Redesigning the OPAC: Moving Outside the ILMS[1]

LLOYD O SOKVITNE

ABSTRACT: The need to develop new OPACs that meet the needs of users better is becoming well acknowledged. How libraries can transform their OPACs to modern web friendly formats is not yet clear. This paper describes the State Library of Tasmania's experience in developing a new OPAC by exporting bibliographic data to a software package outside the ILMS. Issues involved with the provision of a user-oriented, browsable, or facet-driven interface are described, the two most important ones for libraries being how to change cataloguing processes so as to deliver the required data and search functionality, and how to share evaluative and usage data about our resources.

Recent years have seen the emergence of strong competitors to the role of the library as an information provider. These competitors include search engines, online encyclopedias and bookstores, and, unlike libraries, they have been able to develop rapidly and unfettered by past approaches or entrenched service delivery models. Behind these new services are hard commercial drivers that are related to attracting users on the web, among them the need to be fast, efficient, easy, and relevant to users. Libraries have not had such hard business drivers; consequently, our information services have been much slower to adapt.

A belated response across libraries, based on acknowledging that our current toolsets and services are no longer adequate to meet these new world challenges, is now gathering momentum. It is becoming accepted that libraries need to develop new tools that are appropriate to a networked web world increasingly populated by Generation Y users moving in a wide-ranging online social environment supported by technology based on integrated and on-demand content from various websites (mashups).[2]

Current moves that recognise a new web environment (Web 2.0) and the need to develop a new library environment (Library 2.0) extend beyond tools to simply improve remote access to library services. These include initiatives to make library resources more readily discoverable on the general web (e.g. opening library holdings to search engines), to communicate more effectively to users on the web (e.g. blogs), to provide relevant new services (e.g. podcasts, gaming), and to allow increased user interaction (e.g. tagging, user reviews).[3]

Significant debate is beginning to emerge about the methodology and approach necessary to make library tools such as the Online Public access Catalogue (OPAC) more web friendly, easier to use, and compatible with external web services.[4] It is recognised that we need to do more than just 'put lipstick on a pig'[5], but relatively few production examples exist as to what

Lloyd O. Sokvitne, Senior Manager (Digital Strategies), State Library of Tasmania, Hobart, Tasmania 7000. Email: lloyd.sokvitne@education.tas.gov.au

should be done and how it could be achieved.[6] This paper is an exploration of the issues involved with OPAC redesign in the modern web environment, and is based on the real-world issues experienced by the State Library of Tasmania in developing a replacement for its Integrated Library Management System (ILMS)-based OPAC.

OPAC Beginnings

The library OPAC developed as a tool to locate those physical resources that had been acquired and stored by the library itself. Within the ILMS these resources were, first, described within the acquisitions and bibliographic database, and then accessed by the public via the OPAC. The OPAC and the bibliographic database were two aspects of the same overall software. OPAC design and functionality over the past ten years have seen a number of important changes. These include adapting bibliographic database capacity so as to include or reference non-print materials and digital content, such as images and full text on a local network, and, more recently, to link to wide-area networks and open web content.

Access on the OPAC to this new content still required the existence of a descriptive record in the bibliographic database. The extensive work required to create new bibliographic records strictly limited the range of non-library holdings that could be indexed or catalogued and subsequently referenced through the OPAC. For this and other reasons, libraries have frequently developed other resource collections outside the ILMS that have separate indexing and access methods. Examples of these external collections include images and other digitisation datastores, lists of recommended web resources, and online journal collections.

The growth of modern technology and high-speed networks now mean that it is possible to separate the OPAC from the bibliographic database used by the library to manage its holdings. The capacity to extract, move, and import large amounts of data in modern computer systems is well developed. This has been coupled with the capacity of new search-engine type applications to efficiently index large volumes of data to provide powerful real-time searching and results handling. In this new context, an OPAC need not necessarily be part of the ILMS, and as a consequence, need not be limited by traditional ILMS-based OPAC functionality.

The State Library of Tasmania has now embarked upon this path of separating the OPAC from the ILMS. In 2005/06 a tender was issued and completed to acquire new software that could index data exported from the library's bibliographic database, as well as from other datastores managed by the library. This software was required to provide powerful searching capacity and to allow a user-friendly web-orientated approach to searching, results management, and resource access.

