RIDING THE AGGREGATOR WAVE
AUTOMATING INTO A MARC CATALOGUE

ABSTRACT

Aggregations provide unprecedented access to thousands of electronic journal articles for library users. The volume, however, caused a problem that threatened to overwhelm cataloguers, who saw the need to provide swift, cost efficient and easy access to these resources. The traditional approach was rejected as too time consuming and expensive. The preferred option was automatic creation of brief MARC bibliographic records for individual titles in aggregations by using aggregator-supplied lists. Bulk loading of aggregators’ records and weekly reloading of the bibliographic records refresh the library’s catalogue. These processes are presented as the successful application of cost effective automated techniques.

INTRODUCTION

There is a wealth of valuable, scholarly, information available on the web to support teaching, learning and research, but it is often slow and cumbersome identifying and accessing it from the vast amount of unstructured information of variable quality, authority and stability in which it is subsumed. Many students are ill equipped to assess the quality of the information resources they retrieve, and academic staff and students are short of the time necessary to search for and evaluate information resources so, as Sarah Thomas (2000) puts it, the filtering and organising done by libraries has the potential to serve as a labour-saving device and productivity tool for researchers,

as the resource has been chosen for its relevance to past and present foci of the community of which the searcher is a member; [and] … the material possesses authenticity, in that the rigor of the selection process vouches in some way for its scholarly value.

BUDDY INTERFACE

One important way academic libraries add value to the research, learning and teaching activities in universities is to provide quality controlled subject based gateways for academic staff and students. Buddy is the University of Melbourne’s subject based gateway to its electronic resources. Information librarians appreciate the speed of data entry and the flexibility of Buddy. All the information that technical services staff enter in the default descriptors and remarks fields may be overridden by subject librarians for their own Buddy menus if they consider that more detailed or discipline specific information would be of greater value to their client groups. It also “offers many of the desirable features of the catalog, since the bibliographic control over its contents is carefully managed by librarians.” (Thomas 2000)

AGGREGATIONS

Aggregations are collections of publications in electronic form. Aggregators is a term that is used widely and loosely, to refer to a number of different entities:

- Serials vendors such as Swets Blackwell
- Article aggregators such as Proquest
- Publishers bundling all their titles into a set, for example Blackwell Science’s Synergy
- OVID, which aggregates both abstracting and indexing tools and full text journals
- Firms which provide technical support, including web loading for publishers not wanting to load their titles onto their own web pages, such as ingenta and Highwire.

In this paper the term is used to refer to both the databases which are compilations of full text articles, such as Proquest 5000 and Expanded Academic ASAP, and publisher compilations such as Elsevier’s ScienceDirect and Blackwell Science’s Synergy.

CATALOGUING OF AGGREGATIONS
The old saw, “fast, cheap and good – pick any two” no longer holds water at the University of Melbourne, at least in relation to cataloguing of electronic resources.

In these days of instant gratification, students want immediate access to electronic resources. Delays caused by undertaking original cataloguing according to current cataloguing rules would be unappreciated by the intended beneficiaries. Nor do the hundreds of titles on aggregations warrant the extensive work that original cataloguing entails, as these aggregations are volatile in nature, as are their records. Nor is the time for this level of cataloguing available in the present climate of tight staffing resources.

Thomas (2000) has pointed out that,

Some librarians have balked at the introduction of certain types of electronic resources into the catalog, particularly those likely to have transient URLs or which require heavy maintenance. The catalog represents stability, dependability, reliability, and quality. Its holdings have not typically been ephemeral in nature. It goes against the grain for librarians to invest in the creation of an expensive and detailed bibliographic record if the resource for which it is a surrogate, is not likely to endure for the foreseeable future, if not permanently.

Cataloguing rules were formulated in a period when a catalogue record acted as a surrogate for a book. People needed to know as much information as possible before they walked from the catalogue to the bookstacks or even to another library. Now the need to create such detailed records has disappeared, as the resources they describe are available from the desktop at the click of a mouse.

