

Beyond the grave: where to with gen (wh)Y?

Michael Gonzalez
Coordinator Digital Discoveries and Services
University of Western Sydney Library
m.gonzalez@uws.edu.au

Abstract

Through the analysis of the University of Western Sydney's (UWS) federated searching tool usage statistics, the paper aims to highlight potential issues for Libraries choosing to make federated searching, the preferred method for finding scholarly resources. The paper will reflect on the UWS experience and create an awareness of potential issues for future implementations of federated search products.

Introduction

Since the 2005 OCLC *Perceptions* study (Online Computer Library Center 2004), there has been an active push across the Library community to create Google-like interfaces, increasingly known as discovery layers, spanning a range of offerings. Recognising changes in the way 'gen Y' students seek information, the University of Western Sydney (UWS) Library embarked on the implementation of a discovery tool to facilitate discovery of quality, scholarly information. In February 2009 the Library launched its new finding aid, merging the classic catalogue with a federated search tool which offered more than the traditional OPAC. After much conceptualisation the new system, generically known as the Library Search Box; was launched as a 'world first': a purpose-built tool facilitating seamless, intuitive searching of the Library's collections, in house print and remotely accessed electronic materials.

The Library Search Box assumed a prominent position on the Library's WebPages, and thus gained the honour of becoming the primary search facility for students seeking scholarly information. Its popularity amongst clients was unparalleled and the information gathered from user feedback, complemented by usage statistics, has led to a greater understanding of student information-seeking behaviour. With a great wealth of analysable information available, it became important to revisit the initial rationale underlying the implementation of the discovery layer and examine whether the findings of studies such as the OCLC *Perceptions* report were reflected by our own users' experiences.

This paper will examine data drawn from user search logs and feedback and draw comparisons between various user studies and the University of Western Sydney Library experience. It will also outline changes undertaken by the Library as a direct result of this analysis and discuss the effectiveness of changes made in enhancing access to resources and services for students and staff. The analysis and subsequent refinements undertaken highlight the impact of the introduction of a discovery layer on Library operations and traditional Library practices.

The Library Search Box launched in February 2009 is not the Library Search Box that exists today. Over the period of one year it has changed and adapted to better meet the needs of our primary clientele: UWS staff and students. Statistical analyses, which became a daily routine, drove the evolution of the system, and insights gained into client information-seeking behaviour meant that a mind shift from within the Library was required to facilitate the many changes, however radical, which ensued.

UWS Library recognises that the current system is not perfect. Data gathered highlights issues needing to be addressed by next and future generations of discovery layer tools. In a world where students may no longer see a direct need for the Library or Librarians, it is critical that we make available tools which are intuitive, useful, relevant, and which directly support and enhance scholarly research. The UWS Library Search Box is the first step along this road.

Background

Important to the understanding of the decision reached by UWS Library to implement a discovery layer is an awareness of the structure under which the Library operates. UWS Library comprises seven centrally administered Libraries spanning a large geographical area of Greater Western Sydney. UWS students are not tied to one campus and may travel between campuses to complete their studies. Student assistance is also provided through a centralised virtual reference service, Information Central (Pavincich 2007). The Library provides the student community with comprehensive online tutorials (Successful Searching) and associated quizzes, a direct result of the decision in late 2007 to significantly refine face-to-face information literacy delivery.

In recognition of clients' unique study patterns, the Library has adopted an electronic preferred collection policy, complementing a strong culture of electronic resource promotion, as evidenced through the Library's resource usage surveys and awareness building programs (Boness 2009).

In the second half of 2008, the Library launched 'Search Central' (Serials Solutions' 360 Search), an initial federated search tool designed to search a range of electronic resources as well as Library holdings. Search Central was widely promoted as a separate tool to the catalogue, and whilst it led to an increase in electronic resource usage (as evidenced through link resolver statistics), it did not contribute to any significant decrease in searching of the Library catalogue. This meant that there existed the potential for having students visit three areas of the webpage for scholarly resources: database platforms, OPAC and Search Central, ultimately adding to issues of "library anxiety" (Onwuegbuzie, Jiao & Bostick 2004, p. 251).

