Bibliographic Data Migration from LibSys to Koha : Experience at St. Xavier's College, Kolkata.

Sougata Chattopadhyay[1] , Arup Sarkar[2]
Library Assistant, St. Xavier's College (Autonomous), Kolkata
E-mail: sougata_chattopadhyay@yahoo.co.in[1] , arupsarkar46@gmail.com[2]

Abstract:
Data migration is a process of transferring data from existing system to another system and it is divided into two processes: (a) extracting data from an existing system into an extracted file and (b) loading data from extracted file into the new system. Before data migration procedure first you check the database structure and tables of the koha software. After checking the data format, collect data from the libsys software. It includes bibliographical details of the documents, items details, transaction records and patron details. Then using MarcEdit software collected data has been formatted as per the koha software data base structure. Using koha's data import facility data are uploaded to koha database. Last and most important step is the checking data consistency. This step completes the data migration procedure and certified the koha to be run independently.
In this paper we tried to explain the step by step method of bibliographic data migration from libsys to koha, checking data consistency and the problems faced during data migration.

Keywords: Koha, Libsys, MarcEdit

Introduction:
St. Xavier's College, central library has switched from libsys to koha LMS software in 2014. During the migration process we visit many libraries like British Council Library to get suggestion on the problems we have faced. A major problem during the migration was the lack of export facility from libsys. In migration there is challenge but opportunity for rectifying existing problems and redesigning the work flow of the library. Complete data extraction is probably the most important and vital aspect of the migration process. Any mistake in the data migration process can be unrecoverable. Migration should be done without any loss of any valuable data accumulated over years. Since libsys does not provide any comprehensive export facility to extract complete bibliographic data in standard formats such as MARC21 or CSV (comma separated value). Hence all the functional units of the library should be thoroughly analyzed before proper migration plan.

Before implementation we prepare system study to test the feasibility of data migration to koha, installed a demo server in the library and imported entire bibliographic records of BEd departmental library books, along with item information for trial and testing purpose. Transaction history data including member details, books issued and date of return can also be migrated from libsys to koha. After checking data consistency we implement the method in our central library.

Data migration methodology:
Our library has collection around 60000+ documents. The bibliographic data is stored unscientifically in libsys software. We prepare a design for these bibliographic data to migrate from libsys to koha software.
The stapes are as follows-
1. Checking the database and tables of koha software
2. Collection of data from libsys software
3. Shorting the excel sheet data.

4. Rectifying the errors in data.
5. Matching the field and subfield columns of excel sheet as per koha software.
6. Transferring the data into the MARCEditor.
7. Converting excel to .MRC format.
8. Converting .MRC to .MRK format.
9. Exporting .MRK data to koha software.
10. Checking the data consistency

1. Checking the database and tables of koha:
The data base structure of koha is very simple and well documented. The following tables are mainly used to store library holding such as biblioitems (information related to bibliographic records in koha), biblioframewar (information about MARC frameworks) biblio (table that stores bibliographic information), branch transfer (information for items that are in transit between branches), deleted biblio (stores information about bibliographic records that have been deleted), item (holdings/item information) etc.

2. Collection of data from libsys:
Data in source format:
The data we have in libsys is in the format of text file. We have multiple files with accession number as a mandatory filed along with other fields.
File 1. a) Accession number (computerized) – used to identify unique title
b) Title of the book
c) Publisher name and place of publication separated by any delimiter for identification


```
St. Xavier's College,          TITLES BY ACCESSION NO.        Page : 1
Kolkata                        ----------------------         Date : 16/06/2017


      ------------------------------------------------------------------------------
      Accn No.          Title Details                                      Status
      ------------------------------------------------------------------------------
      H1        Indian economy: its development experience / Misra, S.     Reference
                K. --Mumbai: Himalaya Pub., 2009
                BB 330.954 MIS(IND).Ed27
                Accn Date: 22/03/2010;

      H2                            "                                      Reference
                Accn Date: 22/03/2010;

      H3                            "                                      Reference
                Accn Date: 22/03/2010;

      H4                            "                                      Reference
                Accn Date: 22/03/2010;

      H5        Microeconomics : theory and applications / Sen,            Reference
                Anindya. --New Delhi: Oxford University Press, 2007
```
Fig: 1 [Source Data Format]

File 2.   a) Accession number
b) Volume of book and year of publication separated by author any delimiter
for identification

c) Author of book
    File 3. a) Accession number
           b) Classification number
           c) Pages, Edition of book separated by some another delimiter for identification
           and so on.