**CREATION OF BUDDY BIBLIOGRAPHIC RECORDS FROM AGGREGATORS’ LISTS**

Electronic record creation in Buddy is a two-stage process. The Buddy Technical Development Officer has written Perl scripts to enable automatic generation of Buddy records from digital files supplied by the aggregators. The information supplied in these files is limited, frequently comprising only title, ISSN, URL and (sometimes) holdings. Information from vendors is not as complete, consistent or accurate as desired. There is no title authority control between vendors, and missing or incorrect ISSN data are also problems.

Work that manually would take several weeks can be done in a day, making the titles instantly available to staff and students from the Buddy interface. The expectations of academic staff and students for rapid accessibility of information resources are thus met: hundreds of titles from aggregations are available on Buddy within 48 hours of receipt from the aggregator. Changes that occur regularly in these volatile databases can be just as expeditiously signaled.

It is advantageous to create Buddy records from each aggregator database in collection specific batches, as much of the data is the same, for example, name of aggregator, format and mode of access. All the Buddy records in a database are created at once, e.g. the over 700 *ScienceDirect* titles.

**Insert Powerpoints 1 and 2 here: 1. Aggregator record as a Buddy (Electronic database) record and 2. Buddy Record Continued**

Once the Buddy Technical Development Officer has downloaded into Buddy the brief records created from the aggregators’ lists, technical services library assistants flesh out these “skeleton” records by adding natural language subject headings, and default remarks (notes). Checks are also done in Innopac, the library’s catalogue, on the availability of other formats, and this is included in the record. Other information in the default remarks field includes former, varying or alternative title, and general notes on the resources. Original cataloguers match each of the natural language subject headings (default descriptors) to their Library of Congress equivalents to ready the records for loading into Innopac.

**Insert Powerpoint 3 here: 3. Buddy natural language descriptors mapped to Library of Congress Subject Headings.**
The collections within many of the aggregators’ databases are extremely volatile: titles may be added or dropped in each monthly update. The dynamic nature of the databases means that any attempt at original cataloguing beyond the collection level would be a Sisyphean task. The Buddy Technical Development Officer ensures that these updates are handled automatically and that unwelcome delays inherent in any attempts at original cataloguing at the title level are avoided.

The rapidity and ease of such record creation make the bibliographic records available from publishers or vendors of academic interest only. For much the same reasons purchasing records from Kinetica is equally unappealing.

SINGLE VERSUS MULTIPLE RECORD APPROACH

At the University of Melbourne, cataloguers have exercised their minds over the implications of both single and multiple records and both have been trialled. The major factor behind the decision to adopt the multiple record approach at the University of Melbourne was the purchase of a number of aggregations. Many of these databases are dynamic, with titles and date coverage changed by the host provider, sometimes without notification. Another issue is the level of title duplication between aggregations.

The automatic creation of non-MARC Buddy records from aggregator updates involves replacing an entire set of records each month. It makes sense to use the multiple record approach to avoid having to make changes manually each month to a single record, covering both print and electronic versions of the title. The implications on resources are enormous with the one record approach. While one record covering all formats saves duplication, it is more resource hungry, as we would not be able to create records from Buddy automatically if we went down that path.

In some libraries single records are preferred when the online version has significant additional content not present in the print version. The choice of a separate record approach in some cases indicates that the versions are not considered equivalent, and the difference of the online version from the print is significant to users.

Insert Powerpoint 4 here: 4. Multiple titles in Innopac

URLS

The decision was taken to make the URL link directly to the web page of the resource in Buddy, not to a vendor’s or publisher’s page – as that requires several more clicks to access the resource. This vastly increases the patron’s access to the specified resource (Herrera and Aldana 2001, p 246). In cases where all the titles on an aggregator’s site point to only one URL, then the URL is scripted so that each title is linked to its own unique URL.