Drawing on experiences outlined above and in response to the 2005 OCLC *Perceptions* study, the Library began its investigations into the potential implementation of a more user-friendly discovery layer. During the investigations of a range of products, it became evident that the needs of the Library, whilst simple, were not easily achievable. All products reviewed lacked the ability to quickly and simultaneously search the catalogue as well as electronic resources. This fundamental requirement led to the conclusion that a solution would need to be purpose-built, integrating the best of the existing catalogue and the Search Central service, whilst delivering a range of user friendly features.

The Library chose to combine its existing federated search tool, Search Central, with a feature-rich discovery layer product (AquaBrowser). The result was a search interface that simultaneously searched Library holdings as well as a select group of electronic resources. This tool became known as the Library Search Box.

The Library Search Box

In the process of configuring the Library search box, certain informed assumptions were made in relation to what our clients needed from a catalogue. The goal of the implementation team was to emulate a 'Google'-like search which would produce 'Google' type results. These characteristics were drawn mainly from the *Perceptions* study, additionally including elements identified in other work, particularly that of Jansen & Pooch (2001).

The key assumptions made at the beginning of the project included:

- Clients wanted to be able to search the majority of our resources simultaneously and in a timely manner;
- The introduction of the Library Search Box would largely remove the demand for access to native interfaces;
- There was a need to be able to provide natural language searching;
- Clients wanted to enter entire questions and retrieve results;
- Clients wanted to use to use interactive technologies (Web 2.0)

Early in 2009, after six months of development, the Library Search Box was launched. Without fanfare, its appearance coinciding with a website redesign, a catalogue upgrade and the student semester break, the Search Box took a prominent position on the university homepage as a single user-friendly and intuitive search box that amalgamated a range of existing offerings.

Feedback

When launched, client feedback was sought. Resulting information was invaluable in determining what the implementation team had done correctly and what improvement or refinement was needed. Initially feedback received was largely negative and based on the fact that many of the traditional links had been removed and placed at the second level of the webpage. These included links to our electronic resources and Search Central. Comments included "Where on the first page is the direct link to the UWS catalogue? I can't see it", "e-resources link should be on home page" and "It is not immediately obvious how to access voyager". Quickly realising that the Library search box was not a one-size-fits-all solution, a link to the electronic resources page was quickly reinstated, albeit not as prominently located as in the past. The Search Central link, however, remained as a link within the Library Search Box.

In direct relation to the Library Search Box, feedback showed a general acceptance; however, things not obvious to the implementation team were quickly highlighted in the feedback received. The initial inability of the Library Search Box to facilitate phrase searching was identified in comments indicating that the software could not "configure the search for title" and "I can't find anything". This led to further testing resulting in phrase searching being placed at the top of the relevance matrix. Adjustment to relevance required daily analysis, and balancing of the many elements

of the search to produce the best results became a priority, particularly when the team examined the statistics relating to page views.

Positive client feedback included views that “tags [make] it easier to search for related subjects”, the “icons showing whether the result is a book, electronic resource” were useful and “its positively breathtaking”. The team had initially assumed users wanted to use Web 2.0 technologies; however, the level at which aspects of content enrichment, user feedback and user configuration were adopted was surprising and highlights the importance of customised experiences to the “customization generation” (Solomon & Schrum 2007, p. 33). Within the first month, over 400 students had created user accounts and contributed to the system.

The most validating comment received was on the back of an initial complaint. Two months after the launch, a student followed up on a complaint made in the initial stages, stating: “I’ve turned up some works I would probably never have found using an old-style search on the classic catalogue”. Whilst the classic catalogue allowed the student to search for known items, the strength of the Library Search Box lay in the ability to retrieve items previously too difficult to find.

As indicated above, at launch, the feedback received from Library clients was primarily negative, accounting for 70% of the total feedback. An analysis of the data showed that of the negative feedback received, only 20% was directly related to the Library Search Box, with the majority of the remaining feedback being related to the changes to the Library webpage.