A software package from Autonomy called Verity K2[7] was chosen and the State Library is now in the process of designing new discovery services based on that software. In tackling this process, the State Library has decided to initially develop two distinct information search and retrieval services:

- A replacement of the current OPAC, providing focused access to the tangible (physical) resources held or controlled by the library. As the State Library of Tasmania provides a reference and public library network spanning over 50 libraries, access to these physical resources also includes specific user needs to access library and shelf location information, loan status, etc.
- A new search tool or gateway to provide integrated access to the full range of information resources provided by or indexed by the State Library, irrespective of format or material type. These resources include material stored or indexed by the State Library outside the ILMS, and include a digitised images photographic collection, an eHeritage collection, and collections of preserved websites and electronic documents.

Both of these search services will be based on a regular export of data from existing data sources. For the OPAC, this export would occur as a major exercise at the beginning of the process, with regular exports or updates (based on identifying data that had changed in the bibliographic database) on a daily basis. For the gateway, full exports may be received and processed daily for smaller data sources (such as the images database).

Experiments have proved that the capacity of the software to index and update large data volumes was well within the operational requirements of both the OPAC and the gateway. The other OPAC data issue – that of providing current copy and circulation status information – would not be handled by export routines, but by real-time queries into the circulation system at time of need. In a similar way, user requests to place holds on items found on the new OPAC would initiate a transfer by that software of the user into the existing holds system (although ultimately this may be done within the new software as well).

Designing and Developing a New OPAC

When beginning work on the new OPAC, the State Library adopted the philosophy that this redesign would be an ongoing process, and always be a work in progress. The goal was to implement the software as soon as practical, but then monitor and add or change functionality over time, based on user behaviour. Although there were data, ILMS, structural, and resource issues to consider, the decision was taken to begin the design process around the need to satisfy real user needs and behaviours.

Starting with the User

Although recent debates and initiatives have made it abundantly clear that our OPACs 'suck'[8], this is an attribute that has been a part of OPAC operation well before the web emerged. Library OPACs have always been difficult to use and have produced poor outcomes for users.[9] Librarians have tended to avoid this issue by emphasising the importance of user training and the good results that can emerge when OPACs are used correctly (e.g. by librarians). It is not before time that libraries are returning to an analysis of real-life user behaviour as the basis for OPAC design.[10]

When the State Library of Tasmania reviewed how most users searched for library materials in the physical world, it became clear that browsing was by far the most important methodology adopted. For example, over a number of hours of direct observation of library patrons in the Hobart Lending Library, virtually all patrons went directly to the shelves rather than going to the OPAC terminal. at the shelves, they first went to recent returns, then wandered along the shelves (frequently by genre, but sometimes generally in the alphabetical section), looking at book covers and publisher blurbs to finally select the items they wished to borrow. This pattern was repeated in the non-fiction area, although the users browsed a narrower shelf (Dewey number) range. The analysis of user behaviour in the physical library produced two clear objectives for the new OPAC. First, it must enable the user to browse a large collection easily and effectively. Second, it must provide additional contextual information to help the user make a selection. This contextual information could be recent returns data, book covers, item descriptions, reviews, etc.

It should be noted here that the ability to do direct input or known-item searching (eg, by typing an author/title/keyword in an empty search box) was a given for the new OPAC. However, such functionality was not to be the only or main search methodology provided. An earlier internal analysis of user search behaviour across two major websites provided by the State Library of Tasmania had also confirmed the prevalence of browsing as a discovery technique employed by general web users.

Providing Online Browsing Capability

The ability to readily browse online requires the capacity to provide results based on simple choices made by the user, with the secondary ability to continue to display further choices (refinements) as the user navigates through the system. In other words, the user is navigating by responding to options or facets provided by the system, both to initiate and refine searches. This is distinct from having the user create search statements and refinements by articulating them themselves and typing them in.

On the OPAC, facets are content subsets that are provided by the system right from the home page, and can be chosen in any order or sequence, with the choices offered to the user always being a subset of the facets chosen to that

point. Every facet choice provides a set of results, which can be refined further. In this process the user should also be able remove individual facets from a given results set and choose new options accordingly. The Verity K2 software chosen by the State Library can deliver this type of faceted browsing. The issue for the State Library became one of choosing the appropriate facets and identifying and sourcing the data to drive those facets.

An internal working party was established to consider the possible facets to be provided on the new OPAC, and the decision was subsequently taken to provide format (e.g. images, video, audio), popular genres (e.g. romance, biography), and audience groups (e.g. youth, children, researchers) as facets on the home page. Two other possible facets – location and topic – were excluded from home page provision, but will be included as refining options on results pages. Figure 1 provides a screen shot of the internal OPAC prototype home page as first developed; Figure 2 provides a screen shot of the internal OPAC home page prototype at the time of writing and as modified by usability testing to date.