The above snapshot will provide you source data format. Here we have multiple files in the same format.

After combining all the report file, keeping accession number is common a consolidated report is generated. Though the report gives an output in the form of text file, a script is written to convert data in excel format.

3. Shorting the excel sheet data:

Excel sheet is shorted on the bases of title to accumulate all unique titles at the same place.

	A	B	C	D	E	F	G	H	I	J	K	L
1	245		100	700	250		260			300	440	500	
2	245$a	245$c	100$a	700$a	250$a	260$a	260$b	260$c	300$a	440$a	500$a	020$a
3	TITLE PROPER	SOR	MAIN ENRTY	ADDED ENTRY	EDITIO N	PLACE OF PUB	NAME OF PUB	DATE OF PUB	EXTENT OF ITEM	SERIES	NOTE	ISBN
4	August 1914	Alexander Solzh	Solzhenitsyn, A	Glenny, Michael (tr.)		Middlesex	Penguin Books	1972	645p.			
5	Molecular biolo	James D. Watson... [et al.]		Watson, James	6th ed.	New York	Cold Spring Har	2007	xxxii, 841p. : ill.		Includes index	
6	Kaabyasamagra	Jeebonaananda	Daas, Jeebonaa	Gupta, Khetra (ed.)		Kolkata	Bharbi	2001	848p.			
7	Kireetee Amnib	Niharranjan Gup	Gupta, Niharranjan			Kolkata	Amar Sahitya Pr	1985	V.2 & 3 (458p.)			
8	Rachanaabalee	Ramesh Chandr	Datta, Ramesh Chandra			Calcutta	Saahitya Samsad		518p.			
9	Rachanaabalee	Aashaapurnaa I	Debee, Aashaapurnaa			Kolkata	Mitra & Ghoshl	2001	V.5 (vii, 660p.) : ill.			81-7293-708
10	Rachanaabalee	Aashaapurnaa I	Debee, Aashaapurnaa			Kolkata	Mitra & Ghoshl	2009	V.6 (584p.) : ill.			81-7293-750

Fig: 2

In the above screen shot [Fig. 2] defines different fields as like –
- ❖ Column A: Title Proper
- ❖ Column B: Statement of Responsibility
- ❖ Column C: Main Entry
- ❖ Column D: Added Entry
- ❖ Column E: Edition and so on.

4. Rectifying the errors in data:

After shorting all subject and titles next task is to
- I.) Merge all the duplicate entry title on one place.
- II.) Trimmed the extra space between the words, deleted all blank lines, removed headers, footers and blank rows, merge multi line text to single line.
- III.) Checking all the typographical and conceptual errors of the columns.
- IV.) Rectify any error in Indic language.
- V.) Rectify all the error related to classification scheme and cataloguing rules.
- VI.) Rectify all conceptual errors in the author field. For example "Goon", "Gun", "Roy", "Ray", rules of more than three authors etc.
- VII.) Standardize all the subject heading.

3

Now after removing all flaws like duplicate entry, typographical errors accumulate all the bibliographic data in a single file for migration. There is only one data type in each column (including subfields). Save it as Text (tab delimited) excel file with filename.

5. Matching the field and subfield columns of excel sheet as per MARC tag in koha software :

Koha, library automation software is compatible with MARC tags. For transferring bibliographic data in koha software, it needs to map with MARC tag and fields in our excel sheet.