The Buddy URL comprises the Buddy server name and the number of the Buddy record. This Buddy URL is the one that becomes the URL for the Buddy record in Innopac, providing the mechanism for the authentication and authorisation required by the licences for the subscribed databases.

All records in Buddy, including those from aggregations, are checked by the homegrown link checker nightly, and a report is generated dynamically on the web. Technical services staff can see when URLs have come up as 404s and then manually check them and fix problems. For the most part, aggregators’ records have been relatively problem free, although there have been a few hiccups along the way.

BUDDY INTO INNOPAC

University of Melbourne creates MARC records from the listings of full-text titles on the aggregations. The Buddy web database record allows for the creation of interpreted MARC records from an access database. These interpreted MARC records are then run through our homegrown Perl script which acts like the Library of Congress’s MARCMakr programme (http://lcweb.loc.gov/marc/marcmakr.html) to create transmission format MARC records for loading into an online catalogue. The MARC records created are automatically loaded weekly from Buddy into Innopac through a load table purchased from Innovative Interfaces. This has enabled bulk record input and overlay, without disturbing other records in the catalogue.

Insert Powerpoint 5: 5. MARC tags on Buddy into Innopac
It was decided that the best method to achieve the updating of Innopac was by loading a full file each week and overlaying all records as appropriate. The load table has been set for records to overlay on the 001 Tag (BUD number) without creating an item. If a match is found only the bib will overlay. If no match is found, then this process provides a new bibliographic record and item. Any records not matched in this process is removed from Innopac. This process is quicker than editing both the Buddy file and the Innopac file. The load table makes the URL “live” as part of the load process.

In WebPac, time out was extended from 5 minutes to 20 minutes to allow users more time to look at the electronic resources. Most information librarians have been pleased with the time out period, as it allows users to obtain the references they need while not tying up the resource for an extended period. For more intensive use of electronic resources, users are encouraged to transfer to Buddy.

**ELEMENTS OF THE BIBLIOGRAPHIC RECORD**

Decisions had to be made on the bibliographic record to be created, including location codes, call numbers, item status codes, and material types. As all Buddy records from an aggregation are created together, the fixed fields are obviously of some concern. The Innopac BUD record number is the number of the Buddy record. The language is English, bibliographic level is “s”, as most of the items are serials (although that is changing with the growth of e-books, audio lectures and other electronic resources). An encoding level “b” for Buddy records was created, and the material type is “m” computer file”. This could be translated to “electronic resources” in the list of limiters for the Opac user, as the Buddy records are rapidly including all other types of electronic resources.

Each bibliographic record and item record has a location code “UniMelb Internet Resource”. Each item record has a status code of “available”. It was decided not to include a call number.

Each field in the electronic database was mapped to a MARC tag to be loaded into Innopac.

**Insert Powerpoint 6: 6. Aggregator record in Innopac**

**Title Proper**

245 Title|h[electronic]|c(Exp.Acad)

When dealing with the aggregator databases, the title proper was the title used by the vendor. This was the “hook” used for overlaying the record in each monthly update. When a title is available from a number of different aggregators, separate records are created for each manifestation of the title, adding the name of the aggregator as a qualifier, e.g. Theater journal (Project Muse), Theater journal (Exp. Acad).

**GMD**

The issue of the General Material Designation has been a subject of some intensive discussion. “Computer file” has caused confusion in the minds of our users, and the term “Electronic resources” has not yet been approved. We have used the GMD “electronic” to alert our users to the fact that they are looking at a record with Internet access, and have taught our information librarians the meaning of the term.

There still are problems because, due to a decision by Library of Congress, we cannot use a GMD for text. To do so would help users, as they find it confusing to distinguish between formats in a multiformat catalogue.

**022 ISSN**

Often the ISSNs in the databases may refer to either print or electronic versions, or are used interchangeably. However, because this field is also used as a “hook” for any changes, most are accepted as found, unless a line of “00000000” or odd letters appear, in which case the publishers are informed.

**246 Variant titles**

The 246 field comes from the “also known as” from the beginning of the default remarks field. It is created to provide an extra access point.