Figure 1
Initial OPAC Replacement Prototype – Home Page

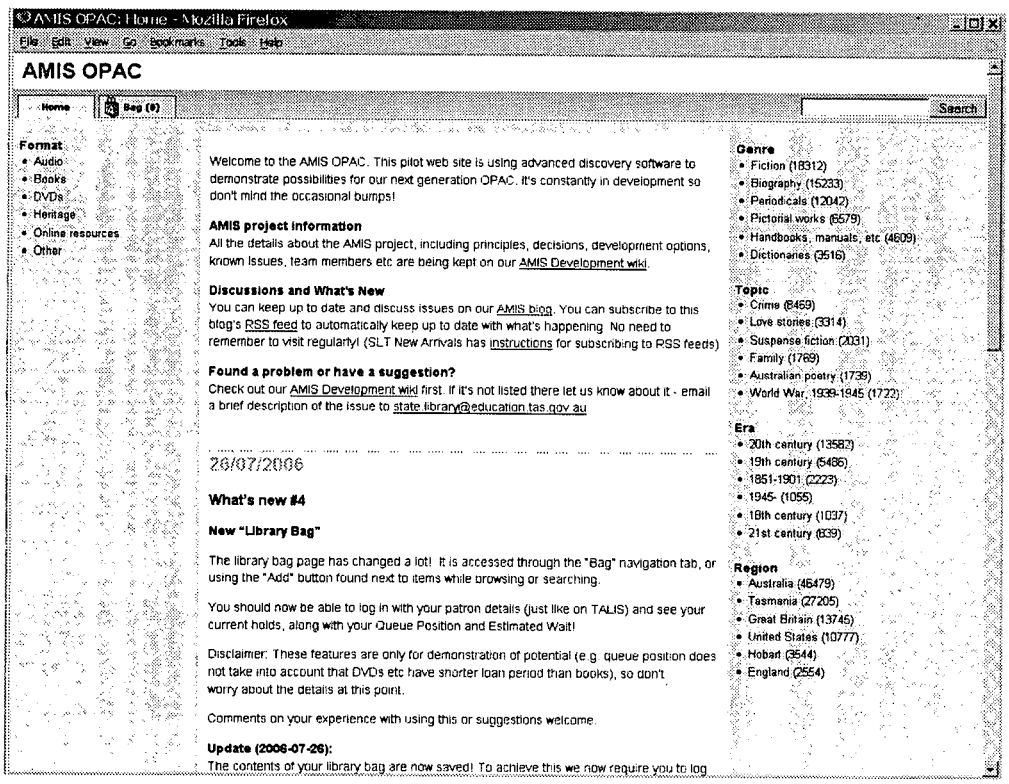
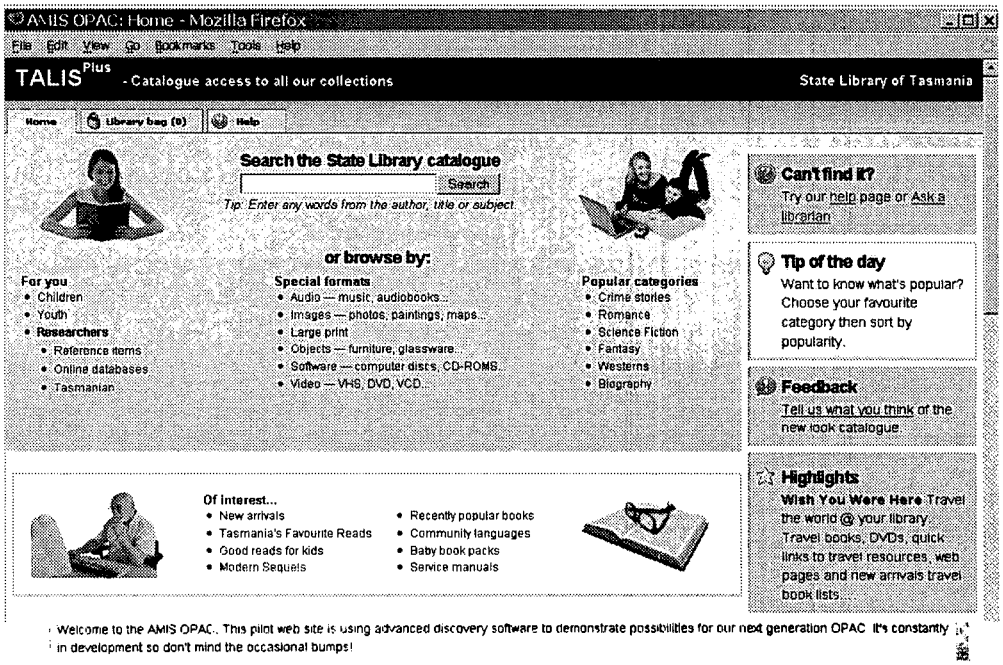