Similarly, staff feedback was also mixed. Beta versions of the Library Search Box were introduced to staff in mid-2008, as a forthcoming enhancement. Staff were encouraged to research and trial different implementations of the product in making their assessments of what the tool could do. As the launch of the Library Search box was timed with an LMS upgrade and subsequent OPAC downtime, the only way for staff to access holdings information was through the new interface. This meant staff, of necessity, needed to familiarise themselves with the new interface.

Feedback received from staff included “I can’t find anything”, “what it doesn’t phrase search!” and “Why can’t I search for a call number”. All feedback not directly related to a functionality issue was met with a similar response: essentially a request for understanding that the Library search box was about the student rather than staff requirements. Similarly, suggested improvements included “a link to the old catalogue”, the development of the advanced search module and the extended development of the help menus, a feature that the implementation team consciously had chosen not to develop, as statistically such features have recorded low use in the traditional OPAC.

The positive staff feedback included comment on the ability to refine results, the speed at which results were retrieved and the ability to extend searches by accessing the article results portion of the results page. An interesting aspect of the positive feedback received was that the majority came from more experienced Librarians, who particularly commented on the natural language searching and ability to export results.

All the feedback received demonstrated the unique challenges faced when developing a tool such as the Library Search Box. Traditionally, Libraries have not

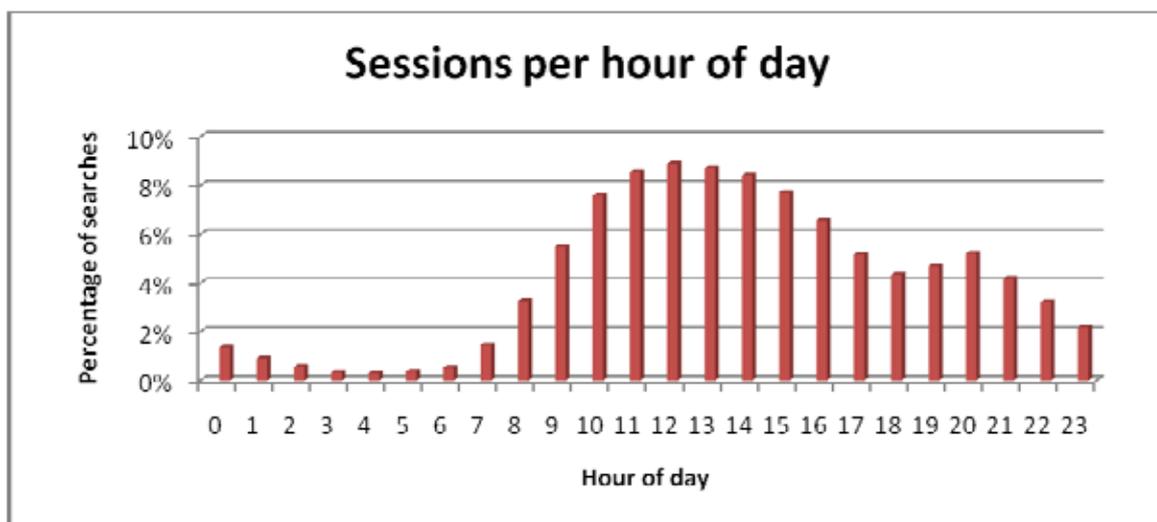
needed to configure the complexities of search engines, they have not had to examine relevance, phrase searching or field searching and they have not needed an understanding of natural language searching. These items were often found in advanced search screen of our OPACs or through the use of Boolean Operators. A study by Lau & Goh 2006 expressed the need for Libraries to work with clients when configuring OPAC type systems, and the successful implementation of the Library Search Box was seeded in the ability to respond quickly and make proactive and in some cases reactive changes to best suit client needs. All feedback was addressed quickly and all suggestions for improvement were given serious consideration.

Search Box Usage

Amongst the advantages of the search box was the statistical information it generated. The information was used to both validate and make defunct some of the assumptions the implementation team had originally made. The ease in which the system generated and presented statistical information meant that the implementation team was able to analyse data on a daily basis and make changes accordingly. Each morning the previous days data was analysed, this included an examination of the search logs for popular and orphaned query strings, items that received a high hit counts, usage patterns of the facets as well hourly usage data.

