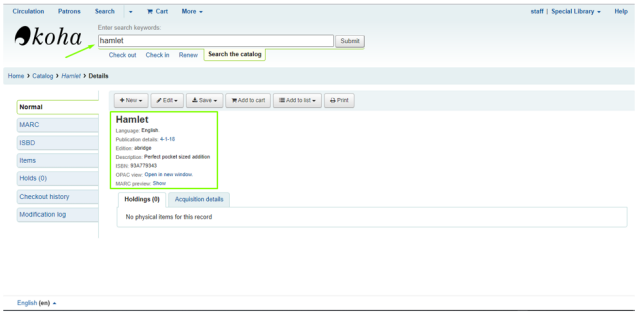

Fig. 17:

1. For successful doing data migration and retrospective conversion projects, there is a need for a detailed plan tempered with realistic expectations. The plan should Identify the objectives, carefully import data procedures and Specify standards to be used.
2. It is possible to examine import data migration to Koha with the help of MarcEdit software and open-source software which helps to access collections of documents very easily
3. At present, the efforts for importing data by clicking the Auto-generate button will automatically record data
4. There are no fixed policies prepared in respect of Retrospective Conversion using MarcEdit
5. The study findings reveal that MarcEdit software fully supports conversion of non-marc data in marc format for effective and accurate import to open source ILMS Koha making libraries independent and empowered.
6. Organize training program to develop an awareness of data migration in Koha and handling technical problems.
7. UGC or the Central government is encouraging the project of data migration and retro convention and providing training to college libraries to migrate library's bibliographic data into Koha OSS to bring grey resources in limelight.

11. Conclusion

Data Migration is a process of moving bibliographic data into a system that usually performed through programs. Technology has been perhaps continued to have a dramatic impact on library operations and services. It is the main force for changing the core work culture of library situation to maintain their record on the accession register into excel sheet of a computer. So, if a library wants automation, then needs to migrate the huge amounts of data into the database of a system. For reducing manpower, time, and money to apply the method of data migration. So, an automation system is very crucial in the library. The trends in technology will certainly find their way into a large library setup. Because, the library must satisfy the expectations of its end users to sustain their goals, objectives, and existence in the present techno-oriented world. In the past, they were compiled manually, a tedious job. Nowadays library systems come with a graphical report interface that enables point and clicks selection of appropriate data elements, periods, frequencies, output devices, etc.

ReCON involves a considerable amount of work, essentially acquiring existing databases and keyboarding all the existing entries of records error-checking, etc., or getting the job done from commercial vendors and creating a bibliographic database and transferring the records to machine-readable form. RECON is a costly and time-consuming task for most libraries. Therefore, it is to be planned properly for completing the task with the day-to-day activities of the library. Hence Cost-effective and efficient forethought and planning are essential.

12. Source of Funding

None.

13. Conflict of Interest

None.

References

1. Library Automation DLIS007. (n.d.). Available from: https://ccsuniversity.ac.in/bridge-library/pdf/Library-Automation.pdf.
2. Kent A. Encyclopaedia of Library and Information Science. New York: Marcel Dekker;. p. 800. Available from: https://www.google.co.in/books/edition/Encyclopedia_of_Library_and_Information/I7J5LGp42XwC?hl=en&gbpv=0.
3. Koha 21.11 Manual (en); 2021. Available from: https://koha-community.org/manual//21.11/en/html/.
4. Features. MarcEdit Development; 2013. Available from: https://marcedit.reeset.net/features.
5. Talend - a Leader in Data Integration & Data Integrity. *Data Migration: Strategy and Best Practices*. 2022;Available from: https://www.talend.com/.
6. Data migration from one library management system to another: A case study in India. Available from: https://egranthalaya.nic.in/Articles/migration.pdf.
7. Dabas B. Retrospective Conversion: Guidelines for Libraries and Information Centers; 2021. Available from: https://ir.inflibnet.ac.in/bitstream/1944/327/1/04cali_19.pdf.
8. Todd C. Librarian as data migrator: a functional pathway from Millennium to Koha - D-Scholarship@Pitt; 2018. Available from: https://doi.org/http://d-scholarship.pitt.edu/35264/1/Todd%2C%20Christopher%20R.%20-%20Librarian%20as%20Data%20Migrator%20-%20A%20Functional%20Pathway%20from%20Millennium%20to%20Koha.pdf.
9. Mitchell H. The Times, They Are Changing. *Tech Serv Q*. 2017;34(2):129–45.
10. Saikia M. Data Migration from Libsys to Koha ILS: An Experimental Study; 2019. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3396855.
11. Mishra VK, Rath DS. Data migration, RFID integration and implementation of Koha: a case study o. *EBSCOhost Ebscohostcom*. 2021;Available from: http://op.niscpr.res.in/index.php/ALIS/article/viewFile/37510/465479108.
12. Sarmah SS. Data migration. *Sci Technol*. 2018;8(1):1–10.
13. Dabas BK. Retrospective conversion: guidelines for libraries and information centers; 2004. Available from: https://ir.inflibnet.ac.in/bitstream/1944/327/1/04cali_19.pdf.
14. Pund M, Jain P. System and process for data transformation and migration from Libsys to Koha. *Int Res J Eng Technol*. 2016;3(4):690–702.
15. Kumar V. Migration from SOUL to KOHA: a learning experience. 2019;p. 33–7.

Author biography

Gulafsha , Student https://orcid.org/0000-0002-6827-4724

Anhad Narad, Student https://orcid.org/0000-0003-0308-6770

Pooja Mishra, Librarian https://orcid.org/0000-0002-2682-778

Atasi Chakravarty, Assistant Librarian https://orcid.org/0000-0001-7036-0827

Rupak Chakravarty, Professor https://orcid.org/0000-0001-5046-1663

Cite this article: G, Narad A, Mishra P, Chakravarty A, Chakravarty R. Data migration and retrospective conversion using marcedit: A study of Koha. *IP Indian J Libr Sci Inf Technol* 2023;8(1):5-13.