Figure 2
Updated OPAC Prototype Home Page after Two Rounds of Usability Testing



On the results pages, the facets of audience, format, region, availability (e.g. on shelf, requestable, reference, online), subject, genre, series, author, and library location would be provided. Figure 3 provides a screen shot of the internal OPAC prototype results page as first developed; Figure 4 provides a screen shot of the internal OPAC prototype results page at the time of writing and as modified by usability testing to date.

These decisions could only be based on an incomplete analysis of current OPAC search patterns, as the current OPAC provides no browse options to measure against. Although it was recognised that usability testing would be undertaken before the OPAC went live, it was also recognised that this usability testing could only reflect options initially created by the library. It would not be possible to undertake a full-scale research project giving users an absolutely clear starting point where they were then asked to articulate their information needs and somehow demonstrate searching behaviours. The State Library was, however, able to use experience gained over six years in providing two major Tasmanian websites, *Tasmania Online*[11] and *Service Tasmania Online*[12], which both provide and utilise facets.

Data Availability

The facets chosen to deliver on the OPAC were those that the State Library felt would be most valuable to the user. Work then began on determining whether the relevant data existed within the bibliographic database to provide those facets, or whether it could be created or translated from existing data. This process brought into sharp contrast the tension between the bibliographic data created to manage physical collections and the data actually needed by the new OPAC to enable simple user-oriented discovery. Much of the MARC record and bibliographic system is geared to meeting the needs of acquisitions, unique title/edition identification, and internal collection management and use. The provision of browsable facets introduces a new demand on this data with two distinct questions.