The bulk of the search box usage was between the hours of 10am and 6pm and accounted for 61.5% of total sessions for the year. Significantly, usage was recorded at all times of the day with the lowest usage period predictably being at 4am, accounting for only 0.3% of total use. Usage spikes found between the hours of 7pm and 10pm accounted for 17.3% of total usage (Figure 1). This data helpfully validates the Library opening hours, particularly those of the virtual reference service, the hours of which mirror peak usage patterns of the Library Search Box.

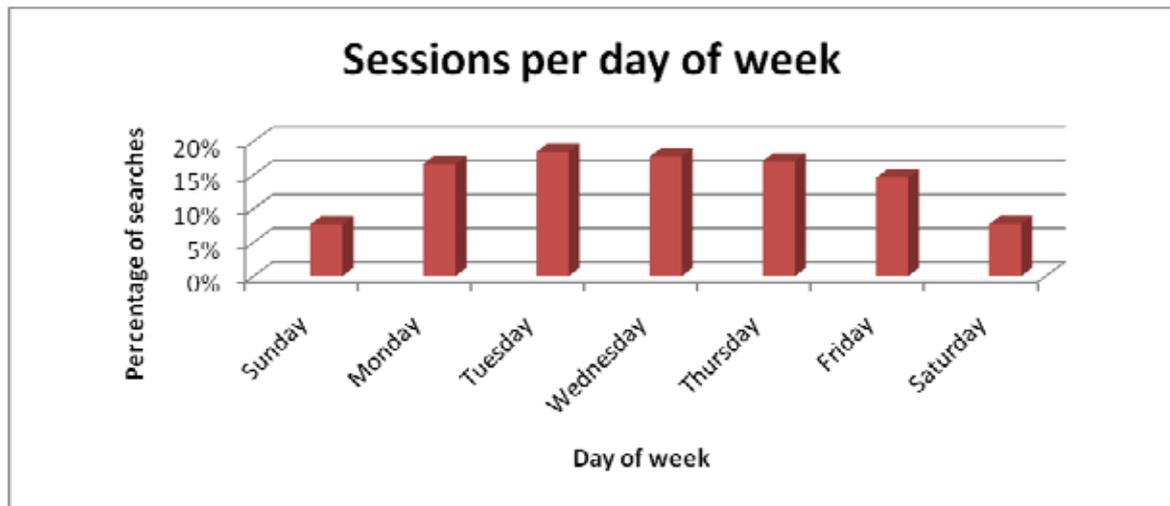
Figure 1



What was demonstrated through analysis of daily usage was that 15% of the weekly usage occurred on weekends, with weekday usage patterns matching attendance patterns of UWS students, with the busiest day being Tuesday. Interestingly, usage patterns did not change during the student semester break periods. Although

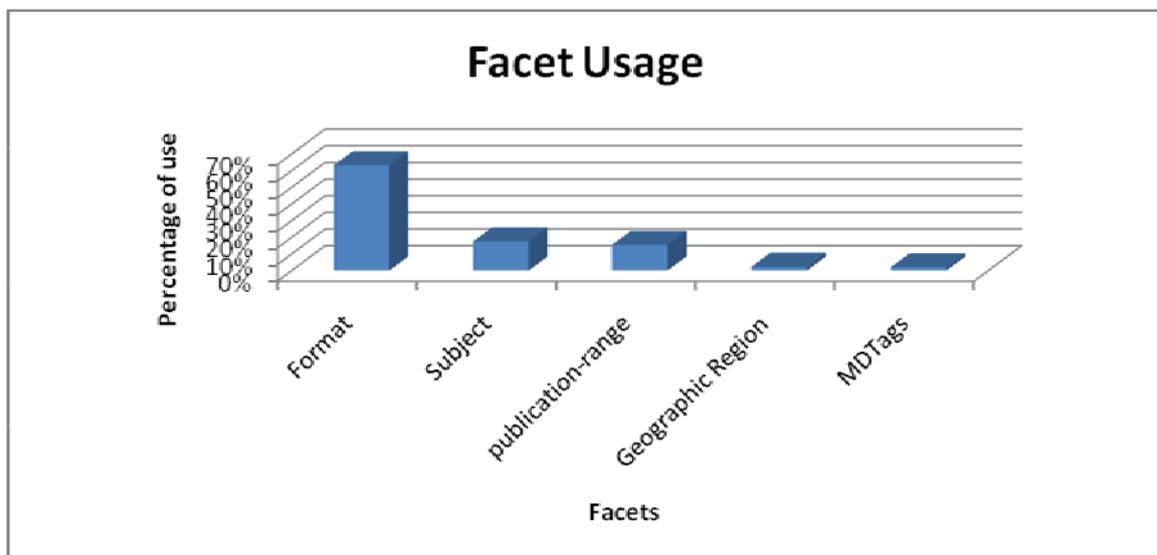
searching was considerably lower, the daily use patterns matched semester patterns with similar proportional use recorded on weekends.

