

Cataloguing Issues and Development of the National Union Catalogue

The goal of contributing cataloguing data to the Jamaican National Union Catalogue and hence the major objective of NIS union catalogue is to facilitate the identification information materials held by the libraries of the national information system. Primarily this means:

1. Retrieval of bibliographic records of known items, indicating which libraries hold copies
2. Retrieval of bibliographic records of items dealing with a particular subject.

In this regard the cataloguing activities undertaken by the various cataloguers contributing records, aimed at describing and subject cataloguing these information materials is the first and most important step in the process. The quality work that is done at this stage will determine the capability of the information retrieval mechanism, be it a manual cabinet of catalogue drawers, an internal computerized database (OPAC) or a unified composite catalogue on the Internet.

The objective of a library catalogue in, whatever form, is much more ambitious than the retrieval capability that is common on the open Web. In the first place, the library catalogue deals with highly structured data, which is what descriptive cataloguing is all about. This formal structuring of the data in terms of an accepted convention of descriptive elements facilitates highly precise known item searches and retrieval. This means, basically that if you are looking for a particular known item, the retrieval mechanism can give you a precise response. It will identify it on the first attempt, that is it will not give you 15,000 possibilities. It also means that if the response is zero hit, you do not have to keep trying.

e.g. if the user is looking for the *FAO report on Jamaica 2005: post harvest crop loss*, published in Washington in 2006, in the series *FAO country reports: MT 2005/83*, the library catalogue is designed to give a precise response. This is

because: title information is in title field; subtitle information is in subtitle field; author information is in author field; place and date of publication are in appropriate fields; series information is in series field, etc.

On the other hand a search for this title in Google basic search yielded 2,170 documents. In the advanced search, asking it to return the exact wording or phrase yielded 10 documents. When the same title information was used in the existing national union catalogue, it correctly returned zero results because the document is hypothetical. Likewise when a real title, Post Harvest Losses in the Caribbean was tried in Google it returned 47,100 hits. When Google was asked to return exact wording or phrase it returned 7 documents. Of course the existing NUC yielded one result for the full title. We are looking at inbuilt precision and resultant confidence.

The concern in this context though is that if the data was not entered correctly in the appropriate field, we could not have had that confidence with regards to the precision of our results. Hence the integrity of data is the first requirement of quality control, relating to the cataloguing data submitted to the union catalogue.

The second functional capability is recall. We need to know that every item dealing with a particular subject, or by a particular author, or coming out of a particular publisher in a particular year has been returned. This is where web search engines would generally be strong, with high recall. However we need to know further that the returned results are relevant to our needs. This is where the intellectual input of the cataloguer will give us the confidence of the relevance of the returned results.

Certain techniques and functionalities on the design and implementation side of the online union catalogue can assist with this required relevance, which I will touch on later.

However, it is so important to ensure that high quality cataloguing data comes into the , that certain defining policy questions regarding the NUC needs to be address. Firstly, is it acceptable to take cataloguing data for the NUC only from libraries that have professional librarians? I do not have at my disposal, analysis of

contributors to the existing NUC that would indicate what percentage are professional librarians, paraprofessionals, subject specialists or clerical. This is a kind of immediate analysis that NLJ could easily arrive at when required.

To restrain the growth this way in favour of quality control, even though it would imply significant improvement, would still not remove the issue. Why? It is commonly understood that cataloguing and for that matter data entry are types of activities that require double checking or proofing. It is also understood that a significant number of special libraries have only one professional. It is also understood that because that librarian is often controlling the whole internal process, including interrogating the catalogue on behalf of the users, they can often function effectively with rough and ready tools. Thus there is still a real need for editing of the cataloguing data that gets to the NUC, even if it is coming from professionals.

Here also there are various approaches to be considered:

1. Insist that the cataloguing data be edited before being sent to the National Library of Jamaica. Obviously this would be slowest method of building the NUC. It also would not address the critical mass of records that are in the existing NUC.
2. Editing be afforded at the level of the sectoral networks, i.e. SECIN, STIN, COLINET, LINET, JADIN, AVIN and JARD. In fact SECIN is taking this responsibility for the SECIN sectoral database. The personnel for this could either be sponsored by local businesses or through an internationally funded project
3. The editing function is to be located at the National Library of Jamaica where coordination could be had with the systems staff responsible for development and maintenance. This could spin off in computerized approaches to some of the editing. This personnel could be either on budget, sponsored by local businesses or through an internationally funded project.
4. No general editing is considered. Each library's catalogue is provided on the network as it is. However individual editing should be facilitated by

the system, allowing professional pride to drive the process. Hence the system would grow up better over time as librarians edit their own databases. Other non – centralized technological models could facilitate this approach, such as Z39.50 interfaces that search a number of participating databases on the fly.

There are two other important elements of quality control that have to be address at the systems level of the NUC. In response to a search, the union catalogue will display multiple copies of records, one from each institution holding a copy of the record. This is tedious for the searcher and initially misleading. A search may return two pages of hits, which could actually turn out to be only three or four records held by many libraries. NLJ markets a duplicate removal utility which could be run against the NUC database to identify duplicates. However there is a need for an editor to look at the duplicates and decide which record has the best cataloguing data, to be used for display in search results. Then the location information can be taken from the others and placed on the best record,

Secondly, the natural inconsistencies in retrieval will obtain when a database is operating on cataloguing data from a variety of libraries using a variety of subject heading lists and thesauri. This is only compounded by multiple forms of personal names and corporate bodies. There is an obvious need for authority files to assist the contributing cataloguers in deciding on forms of personal names and corporate bodies. NLJ in fact have personal name, corporate name as well as subject authority files that could be made available online to these cataloguers. This would encourage them to bring these standards to the cataloguing data that will come into the NUC.