Figure 3
Initial OPAC Replacement Prototype – Results Page

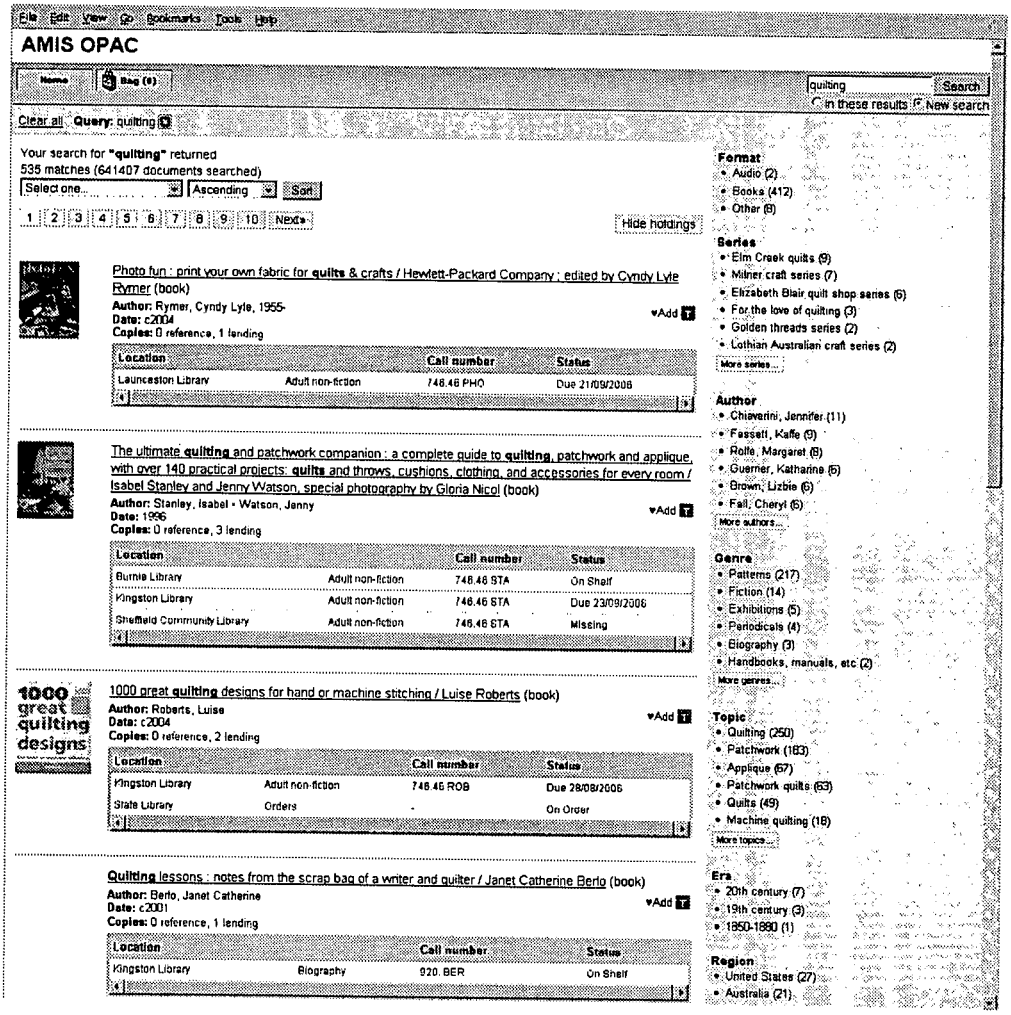
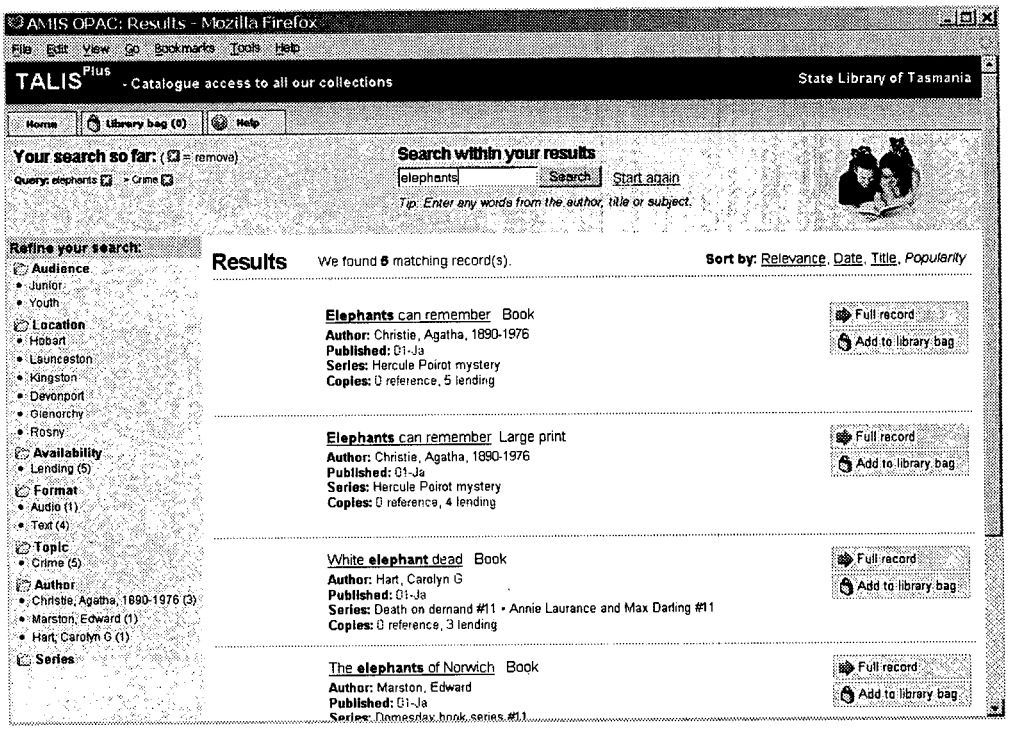


Figure 4
Initial OPAC Replacement Prototype – Results Page after Two Rounds of Usability Testing
(Book Jacket Display Temporarily Disabled)



First, does the required facet data exist? For example, for the provision of a temporal or chronology-based facet (an 'era' facet), the bibliographic record had to contain information describing the chronological coverage of the item. This information was found as chronological subdivisions in Library of Congress (LC) subject headings. Consequently it was a fairly straightforward process to extract those subdivisions and transfer them to the new OPAC. For another possible facet, that of audience, the data is harder to find.