Figure 2



Much time was spent by the implementation team in designing the facet lists that would provide useful refinement options for Library clients. These facet lists were designed with users in mind, drawing on reference service scenarios to determine how facets would be worded and presented. The main facet used by students was the format option (book, journal etc); this facet accounting for 63% of the total usage. Interestingly, usage of the user tags (MD Tags) accounted for only 1.9% of total usage and was consistently the least used refine option (Figure 3).

Figure 3

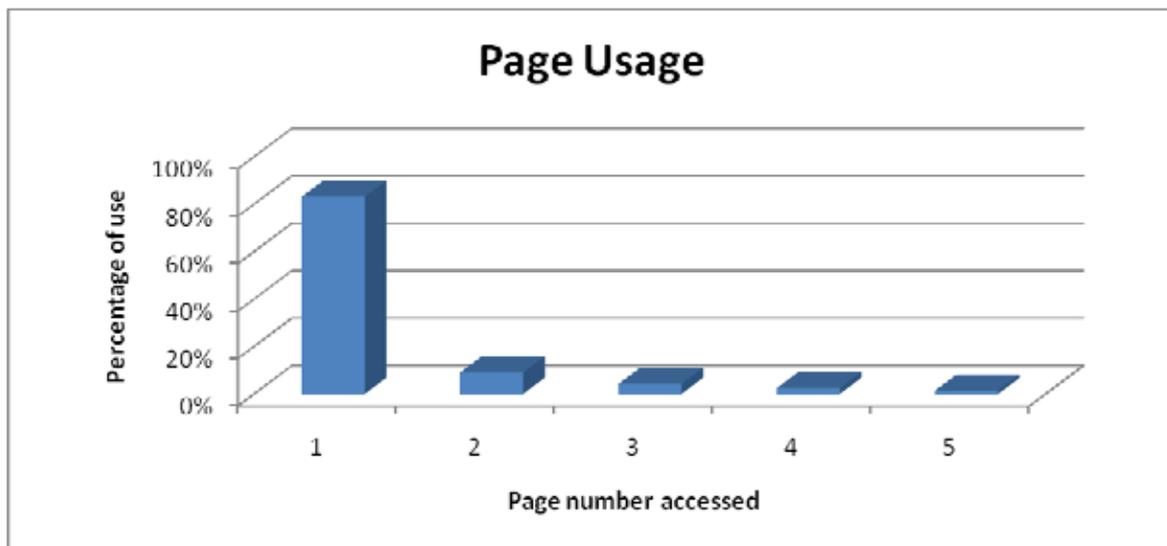


Dominance of the format facet demonstrated an obvious need of our clients to granulate the item types being presented. This led to the facets being redefined in mid-2009, and also accommodated a new facet (eBooks), adding greater specificity to those existing. This process highlighted the importance of consistent and accurate holdings data.

A further interesting aspect demonstrated through usage statistics was use of the help menu. As previously stated, the Library had consciously decided not to develop the module. This decision was validated, with statistics showing that the help menu was accessed only once every 840 sessions.

