Discovering the F-Word

Paul Mellinger explains why University of the Arts London’s library service decided to move to a next-generation federated search engine, at a time when most academic library services are moving to, or already using, a web-scale discovery product.

THE dual status of UAL Library Services as both academic library and specialised arts library, along with our unusual and historical collegiate structure, often directly influences our policies and decision-making, particularly in the areas of customer services and collection development and is almost certainly the cause of an oft-repeated refrain that we are ‘special’.

A perfect example of this was our decision to move away from using a web-scale, index-based discovery service, in favour of a next-generation federated search engine, Explorit (produced by Deep Web Technologies www.deepwebtech.com), precisely at the time when almost every other mainstream academic library service was either already using a web-scale discovery product or was in the process of procuring and implementing one.

An extensive report for UKSG in 2013, Impact of library discovery technologies,1 which primarily sought to assess the impact of library discovery technologies on the usage of academic content, found, in responses to a web-based survey of 62 UK HE libraries, that 77 per cent of survey respondents had already implemented a Resource Discovery System (RDS) at their institution, and a further 11 per cent were in the process of doing so.

Slow and clunky?
The very phrase ‘federated search’ had become something of a dirty term in library circles, conjuring images of outdated user interfaces and memories of slow and clunky Z39.50 ‘state-ful’ standards, where results are not returned until the slowest server gives a response. Indeed, these were the primary reasons why we had been swept up by the euphoria of the web-scale discovery revolution and decided to swap our original federated search engine, Metalib, for an index-based RDS in 2011.

The lure of Google-like search
Why wouldn’t a library service, conscious of the increasing reliance that its undergraduates were placing on Google to find their

Concerns about the efficacy with which our ‘e-Library’, as it was then called, could reliably return relevant full-text content on a regular basis, forced us to conduct a full-scale review.
primary sources of research information, be ecstatic that it could at last compete with the internet behemoth, at least in terms of UI design and responsiveness? This was borne out by one of the key findings of the aforementioned report for UKSG:

‘Increased usage is not the primary motivation for moving to a discovery technology – libraries are more concerned with user experience and providing a single search interface linked to full text. Undergraduate students are seen as the primary users and beneficiaries of library discovery technologies.’ (1, p. iv)

The problem, though, with modelling the look and feel of an RDS on Google is that this approach is in danger of elevating form over substance and, ultimately, risks offering a system built on a fallacy. Whilst an RDS may well use very similar algorithms as Google to search through its index of metadata in order to determine the relevance of the information in the index to what the user is searching for, it is the way that the metadata is harvested and the index compiled that is different and that creates an inherent flaw in the model.

Rather than sending out a crawler to read information from websites based on a search query, as well as determining the inter-relation of those sites with all the other sites on the internet, an RDS simply sends the query directly to its bank of pre-harvested metadata. Admittedly, this bank is vast and ever-expanding but, at any given time, it is finite and, crucially, the accuracy of its contents is always dependent on an equally ever-expanding array of sources.

Disenchantment and review

The disenchantment with the RDS model at UAL started to creep in about half-way through a three-year contract and it was at this point that the concerns about the efficacy with which our ‘e-Library’, as it was then called, could reliably return relevant full-text content on a regular enough basis, forced us to conduct a full-scale review.

This e-Library review, which collated feedback from students, academic staff and library staff, was conducted in the spring of 2013 and, from this, several key issues with the RDS were identified. The most obvious was that the occurrence of supposed full-text results leading to dead links was far too regular to inspire confidence in the underlying search system. Not only did it cause frustration to the end-users but it caused a great deal of discomfort and embarrassment to those members of library staff involved in information skills delivery programmes.

The dead links, as we now all know, were either caused by errors in the metadata, due to the harvesting and normalisation process; errors in back-end collection activation, due to variations in package contents from database providers; and technical issues with link resolvers working incorrectly.

These technical linking issues hint at the risk of a conflict of interests when compa-

nies who provide content also provide the discovery tools. The National Information Standards Organization (NISO) Open Discovery Initiative (ODI) was set up in 2011 to tackle just such problems, one of whose stated aims was to ‘facilitate increased transparency in the content coverage of index based discovery services and to recommend consistent methods of content exchange or other mechanisms’ 2 This culminated in the publication of a set of recommended practices (NISO RP-19-2014), which includes a conformance checklist for both discovery service providers and content providers to complete and publish.