The second, and equally important, problem was: does this data make sense when transposed into the new environment? Using the above 'era' facet example, chronological options readily appear in the new OPAC, with sub-options such as '20th century', '1900-', '1945-', etc. But it then became clear that these sub-options often overlapped, and offered our users a confusing rather than a clear choice (eg, when both '20th century' and '1900- 'appeared as options under the 'era' facet). This had not been a problem within LC subject headings, because the chronological subdivisions always appear in context.

Other data problems were found, as, in some cases, the use of certain data fields in the bibliographic database was inconsistent, and the use of facets now made this inconsistency obvious to the user. For example, fiction subject headings for genre had only been added since a certain date to the bibliographic database; consequently, a large number of resources do not appear under that facet when it is presented in the new OPAC. This, in turn, reduces the value of that facet to the end-user, possibly to the point where that facet cannot be provided.

Most of these problems can be resolved through data manipulation when importing the data, or through cleanups of the existing data within the ILMS. This workload raises the question as to whether it would make more sense in the future to alter the nature of cataloguing activity, so as to focus on these discovery data needs. One can argue that these needs are more important than creating extensive bibliographic records where the data may only be rarely needed, especially where that bibliographic detail is available elsewhere.

The other key data needed by the new State Library OPAC was shelf location and copy availability. The new OPAC has to provide the functionality to directly query, at the time of need, the circulation record of the items found by the user. This process, performed using real-time queries directly into the circulation database, has proven to be quite simple, with users offered the option to display holdings information on a page-by-page basis when viewing results. The performance demands of this type of access have not been a problem.

Facet Operation

Another aspect of facet functionality to be resolved in the OPAC redesign process is the actual operation of the facets themselves. Should facet options be provided in alphabetical order, or in an order based on number of occurrences? Should they be provided in small groups, or in longer lists? How should the user expand or move through possibilities within one facet? How should hierarchies within a facet be organised and displayed? Some questions may be answered through usability testing later in the project, but it is probable that the new OPAC will go into production with some issues still unresolved.

It is against this context that the vision of the first OPAC release being a development or version 1 release is important. The State Library will in many ways only really begin the true design process as it starts to monitor and enhance or alter the operation of the OPAC over time, as real-life reactions and user behaviours emerge. Although it would be ideal to have really firm behaviours to work with from the beginning, this 'develop on the fly' approach is, in fact, very common on the web, where many successful search services have been in beta mode for extended periods.

Enhancing Results Provision on the New OPAC

Once a working framework for facet provision existed, the design process moved onto the next discovery stage – that of providing a results list so that the user could select an actual item for further investigation or access. This is an area where library OPACs have traditionally provided simple results listings, often an alphabetical list, but sometimes a list ordered according to when items were added to the database. Library OPACs have fallen behind the results outcomes delivered on the web, where relevance ranking is the accepted norm. This can best be highlighted by noting that Google is not so much a superlative search engine as a superlative ranking engine.

That users tend to choose the first item in a results list has been recognised and described by Steve Krug, using the concept developed by Herbert Simon, as 'satisficing'.[13] For libraries 'satisficing' behaviour holds two key messages: users do not read a list logically from beginning to end and then choose the best option, and, users are likely to be satisfied by the first item they see. Users do not necessarily evaluate a range of competing resources to find the very best match. This behaviour has been consistently evident when the State Library has analysed the searching behaviour on a number of its websites.[14]

This behaviour means that the new OPAC must provide good relevance ranking, where the user must be able to find something they think may satisfy their query in the first page of results. For this reason, the State Library decided to provide relevance ranking as the default results display option on the new OPAC, with date, title, and other sort options as alternative choices. At this point in time, investigations are still under way to determine whether word occurrence ranking or item popularity (as determined by the number of copies/ reservations/loans) will produce the best relevance ranking for normal user enquiries.

Another key aspect to providing an effective results display geared to user behaviours is the reduction of unnecessary complications or choices. The unnecessary duplication of items in a results list is a major issue for public libraries, as most users are rarely, if ever, likely to care about the edition or minor format variation of the item they are seeking. The development of Functional Requirements for Bibliographic Records (FRBR)[15] recognises this need and moves to overcome the balkanisation of bibliographic records are underway. The prototypes so far developed, however, demonstrate that it is hard to deliver consistently good results.[16] For some time the State Library has amalgamated copies of similar fiction title editions onto the same bibliographic record. This pre-catalogue 'FRBR-isation', although horrifying to cataloguing purists, is now paying dividends on the new OPAC in terms of simpler results pages.