**260 Publisher**
In cases where there are publishers and holdings in the Buddy database these will automatically be entered into Innopac. It was decided to include holdings in the 260 c subfield so that they display in the second title screen.

**500 Note**  
It was decided that a note indicating that an electronic item is only available through WebPac would be very helpful to the many users of the catalogue who still use the Telnet access. The wording of the note is: “This is a record with electronic access and is only available through the Library's Web catalogue or through the Internet.”

**520 Summary**  
Since the contents of this field are available for keyword searching, it was decided to include the following in all Innopac records:

- Default remarks - (not including the “also known as” which will be in the 246 field.) This field includes information about the journal, its subject and scope.
- Service provider (this includes the name of the publishers and aggregators)
- Buddy descriptors, prefixed by the word “Buddy descriptors”. Some information librarians wanted these kept in the record, even though the words had been mapped to Library of Congress Subject Headings.

**655 Genre notes taken from the Coverage field in Buddy**  
E.g., 655 Web sites  
The coverage field in the database was reserved for genre notes to group 16 different formats, ranging from archives, bibliography, and directory to lectures, tutorials, and web sites. Items can be retrieved in the Buddy database by format. This field is also used to automatically create the genre notes.

**965 Institution Specific Subject Heading**  
An institution specific subject “Internet access to Buddy formats” was introduced as a method of alerting users to the different formats in the electronic database, and has been warmly welcomed by the information librarians.

**Insert PowerPoint 7: 7. Institution Specific Subject Headings**

**530 Additional physical form entry. Other physical formats taken from bib numbers**  
E.g., Library also holds print, CD-ROM formats  
530 note field is automatically added to all versions of the bibliographic record with text about the availability of other formats. This information is taken from the Innopac format notes in the Buddy record.

**538 Mode of access (taken from Mode of Access in Buddy)**  
A “mode of access” note is automatically created for each record to explain the means by which the electronic resource can be accessed. The wording of the note is: Mode of access: Internet

**650 Library of Congress Subject Heading**  
A separate Library of Congress subject heading is matched to each Buddy natural language descriptor. When the Innopac record is created, the 650 fields are automatically entered in the record.

**856 URL**  
The 856 field is used as a “hot link” to connect the bibliographic record directly to the online resource. For example,  
856 40 [uhttp://xena.lib.unimelb.edu.au/cgi-bin/buddyrun.cgi/NET/3940|zConnect (University of Melbourne only)]

The URL is that of the number of the record in Buddy. This URL has the added protection of authentication, authorisation, and automatic link checking. In Innopac the default wording is “Click on here to”. As our z subfield or local note we have decided to have “Connect (University of Melbourne Only)” for titles restricted to University of Melbourne staff and students. To provide the licensed resources, the library must ensure that remote access to the aggregators is limited to staff and students of the University.
ACCESS MANAGEMENT

Authorisation and authentication are key concepts in digital library provision. The words are usually conjoined, and used almost synonymously, but in fact have distinct meanings. Both are subsumed under the wider term of access management. Authentication procedures are used to ensure that users are who they claim to be, whereas authorisation procedures determine whether users are entitled to particular services. Peter Wynne et al (2001) assert that, “The twin issues of authentication and copyright are arguably the greatest barriers to exploiting the potential of hybrid or electronic libraries”. They go on to state that

The concepts of authentication (and authorisation) and copyright are poorly understood and therefore resented. When licensing restrictions are added into the equation, it can appear that it is the library service itself which is causing problems.

The goal in access management is to let authorised staff and students access information from anywhere at anytime without making the process too difficult for the library, the information provider or, most importantly, the library’s customers, while complying with the legal requirement of the publisher that the resource is secured against unauthorised access. With the increase in numbers of students studying from off campus locations, the problem of allowing authorised users to access licensed information products from anywhere at anytime has become crucial. (The latest statistics from Buddy showed that 68% of accesses were from outside the library. This percentage has been growing steadily since statistics were first kept.) We use the Innopac patron file as the means of authentication as this is updated regularly from Human Resources.