Field Name	MARC Tag	Excel Sheet Column
Title Proper	245$a	A
Statement of Responsibility	245$c	B
Main Entry	100$a	C
Added Entry	700$a	D
Edition	250$a	E
Place of Publication	260$a	F
Name of Publisher	260$b	G
Date of Publication	260$c	H
Extent of Item	300$a	I
Series	440$a	J
Note	500$a	K
ISBN	020$a	L
Price	020$c	M
Call No	082$a	N
Language	041$a	O
Other Physical Details	300$b	P
Subject Heading	650$a	Q
Accession No	925$p	R

Table: 1 [Filed name and equivalent MARC tag in koha]

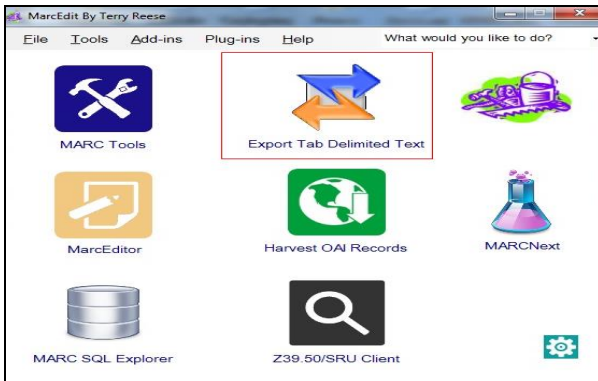## 6. Transferring the data into the MARCEditor
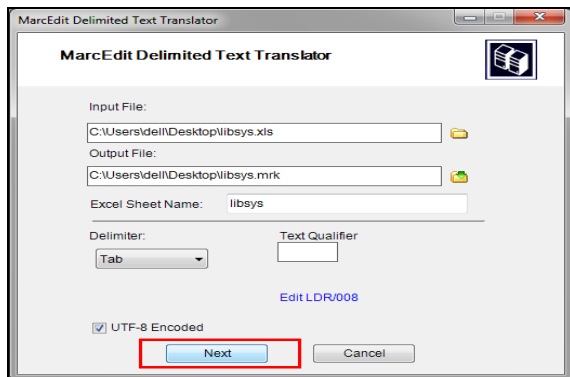

Fig: 3 [MarcEdit Tool]   Fig: 4 [Convert Excel Sheet into .mrk file]

Open MarcEdit 6.0 and select "Export tab delimited Text" icon [Fig 3]. Then browse excel sheet on source file which contains original data and specify output file name where you want to save the file, excel sheet name etc. And also specify LDR/008 as book.[Fig: 4]

7. Converting excel to .mrk format :

After completing field LRD/008 next task is to map excel sheet column to MARC field and sub field. For example field 0 that is fist column I map as 022$a (Valid ISSN for the continuing resources) and then click on apply. You can also customize indicator, term, punctuation etc. Similarly map all other fields and then click on Finish. It will bring next screen that Marc Text File (.mrk) has been created. [Fig 5]
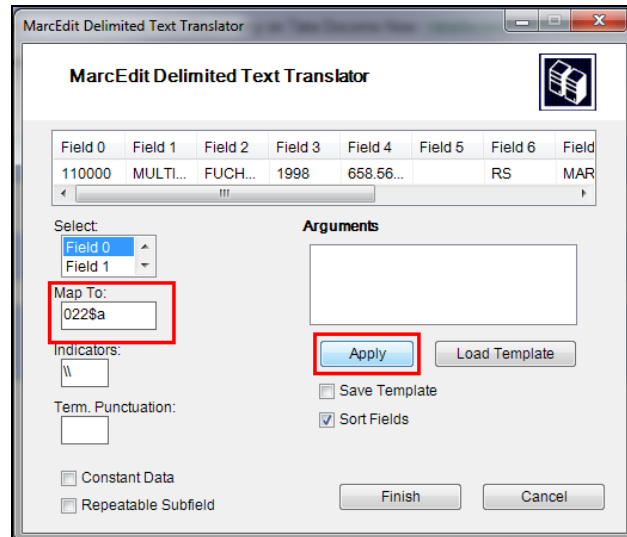

Fig: 5 [Mapping with marc tags]