The final component of tailoring the OPAC to user behaviour is providing the additional information so often helpful to the user when selecting an item for further investigation. In this context, the presentation of the book jacket

where available is considered essential, as is the provision of other supplementary information such as a book synopsis, reviews, recommendations, ratings, and popularity. This is an area where the traditional ILMS bibliographic record provides very little assistance. In the emerging web world of loosely coupled services and mashups, however, it becomes possible to link to other services that provide such data. For example, the OPAC can now bring back data from the web (such as book jackets from Amazon) into the results display process effectively and in real time.

Whether all of these services should be sourced from sources external to the library and the library world is an issue that begs to be addressed. Ultimately it would be better if libraries could create and share this type of data amongst themselves, and thereby provide a commercial-free source of evaluative data and information. It would be easy to argue that this type of data sharing and reuse among libraries would be more valuable in the web world that the recurrent sharing of unnecessary bibliographic data.

The operation of the new OPAC when delivering results will also take into account the operational support standards and examples set by successful external web search services. These include the provision of functionality to ensure that users do not fail, but are always given options, even if they take the user to other possible solution providers. The OPAC will also acknowledge that users (and authors) may not necessarily spell or articulate words and needs correctly, and provide spell-checking, stemming, automatic 'anding', in context assistance, etc.

Developing a Gateway Searching Service: OPAC 2.0

The ability to use non-ILMS software to import and index the data contained in the library database also introduces the capacity to include material extracted from other sources as well the library database. This expansion provides access functionality that goes beyond the resources described in the library database and can extend to the full range of digital resources that can be delivered to the desktop. As such it raises additional design questions concerning interface functionality and operation. In this context the gateway may well be moving beyond the traditional OPAC and may be seen as entering into what could reasonably be called OPAC 2.0 design.

The gateway service to be provided by the State Library of Tasmania aims to bring together resources from different locations, and often of different types, to allow a single search interface. Although federated search services have been developed to provide this type of functionality, this type of solution was seen as undesirable to the State Library because of the performance network latency issues involved in distributing searches and collating search results. It was also recognised that there were inherent difficulties in amalgamating result sets that came from different indexing systems.

The State Library has, instead, decided to use an alternative architecture, where data held in other affiliated systems is imported into a single datastore, which is then indexed by the search and retrieval software as a whole. Rather than use Z39.50 to access the bibliographic database, this data too can be exported and re-indexed easily by an aggregated metadata indexing system.

Standards such as those developed and promoted by the Open Archives Initiative have been developed for exporting and harvesting metadata.[17] In the State Library's case, the initial gateway will include only datastores managed internally. Future developments, however, will move beyond such resources and aim to include content from other affiliated or complying sources of Tasmanian information.

The State Library of Tasmania is not as far advanced in this process as it is with the OPAC, which is scheduled to live in the first few months of 2007, whereas the gateway will not go live for a number of months after that. A major concern with the implementation of the gateway (which as yet has no formal name) is that gateway searching behaviours are not yet fully understood. It is unlikely that it will be appropriate to simply provide the same discovery and functional elements on the gateway as provided for the OPAC replacement.

From a data aspect, it is obvious to the State Library that there must be a minimal set of shareable data common to all the composite datastores to allow effective searching to occur on the gateway. Identifying this shareable dataset is under way, but it is clear that this will be a small dataset, which in turn will reduce the variety and type of facets that can be sensibly provided.

Although the provision of a simple home page with browsable facets is seen as necessary, it is unclear to us at this stage as to how browsing would work in a gateway that incorporates results from a variety of locations. This is linked to the issue of results handling and relevance ranking. For example, how does one measure the likely relevance to the user of a digital image of Mount Wellington against a book about Mount Wellington. In this case, the intent of the user and the query is likely to be very important. This is where the initial entry screen may well attempt to elucidate that intent, by offering format and availability options before offering further searching or refinement options. Relevance ranking may then be delivered as appropriate to the format or intent. The range of facets provided on subsequent results pages may also vary according to initial intent.

The State Library has little experience in providing a consolidated gateway, but will be able to build on the knowledge of user behaviour gained by having a facet-driven OPAC in operation before the gateway is delivered. We will also conduct a range of usability tests and user consultations before a first version of the gateway is released. Our key design principles will be to keep the gateway simple and to follow a process of continual development and refinement that continues after the first release.

Conclusion

This article has described the processes undertaken by the State Library of Tasmania so far in developing a new OPAC. It must be acknowledged, however, that creating a new OPAC is very much a process of discovery, where the combination of having the right data and using it effectively with new software is uncharted territory. Despite this uncertainty, it is clear that the data issues that have emerged will be resolved by more than just 'making our data work harder'[18], and may require libraries to review what data we actually need to create and share.