Some of the library’s most important (and expensive) databases are only available via IP recognition. This denies access to staff and students using a commercial ISP so, after consultation with information technology staff, the library has purchased a licence for EZProxy. EZProxy ([http://www.usefulutilities.com/ezproxy/](http://www.usefulutilities.com/ezproxy/)) is an easy to set up and easy to maintain software tool for providing staff and students with remote access to web-based licensed databases. The Buddy Technical Development Officer has channelled these databases through the EZProxy server to enable access from off the IP range, while still ensuring that users must authenticate themselves.

LICENCES

Academic libraries require flexible licensing arrangements. These include a broad definition of an institution which is not geographically specific - recognition that a university may have students studying anywhere in the world. Present licensing difficulties include arrangements which require a separate licence for each campus, restrictions on access by members of the institution from sites which are not part of the institution, limitations on access by non institutional clients using electronic journals in the library and restrictions on provision of interlibrary loan through use of electronic journals. Evaluating and negotiating licences is time consuming and tedious, but essential to ensure that they meet the library’s requirements.

The Electronic Information Coordinator maintains a licence file on a shared network drive with details of each database, indicating who may use it, and whether it allows interlibrary loan, coursepacks, etc. This is accessible to library staff.

PASSWORDS

After authentication to use Buddy, staff and students do not usually need to enter further passwords to access individual resources. Where it has not proved technically possible to achieve this and passwords are required, these can be obtained by following onscreen instructions.

CONCLUSION

With the inclusion of Buddy records into Innopac, a one-stop shop has been achieved for our customers and staff. User comments have been favourable, and information librarians like having one interface for all University of Melbourne resources.
The techniques developed in consultation with the Buddy Technical Development Officer have ensured fast, simple and cost efficient access to electronic resources. Our print and electronic resources have been integrated into one catalogue, in a timely and affordable manner. Our catalogue no longer merely reflects the physical collection but is also a gateway to the library’s electronic resources.

The conservative “business as usual” perspectives of libraries must shift to “business unusual” – radical changes in the catalog, its role and its composition, are needed for the library to endure in the digital age. (Lagoze 2000)

REFERENCES


1. Aggregator record as a Buddy (Electronic database) record
2. Buddy Record Continued

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode of Access</td>
<td>WWW IP based Login</td>
</tr>
<tr>
<td>Service Provider</td>
<td>SEARCHBANK EXPANDED ACADEMIC ASAP</td>
</tr>
<tr>
<td>Default Description</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Default Remarks</td>
<td>Also known as: Curr. Anthropol. &lt;br&gt;Former Title: Yearbook</td>
</tr>
<tr>
<td>Expiry Date</td>
<td>[41]</td>
</tr>
<tr>
<td>Licence</td>
<td>[20]</td>
</tr>
<tr>
<td>Subscription</td>
<td>[42]</td>
</tr>
<tr>
<td>Password</td>
<td>[63]</td>
</tr>
<tr>
<td>Default Help Files</td>
<td>[16]</td>
</tr>
<tr>
<td>Help Files</td>
<td>[18]</td>
</tr>
<tr>
<td>Where Held</td>
<td>[17]</td>
</tr>
<tr>
<td>Remarks</td>
<td>[21]</td>
</tr>
<tr>
<td>Comments: gb</td>
<td>[22]</td>
</tr>
<tr>
<td>Comments: eng</td>
<td>[23]</td>
</tr>
<tr>
<td>Comments: ban</td>
<td>[24]</td>
</tr>
<tr>
<td>Comments: med</td>
<td>[25]</td>
</tr>
</tbody>
</table>
3. Buddy natural language descriptors mapped to Library of Congress Subject Headings