The very fact that such steps were seen to be required could be seen to vindicate our desire to use a different sort of discovery tool altogether. At a time when students are already sceptical of the need for a dedicated library discovery tool, promoting something that looks like Google but doesn’t perform like Google only serves to increase their scepticism. But it wasn’t just the unreliable linking and sometimes seemingly spurious relevance ranking that created problems for us.

Which database?

The other key problem, which definitely highlights UAL’s ‘specialness’, was that there was no stated provenance for the results delivered by the RDS. For full-text articles, the title of the e-journal was given but what our users wanted to know was which databases the sets of results actually came from.

Despite the assurances of RDS vendors about the decreasing importance of databases as brands, our students, because they were being directed to specific databases by their teachers, wanted to know why, if they performed a certain search in a named database, it tended to produce very different results from the same search performed via the RDS.

Since federated search engines actively showcase databases as information sources, they are seen by academic staff as promoting the trusted, curatorial heritage of those authoritative sources that they themselves recognise and promote. Perhaps it is a symptom of the relatively slow pace of arts education, compared to the sciences, but Explorit seems to foster a sense of familiarity and feels congruent with the pedagogical practice at UAL.

Information literacy

This familiarity, in turn, facilitates the delivery of the information literacy programme. The simple ability to select which specific databases to cross-search seems to make it conceptually easier for students to engage with the research process. Furthermore, the ability to actually see the databases being searched in real-time not only grounds their understanding of the provenance of results but points to another vital benefit of the federated search model – that it delivers real-time results, rather than results that were harvested from suppliers some time ago and which may not be fully up-to-date.

The below example of an infographic used in some information skills sessions at Central Saint Martins shows how the concept of a deep, or invisible, web (‘Articles Plus’ is the branding currently used for Explorit) can yield more useful results than the visible web resources. Coincidentally, this utilises the same metaphor that DWT use (as shown on p. 30) of the predominantly submerged iceberg:

Profile of the library service

This explanation, by accentuating the latent value of library-purchased e-resources, also helps students to understand where a portion of their tuition fees are spent. Transparency in Library Services budget devolution is becoming increasingly important due to the Customer Services Excellence framework within which we operate. The below example shows why an authentication process is necessary to unlock these otherwise hidden treasures.

This model converges with the more traditional, analogue model of libraries as collections. Another quirk of arts libraries is the unusually high importance accorded to hard copy collections. This was one of the reasons why we have continued to use a traditional LMS and another reason why hiding the catalogue behind an RDS is not as important to us as it might be to other institutions, as explained by my colleagues, Jess Crilly and Karen Carden, in an Update article. 3
that information is consequently more easily understood and the profile of the department as a whole is therefore elevated.

Feedback and tweaks
When we went live with Explorit, we were in the fortunate position of being able to run it concurrently with the RDS for a year, so that we could conduct a second e-Library review project, in the 2014 spring term, eliciting further feedback from academic staff, students and library staff, through site visits, an online survey and academic liaison.

Apart from some feedback from a proportion of undergraduates who stated a preference for the simpler look and feel of the RDS, most of the collated feedback was in favour of keeping Explorit. Since then, we have liaised with Deep Web Technologies to implement a number of improvements, such as full-faceted navigation and a date slider and these have been incorporated into the new version, called Explorit Everywhere! (the exclamation mark is part of the branding).

More recently, in March 2017, we started to conduct a task group, looking at the layout and search options presented in our Library Search portal, which incorporates options for a catalogue search and an ‘Articles Plus’ search into an integrated, single search box.

Accepting our ‘specialness’
Ironically, this has highlighted a preference from library staff to extricate the two functions, in order to emphasise the difference between the two underlying systems and surface the full functionality of Explorit Everywhere! We intend to consult more widely with students in order to validate some of the re-design ideas that were discussed in the task group but it seems that our original insistence on trying to replicate the Google-style, single search box of a RDS may have been a mistake in Library Search. If our ‘specialness’ demands a special system to deliver our sometimes special resources, then we should accept and demonstrate that specialness, rather than hide it. [9]

References
4 https://libsearch.arts.ac.uk