For many years we have assumed our bibliographic databases contain the discovery data required by our users, but this is an assumption that is questionable once the OPAC operates on different software. It may be too early to answer fully the question about just what discovery data is needed. This will emerge only through real-life analysis of user behaviour on new OPACs. It is, however, certainly time to question whether all the current information we create and maintain on the ILMS is relevant.[19] In time libraries may be in a position to identify the new types of data needed by users and to then develop processes to create or access that data – and to cease creating unneeded data.

The integration of a new OPAC in the web environment also raises the question of the data that libraries share. Collectively, with so many libraries professionally evaluating information content and possessing so much information about item usage, we must have one of the best recommendation datastores possible. If only we could share and access that data, we could deliver the type of advisory, recommendation, and supplemental information that is now expected by our users to augment bibliographic data. This type of data sharing may be more important in the long run, in terms of keeping our services relevant, than any amount of sharing of bibliographic data.

Notes

1 This article is a revised version of a paper given at the Australian Committee on Cataloguing Seminar 'Beyond the OPAC: Future Directions for Web-Based Catalogues' Perth WA 18 September 2006

2 S Abram *The Kids are Alright: Millennials and their Information Behaviors* Powerpoint presentation at SirsiDynix 2006 www.sirsi.com/Resources/abram_articles.html [accessed 8 Sept 2006]

3 For example, see M Stephens *Tame the Web: Libraries and Technology* weblog, at tametheweb.com/ [accessed 8 September 2006]

4 *Ibid*

5 R Tennant 'Lipstick on a Pig' *Library Journal* 15 April 2005, at libraryjournal.com/article/CA516027.html?display=Digital+LibrariesNews&ind [accessed 8 September 2006]

6 The recent OPAC alternative provided by North Carolina State University is an important milestone: NC State University *NCSU Libraries* OPAC, at

www2.lib.ncsu.edu/catalog/ [accessed 8 September 2006]

7 For more about Autonomy's products, see their website, at www.autonomy.com/content/home/ [accessed 8 September 2006]

8 K G Schneider 'How OPACs Suck, Part 1: Relevance Rank (Or the Lack of It)' *ALA Techsource* Weblog 3 April 2006, at www.techsource.ala.org/blog/2006/03/how-opacs-suck-part-1-relevance-rank-or-the-lack-of-it.html [accessed 8 September 2006]

9 J Macaulay 'OPAC Blog Posts: A List' *Life as I know it* weblog 5 July 2006, at scruffynerf.wordpress.com/2006/07/05/opac-blog-posts-a-list/ [accessed 8 September 2006]

10 C L Borgman 'Why are Online Catalogues Still Hard to Use?' *Journal of the American Society for Information Science* vol 47 no 7 1996 pp493-503

11 State Library of Tasmania *Tasmania Online* website, at www.tas.gov.au/tasmaniaonline [accessed 8 September 2006]

12 State Library of Tasmania *Service Tasmania Online* website, at www.service.tas.gov.au [accessed 8 September 2006]

13 S Krug *Don't Make Me Think: A Common Sense Approach to Web Usability* 2nd ed New Riders Press 2005

14 Research on *Tasmania Online* and *Service Tasmania Online* has led the State Library to move key resources up in results lists to ensure they are in fact used. When key results appeared down in the lists they were ignored by users in preference to items found earlier in the list

15 W Denton *The FRBR Blog* Weblog, at www.frbr.org/ [accessed 8 September 2006]

16 S Weibel 'Dangerous Waters' *Weibel Lines: Ruminations on libraries and Internet Standards* weblog 26 Aug 2006, at weibel-lines.typepad.com/weibelines/2006/08/dangerous_water.html [accessed 8 September 2006]

17 Open Archives Initiative *Open Archives Initiative* website, at www.openarchives.org/ [accessed 8 September 2006]

18 L Dempsey 'Making Data Work: Web 2.0 and Catalogs' *Lorcan Dempsey's Weblog: On Libraries, Services and Networks* 4 Oct 2005, at orweblog.oclc.org/archives/000815.html [accessed 8 September 2006]

19 University of California Libraries Bibliographic Services Task Force *Rethinking how we provide Bibliographic Services for the University of California* 2005, at libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf#search=%22rethinking%20how%20we%20provide%20bibliographic%22 [accessed 8 September 2006]