Paralleling low use of the help menus was the fact that 80% of clients did not venture beyond the first results page, with 95% of all clients not venturing beyond the fourth page (figure 4). This aspect of client information-seeking behaviour was also identified in a study on user behaviour that showed a drop in interest beyond the first page (Granka, Thorsten & Geri 2008); however, the Granka study was based on internet searching rather than traditional OPAC searching. What was deduced showed that Library Search Box users were not displaying OPAC searching behaviour; rather they were using the Library Search Box in the same way they would undertake an internet search. The change in expected searching behaviour was noticed early in the implementation, and changes such as automatic phrase searching and adjustments to the short results layout were subsequently made in order to make the first page of results more meaningful.

Figure 4



Statistics shown in figure 4 highlight a potential issue with the current system. Although students were conducting a simultaneous search of both the catalogue and a select group of electronic resources, they were still required to click through to see journal article results and then required to again click through to see the article itself using the Library's link resolver. Analysis of link resolver usage was undertaken, demonstrating an increase of only 8.19% in use over usage recorded in 2008. Significantly, usage in 2008 had recorded an increase of 63% from the 2007 period (attributable mainly to the introduction of Search Central). From this information came the deduction that the federated journal search portion of the Library Search Box was being used; however, it did not lead to further increases as originally anticipated.

The most interesting data produced by Library Search Box data analysis were the user search logs. These user search logs, from time of implementation, provided robust indications of what Library clients were searching for, how they were searching and, more importantly, what they were not finding. Among the

assumptions that had been made by the implementation team was that Library clients wanted to enter longer statements, incorporating natural language terms. However, this was quickly found to be incorrect: reflecting findings of the study conducted by Jansen & Pooch (2001), Library Search Box users tended to use shorter queries to retrieve results, recording an average of 3.5 searches per session. Log analysis showed that users typically started with a single word, examined a page of results, then entered another word to narrow the search. On average, users would extend their initial search three times before either entering a new search term or using one of the facets (facet usage was recorded once every four queries). The shorter queries also meant that some of the top searches were terms as simple as “book” or “journal”.

The log of queries which retrieved little to no results showed that the Library Search Box was almost immediately treated as a generic search. Of the top 50 orphaned searches, 26% of searches were not resource-based, with a high proportion of the searches related to citation assistance. Such revelations were used by UWS to make changes on the homepage in attempting to make such information more intuitively accessible. User feedback was also incorporated into the results page of the Library Search Box; as a feature designed to direct students to an appropriate service/resource directly related to their initial query. If a client searched for “Library Hours” for example, they would receive a message in their results directing them to an appropriate page on the Library website.

Generic use of the search box was used to reinforce the commitment on behalf of the Library to provide better access to its online Librarian service, which in late 2009 took a prominent position on the front page, in the form of a widget. This enabled students to enter those generic queries they were previously entering into the search box and which often yielded no results. Analysis of the types of queries received by the virtual service showed no change in the proportion of information queries received, which remained at 17% of total queries received in the 2008 and 2009 periods.

Wider Observations

The implementation of the Library Search Box had implications in the area of resource usage measurement. Previous to implementation of the Library Search Box, resource value was measured in terms of either cost per download or search. Historically, usage has climbed with the introduction of federated searches, as evidenced in studies conducted by Newton & Silberger (2007) and Boyd et al (2006). This trend was paralleled when UWS initially introduced federated searching (Search Central) in 2008. Unsurprisingly, the spike in usage was also seen with the introduction of the Library Search Box; however, an alarming rise of over 2000% in usage of some previously unpopular resources led to concern and further investigation. A study conducted by Kinman (2009) alluded to possible issues with current statistical measures, almost mirroring UWS’ experience. The issue was identified in the configuration of the current standard (COUNTER) for recording web based statistics, particularly in relation to the way short records are retrieved. Federated searching increased usage because shorter record retrievals were counted as full text downloads, leading to every search in the Library Search Box

recording downloads from a number of providers. The wider implication of identifying such an error is found in the assessments many Libraries have already made in regard to the utility of a federated search layer as well as subscriptions. Subscriptions and purchase plans of the future are seeded in assessments made on existing resources, and using cost per download in the current federated search environment has the potential to negatively and erroneously impact upon decision making processes. As the technology behind the Library Search Box relied on the use of the Library's link resolver (360 link), increases in usage were measured by examining its usage patterns. Although useful in determining how much usage increased, it was not exact as it did not allow the Library to specifically target usage recorded through the Library Search Box.