Models for Development

As best practices keep changing in response to available technologies, the way forward will involve looking out at some of the union catalogue models that

are often employed to optimize the retrieval of highly structured cataloguing data.

COPAC

The existing Jamaican NUC model is based on cataloguing records being exported from the online catalogue of contributing libraries to the union catalogue. This model is also used by the COPAC Library union catalogue in UK. This composite library catalogue “gives free access to the merged online catalogues of major University, Specialist, and National Libraries in the UK and Ireland, including the British Library”. According to its web site, “COPAC has 32 million records, representing the merged holdings of:

- members of the [Research Libraries UK \(RLUK\)](#). This includes the catalogues of the British Library, the National Library of Scotland, and the National Library of Wales / Llyfrgell Genedlaethol Cymru.
- increasing numbers of specialist libraries with collections of national research interest, as well as records for specialist collections held in UK academic libraries”

One of the major problems with this model generally, that of adding new libraries, often requiring complex export mechanisms involving record conversion , has been pre-empted in Jamaica by the use of a common software platform – CDS/ISIS for windows (WINISIS). Another problem with this model is the amount of time between a record entering a particular library’s catalogue and it getting represented in the union catalogue. This is more a real world problem than a structural problem as theoretically libraries could send small incremental files to central database weekly or even daily. The other attendant problem relates to updating amendments and deletions. This model is most appropriate in situations like in our national information system where the member libraries would want only minimal IT responsibilities. The system support would primarily be left to the central organization.

OLIS

Another approach, which could be appropriate for some of the other major database systems of the National Information System such as Jamaica Library Service and the University West Indies database system, would see the union

catalogue as the main catalogue for all the libraries. Aside from centralizing the cost of developing and maintaining the catalog, it assumes an acceptance of cataloguing and indexing policies by all, with no inconsistencies. It has attendant benefits of a single catalogue that is accessed by participants – reliability and easy maintainability. Of course if the union catalogue goes down, every participating library is without a catalogue. This is the model used by the Oxford University Union Catalogue, known as OLIS.

OCLC

The model used by OCLC is one where records are catalogued into a central catalogue and then imported by the local participating catalogue. All cataloguing and indexing policies are centralized. The local libraries decide which records they want. Issues related to the importation of records are resolved locally, therefore there is a need for some local systems maintenance capacity.

Virtual Union Catalogues

A model that is becoming increasingly common is the virtual union catalogue model. This involves the presentation of a single user interface, to the users, which searches the catalogues of the participating libraries and presents the results as if it is searching a single catalogue. The z39.50 client/server protocol, popular in the international library community, is often used to create this mechanism. This protocol allows a client software on a local computer to query a remote library catalogue running the protocol and return the result. This approach however requires local system support and it has all the recall and precision problems, typical of the open Web. This is because the system is actually searching across many databases with their own cataloguing policies and indexing files.

Features

Search Interface

There are certain features that are typically included in union catalogues to capitalize on the precision and recall capacities engendered by the structured data files and indexing. Most of these are also typical of ordinary online catalogues, which often allow direct transition of databases to union catalogues and a familiar search experience.

Basic and combination searches. (Implemented in NUC)

Boolean operators. (Implemented in NUC)

Field specific searches. (Implemented in NUC)

This is related to the indexing files used by particular databases. Even though use of same cataloguing codes imply similar index files this is not necessarily the case. Often certain databases may include related files in a single index file whereas another database may retain separate indexes for each cataloguing element of description. For example, a database designer may put all statement of responsibility data into an author index, whereas another may have placed personal name main entry data into a separate index file from personal name added entry data. This could mean that if you are searching across databases with an external search interface like Z39.50 you could return inconsistent responses. However in a model like that used by the Jamaican NUC the data files are re-indexed when the database are merged forcing consistency response across all collected datafiles.

Limits or filters (This feature is being implemented in the NUC as one of critical importance)

Many online catalogues and union catalogues (see COPAC) allow searches to be limited or narrowed by particular facets such as date published, place published, material types etc. However the limit or filter that is particularly useful to union catalogues is limit by particular contributing library. This enable the user to restrict the search to a particular library, often to bring a certain subject orientation to searches of composite databases that could not otherwise be easily achieved.

Sort results. (It is not implemented in the NUC)

This feature is of obvious utility in all types of online catalogues.

Display Features

Number of records displayed per page. (It is implemented in the NUC)

Locations symbols

This is critical for a union catalogue. (It is implemented in the NUC)

Marking or Selecting records. (It is not implemented in the NUC)

This allows records retrieved to be selected for saving, printing or to be emailed

Brief/Full Record Display. (It is implemented in the NUC but the choice must be made before the search.

Refine Search. (It is not implemented in the NUC)

This allows a search to be further refined or filtered when the displayed records are examined.

Hot Key or Clickable Subject Headings. (It is not implemented in the NUC)

This allows a search to be conducted on any subject heading in a record by clicking on it.

The conclusion

The groundwork for a National Catalogue has been laid. A critical mass of cataloguing records have been uploaded to a competent search platform. Much systems development work needs to be carried out to achieve a world class set of features. Duplicate records need to give way to a single record with holdings by libraries indicated. Editing of the records contributed by libraries is a critical need.