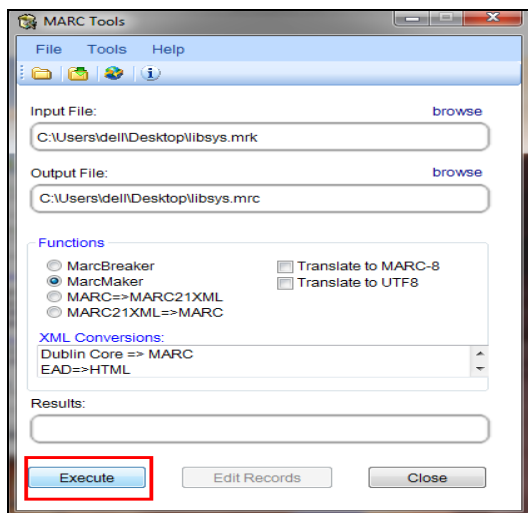## 8. Converting .mrk file to .mrc file format:

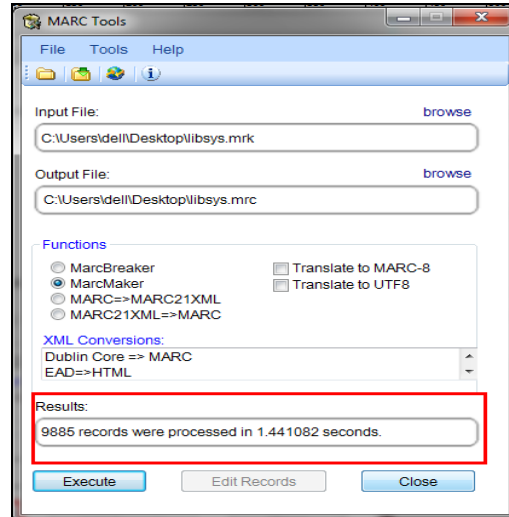

Fig : 6 Use MarcMaker   Fig : 7 Convert .mrk file to .mrc file

For these reopen MarcEditor and selected MARC tools then browsing input file and given output file name. Next select MarcMaker to convert .mrk file into .mrc format then click execute, it will show result. Click Close and now we have raw marc records wih us (.mrc file). [Fig : 6 , 7]

9. Importing .mrc file to koha LIS software

First login with **Koha User ID and Password** and get the main Interface of Koha. Select the option **'Tools'** and Choose the Option **'Stage Marc Records for Import'.** Now, choose the option and click on it. This tool can be used to import both bibliographic and authority

records that are saved in MARC format. Importing records into koha include two stapes. First to stage for import, Next options for matching and imports. So browse file to stage records into the reservoir.

| Fig: 8 | Fig: 9 |

Then upload file. Enter comments about file to identify your upload when going to manage staged MARC record tool. Select records type i.e. bibliographic or authority file record.

Next choose whether or not to import the item data found in the Marc record. Form here you can choose to always add items regardless of matching status, add then only if matching bib was found, add items only if there was no matching bib record, replace item if matching bib was found or ignore items and not add them.
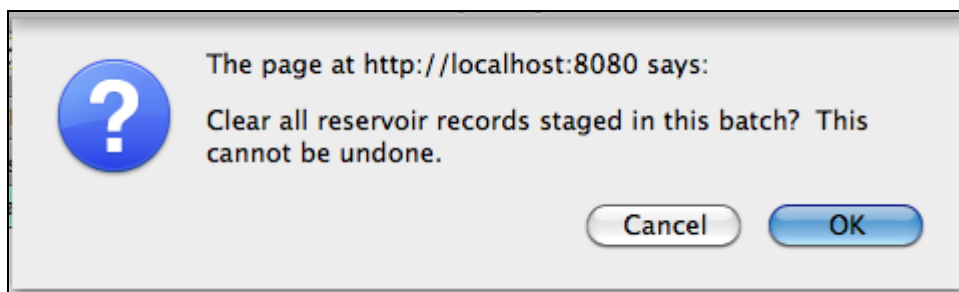