The introduction of federated search interfaces also brings into question the value of the index. One of the key findings in a study conducted by the OCLC into online catalogues (Online Computer Library Center 2008) showed that clients want "direct links to online content – text and media formats". This finding is also mirrored in a report by Arendt (2008), in which the value of the index is questioned, as students prefer a 'Google Scholar' interface which highlights the availability of the items being sought. The next generation of federated search technologies will offer a pre-indexed database of full text content, and a more critical approach will need to be taken of the subscription commitments made by Libraries. The article database categories of the Library Search Box were revised early in 2009 in response to users concerns over the lack of full text results, with refinements removing many of the smaller indexes and emphasising full text publisher content. The next generation of discovery tools will see the rise of pre-indexing and Libraries will need to evaluate their subscriptions more closely and begin to develop new paradigms for determining value.

Although the absence of phrase searching in the original implementation caused the accuracy of the results to be questionable, information queries did not show any proportionate change over the 2008 – 2009 periods. Feedback received from staff in the latter part of 2009 showed that staff did not perceive the Federated Search to be effective in meeting the information needs of our clients, particularly as this feature was not instantly activated off-campus, requiring users to authenticate themselves via login, with one staff member commenting that students were "just not getting it". Other comments from staff included "it is good when you are not sure what you are looking for" and "I still use the Classic Catalogue to find what I want". Of the staff that provided negative feedback in the initial stages of the project, less than half admitted to using the system in preference to the classic catalogue. In discussions with staff it quickly became apparent that there was a lack of understanding of how the searching behaviour of Library clients has changed, and that speed of future catalogues will fundamentally change student research processes across the sector. Information Literacy paradigms are changing and there is an urgent need to understand new and emerging research processes.

With the speed of development of the next generation of discovery layer technologies, the project team realises that there is an ongoing need to directly compete with traditional search engines in regards to how much time and effort is put into adjusting our systems in order to make them more relevant and accurate. In the findings of a study conducted by Lossau (2004), the author found that a student's use of a catalogue was seeded in the reasoning that "for this information type, they don't have an available alternative". If the ultimate aim is to be more 'like Google' we

must be prepared to provide alternatives that are just as responsive, intuitive and predictive. Library staff should ignore what is traditionally appropriate and examine user needs. Much of the success of the Library Search Box derives from the ability to analyse usage data and make changes to the system to keep it both relevant and responsive.

Unlike our search engine competitors (with regional offices) an issue encountered during implementation was the level of technical experience available locally. With the development and support at the time being based in Amsterdam, the team was found providing feedback in the evenings and early mornings allowing the development process to be expedited. Initial support was provided from Amsterdam meaning that many local issues were not always resolved instantly with a response time of up to 24 hours. On call support was available, however the minor issues which the library would traditionally resolve using internal resources were now issues that required a 'support ticket'; a reality of adopting technology on a software as a service basis.

Conclusion

The UWS Library Search Box is by no means the perfect solution to clients' information needs and, like any new technology, it will not meet the needs of all users. UWS Library did succeed in providing an intuitive interface to most of its users; however, the Library recognises that until there are significant improvements to current search interfaces, there exists the need to provide access to traditional search tools.

What the current generation of discovery layers provide the Library with is a greater understanding of the needs of its clients and their information-seeking behaviour. This information will equip the Library with the tools to make better informed decisions about what its clients need in future interfaces as well as how they will interact with the Library.

Vendor support for future products needs to be vastly improved. Traditional catalogues have provided the Library with control over the technical aspects of a system. The new generation interfaces offer black box solutions that oftentimes require outside assistance. Libraries need to recognise this and demand better support. Support at both technical and operational levels is required; geographic location should not disadvantage implementation or use, and Libraries and the systems must be as responsive and reliable as the search engines they are now directly competing with.

Libraries will need to allocate staff and resources to such endeavours to ensure that new interfaces remain relevant. Discovery Layer products place Libraries in the competitive search engine and content provider environment. Consultation times will need to be expedited and a more reactive approach will need to be taken in order to remain competitive in the new environment.