<table>
<thead>
<tr>
<th>Buddy Descriptor</th>
<th>Library of congress mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accounting</td>
<td>Accounting</td>
</tr>
<tr>
<td>Administrative Law</td>
<td>Administrative law</td>
</tr>
<tr>
<td>Advertising</td>
<td>Advertising</td>
</tr>
<tr>
<td>African Studies</td>
<td>Africa</td>
</tr>
<tr>
<td>Agriculture</td>
<td>Agriculture</td>
</tr>
<tr>
<td>Airlines</td>
<td>Airlines</td>
</tr>
<tr>
<td>American History</td>
<td>United States -- History</td>
</tr>
<tr>
<td>American Studies</td>
<td>United States</td>
</tr>
<tr>
<td>Anatomy</td>
<td>Anatomy</td>
</tr>
<tr>
<td>Animal Health</td>
<td>Animal health</td>
</tr>
<tr>
<td>Animal Husbandry</td>
<td>Animal culture ; Livestock</td>
</tr>
<tr>
<td>Anthropology</td>
<td>Anthropology</td>
</tr>
<tr>
<td>Archaeology</td>
<td>Archaeology</td>
</tr>
<tr>
<td>Architecture</td>
<td>Architecture</td>
</tr>
</tbody>
</table>
4. Multiple titles in Innopac

<table>
<thead>
<tr>
<th>Num</th>
<th>Mark</th>
<th>Title</th>
<th>Collection</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Current Anthropology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Current Anthropology / [Microform]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>Current Anthropology / [Computer File]</td>
<td>[1991-]</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Current Anthropology / [Electronic]</td>
<td>2/1992-</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Current Anthropology / [Electronic] (PQD)</td>
<td>12/198+</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>Current Anthropology / [Electronic] (UCP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Multiple titles in Innopac
5. MARC tags on Buddy into Innopac

<table>
<thead>
<tr>
<th>LOCATION</th>
<th>CALL NO</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UniM INTERNET Resource</td>
<td></td>
<td>AVAILABLE</td>
</tr>
</tbody>
</table>
6. Aggregator record in Innopac
7. Institution Specific Subject Headings

Authors.
Eve Young, Electronic Resources Cataloguer, University of Melbourne
Shirley Sullivan, Electronic Information Coordinator, University of Melbourne.

---

[Note 1] A list of natural language subject headings was compiled by information librarians, as they preferred these to the more formal LCSH. Technical services staff allocate subjects to the Buddy records from this list.

[Note 2] To do that would entail extra work: searching for the record, downloading it, adding a note to the checkin record on Innopac, adding the URL to the record, adding holdings, and a call number. Then, when the aggregators delete that title from the database list, or if the holding is changed, the record requires editing. All this activity is labour intensive.

[Note 3] The way that the programme is set up, one can get a MARC or XML view of the record. With XML there is no associated DTD (document type definition) or style processing instructions and it requires an XML capable browser (Internet Explorer 5). This has implications if we decide to go with a common user interface programme.

[Note 4] When a uniform title is needed to distinguish an online version from a print version, it is conceivable to use the qualifier, “Online” but using this qualifier with multiple titles from different aggregators could create more confusion than clarity, so the decision was taken not to use them.

[Note 5] Changes are not made inhouse as that could prevent the regular updating of the record.

[Note 6] Remote access resources may contain variant titles such as: “at head of title” phrases, running titles, or abbreviated titles in header information.
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s: 
YOUNG, EM; SULLIVAN, SA

Title: 
Riding the aggregator wave: automating into a MARC catalogue

Date: 
2001

Citation: 
YOUNG, EM; SULLIVAN, SA, Riding the aggregator wave: automating into a MARC catalogue, Seachange: cataloguing in a dot com world - 14th National Cataloguing Conference preprints, Geelong Waterfront Campus, Deakin University (8-9 November 2001), 2001, (2001), pp. 80 - 91

Publication Status: 
Published

Persistent Link: 
http://hdl.handle.net/11343/33736

File Description: 
Riding the aggregator wave: automating into a MARC catalogue

Terms and Conditions: 
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.