Manage staged MARC records

#	File name	Comments	Type	Status	Staged	# Records	# Items	Action
10	MarcRecords.mrc		Bibliographic	Imported	2016-03-09 17:44:44	27	0	Clean
7	PG-PHY_sougata_16-12-13.mrc		Bibliographic	Imported	2014-06-25 06:37:35	33	134 (Create label batch)	Clean
6	REPORTS(rectified).mrc		Bibliographic	Imported	2014-06-25 06:36:57	207	229 (Create label batch)	Clean
5	CD-DVD(rectified).mrc		Bibliographic	Imported	2014-06-25 06:34:33	819	1815 (Create label batch)	Clean
4	FINAL KOHA TRANSFER-21.06.14.mrc		Bibliographic	Imported	2014-06-25 06:03:04	40139	58509 (Create label batch)	Clean
3	BOOKS-BED(JHUNU)26.06.14.mrc		Bibliographic	Imported	2014-06-25 05:27:23	5390	7766 (Create label batch)	Clean
1	BOOKS-BED(JHUNU)12.03.14.mrc		Bibliographic	Reverted	2014-03-15 01:26:42	5390	7766 (Create label batch)	Clean

Fig: 10

- From the list of staged records, click on the file name that you want to finish importing
- You will note that records that have already been imported will show under 'Status'
- A summary of your import will appear along with the option to change your matching rules

To clean items out of the 'reservoir':
- Visit the main screen of the Manage Staged MARC Records tool
- To clean a batch, click the 'Clean' button to the right
- It will be presented a confirmation message

6

The page at http://localhost:8080 says:

Clear all reservoir records staged in this batch? This cannot be undone.

Cancel   OK

Accept the deletion and the records will be removed from the reservoir and the status will be changed to 'cleaned'.

Marc import is completed. It will import all the record into database. Finally build the index using following command. **sudo koha-rebuild-zebra –v –f library**

It will re-index and record will be available on the **Online Public Access Catalogue**.

10. Checking the data consistency:

After importing all bibliographic and item data a report i.e. accession register is generated by the koha's report generation tools. The report is checked randomly with libsys and koha software. Physical data checking is also done. This completes the data migration procedure.

Conclusion:

Converting data from existing library management software to another is not a simple task and can provide many challenges. Libsys does not provide any access to source code for users and the data can only be manipulated with front end programme.

During the data conversion, some of the records may get lost in part or full due to unstructured output from source file, absence of the key data field such as title, author, accession no etc. In our library, all bibliographic data and item information are transferred, except item information of serial i.e. vol. issue no are not transferred. In case of circulation only documents check out information are transferred but pending fines amount are not transferred. Hence all pending fines become excused. The above case study describes step by step process taken for data migration from libsys to koha with the help of web resources, checking data consistency and problems faced during data migration.

Acknowledgement:

The authors gratefully acknowledge for the valuable suggestions received from Dr. Xavier Savarimuthu, Asst Director, Central Library and Dr. Chinmoy Mukhopadhyay, Librarian, St. Xavier's College.

Bibliography:
1. http://www.koha-community.org
2. http://kohageek.blogspot.in
3. http://opensource-ils.cci.utk.edu/content/koha-migration-process
4. http://marcedit.reeset.net/about-marcedit
5. http://schema.koha-community.org
6. http://manual.koha-community.org/
7. http://www.libsys.co.in/

8. Karak, S & Dutta, K (2017), Bibliographic Data Migration in KOHA 3.18 from Existing Data Sheet : A Practical Experience*, International Research: Journal of Library & Information Science,* 7(1). Retrieved July 15, 2017, from *irjlis.com/wp-content/uploads/2017/05/8-IR-389.pdf*

9. Pund, M & Jain, P (2016), System and Process for Data Transformation and Migration from Libsys to Koha, *International Research Journal of Engineering and Technology (IRJET),*3(4). Retrieved July15, 2017, from *https://www.irjet.net/archives/V3/i4/IRJET-V3I4137.pdf*