Statistically accurate information is another requirement of future discovery layer products. The success of the search box at UWS derives from the ability to respond to statistical information on a daily basis. As an extension to this, we need to be able

to better account for usage and draw upon this information to inform changes to Library acquisition and expenditure. Similarly Libraries need to be prepared to ask the difficult questions surrounding subscriptions to niche products accompanied by niche interfaces. If we expect our users to be able to use a single search box, we need to be flexible with what is still considered a relevant resource for our users. Newer discovery layers carry the indexes to full-text collections; there is a need to be able to determine the value of having resources that will not provide the instant gratification of a full text results page to Library users.

The UWS Library experience evidences the need for extended implementation times. Libraries should not be concerned with how quickly a system can be implemented, rather determining how much time after the launch is needed to make sure that system is complete should be the priority. The providers of such systems need to be as flexible as the administrators involved and understand that with use, information on what Library clients actually need will become clearer.

In a 2001 study conducted to assess the usage of a Library OPAC, one professor commented that “a good catalog is the only tool that I really need from librarians” (Cook & Heath 2001, p. 574) and with a new generation of professors will come a new set of information needs. The UWS Library is well placed to meet this challenge.

References

Arendt, J 2008, 'Imperfect tools: Google Scholar vs. traditional commercial library databases', *Against the grain*, vol. 20, no. 2, pp. 26-30, viewed 4 September 2008, <http://www.against-the-grain.com/TOCFiles/v20.-2_Julie_Arendt.pdf>.

Boness, M 2009, 'E-NHANCING E-RESOURCEFULNESS', paper presented to ALIA Information Online 2009, Darling Harbour Exhibition and Convention Centre, Sydney, <http://information-online.com.au/sb_clients/iog/bin/iog_programme_times_are_e_changing_A4.cfm?vm_key=7029620E-1422-72D3-A090B6A9A422BA70>.

Boyd, J, Hampton, M, Morrison, P, Pugh, P & Cervone, F 2006, 'The one-box challenge: providing a federated search that benefits the research process', *Serials Review*, vol. 32, no. 4, pp. 247-54.

Cook, C & Heath, FM 2001, 'Users' Perceptions of Library Service Quality: A LibQUAL+ Qualitative Study', *Library Trends*, vol. 49, no. 4, pp. 548-84.

Granka, LA, Thorsten, J & Geri, G 2008, 'Eye-tracking analysis of user behaviour in WWW search', paper presented to Annual ACM conference on research and development in information retrieval, Sheffield, United Kingdom.

Jansen, BJ & Pooch, U 2001, 'A review of web searching studies and a framework for future research', *Journal of the American Society for Information Science and Technology*, vol. 52, no. 3, pp. 235-46.

Kinman, V 2009, 'E-Metrics and Library Assessment in Action', *Journal of library assessment in action*, vol. 21, no. 1, pp. 15-36.

Lau, E & Goh, D 2006, 'In search of query patterns: A case study of a university OPAC', *Information Processing and Management*, vol. 42, no. 5, pp. 1316-29.

Lossau, N 2004, *Search Engine Technology and Digital Libraries: Libraries Need to Discover the Academic Internet*, viewed 6 September 2009, <<http://bieson.ub.uni-bielefeld.de/volltexte/2004/531/html/06lossau.html>>.

Newton, V & Silberger, K 2007, 'Out-googling Google: Federated Searching and the Single Search Box', in *ACRL 13th National Conference*, Baltimore.

Online Computer Library Center 2004, *Perceptions of Libraries and Information Resources*.

2008, *Online Catalogs: What Users and Librarians Want*, OCLC

Onwuegbuzie, A, Jiao, Q & Bostick, S 2004, *Library Anxiety: Theory research and applications*, Scarecrow Press, Maryland.

Pavincich, M 2007, 'Information Central: Not just a call centre...', paper presented to ALIA Information Online 2007, Darling Harbour Exhibition and Convention Centre, Sydney,
<<http://conferences.alia.org.au/online2007/Presentations/30Jan.A2.information.central.pdf>>.

Solomon, G & Schrum, L 2007, *Web 2.0 : new tools, new schools*, 1st ed, International Society for Technology in Education, Eugene, Or.