The title of the 'Antaeus' column derives from the name of the mythical giant, Antaeus or Antaios. The son of Gaia (whose name means 'land' or 'earth'), Antaeus was undefeatable in combat so long as he remained in contact with the earth. Once grounded by contact with the soil, he vanquished all opponents. However, in order to disempower Antaeus, Heracles simply lifted him from the earth, overcoming him totally. Thus, many times through the centuries, Antaeus has been used as a symbolic figure showing how any human aspiration must remain grounded in order to succeed. LIS research must therefore retain its contact with the 'ground' of everyday practice in order to fulfil its potential as a sophisticated research discipline – it must remain empowered by its relevance to practitioners.
URLS in the OPAC: comparative reflections on US versus UK practice.

Abstract

Purpose of this paper
To examine whether placing URLs into library opacs has been an effective way of enhancing the role of the catalogue for the contemporary library user.

Design/methodology/approach
A brief review of the literature combined with an analysis of publicly available statistics for library use in the US and the UK.

Findings
That certain ways of placing URLs into the opac are loosely associated with a successful library environment, i.e., with constant or increasing levels of stock circulation and opac use, while other forms of hyper-linking opac records are loosely associated with declining levels of library use.

Research limitations/Implications
The loose association between different opac management practices and apparent statistical trends of library use could be investigated in greater depth by further subsequent research, but along the lines and methodology suggested herein.

Practical implications
Firm suggestions on how to place and manage URLs in the online catalogue are made.

What is original/value of the paper?
This paper takes certain catalogue enhancement practices which are identified with the US library environment and investigates them in a UK, and specifically Scottish context, to shed light on the original US ideas behind these practices.

Paper type: General Review

Keywords: Libraries; catalogues; online catalogues; United States of America; UK; Scotland.
**Introduction**

Librarians have been worried for some time that electronic services beyond the library building (such as commercially available online services, search engines like Google, and “the internet” in general) are taking users away from libraries, print collections and the library catalogue. The library building, its stock and its catalogue are mutually dependent, and so are bundled together in a spiral of decline: “declining demand for today’s catalogs reflects diminishing interest in already low-use research library collections...[but] the legacy of the world’s library collections is for the time being tied to the future of catalogs.” (Calhoun, 2006).

In spite of this gloom, there is plenty of discussion in the LIS literature disputing the evidence either way. In the UK for example, some have noted that physical visits to British academic libraries are in long term decline (Akeroyd, 2001), which in turn implies that catalogue use is in decline. But by contrast others note that UK HE library stock circulation is buoyant (LISU 2006). Similarly, borrowing from UK public libraries has been declining, but on the other hand visits to UK public libraries are holding up well (ibid.). This brief paper will look at a particular aspect of this complex set of trends in library use and examine whether one particular response by librarians to the perceived threat of the networks has been an intelligent and successful one: the practice of adding URLs for free internet resources and websites to library opacs.

**New internet cataloguing practices**

Librarians have always used their catalogues as finding guides to the stock kept within the four walls of their library buildings. Logically therefore, if the physical collection and library building are becoming less and less used, then the catalogue itself will be less and less used. It is difficult to cite any definitive proof of this, but some writers have pointed to local trends of declining catalogue use in their own libraries (Banks, 2000) while others have simply invoked gut instinct (Murray, 2006). One powerful voice confirming the significant trend towards decline in opac use is the Library of Congress (Calhoun, op. cit.). A recent report commissioned by them seems convinced of the catalogue’s dire predicament:

> “Today, a large and growing number of students and scholars routinely bypass library catalogs in favor of other discovery tools, and the catalog represents a shrinking proportion of the universe of scholarly information. The catalog is in decline, its processes and structures are unsustainable, and change needs to be swift.”

One answer to this perceived decline in catalogue use which became popular in the 1990s was to reinvent the catalogue in its online form by putting hyperlinks into opac records for ‘non-library’ electronic items, especially non-library materials. The 856 field in the MARC record is the field allocated to the url in contemporary cataloguing practice: the hyperlink is displayed to the browser from this field and, when clicked, full text can be pulled through the bibliographic description onto the screen by the user. Rather than losing readers to the net, net resources are brought into the library opac and net users’ electronic information seeking may spill over into use of the otherwise ignored print collection.

US libraries with extensive free subscriptions to government serials initiated this trend in the 1990’s by gradually starting to add hyperlinks to their opacs as and when government publications changed from print to online. It seemed logical to
indicate the continuity of content from one medium to another by enriching the record in this way, with a hyperlink.

As a result, the practice grew in ambition: the study of this form of internet cataloguing quite soon spawned its own scholarly literature with a dedicated journal (Riemer, 1997- ). Gradually the interest extended to creating:

- hyperlinked opac records for free-standing internet resources which only exist on the web

rather than ‘inertia cataloguing’ of electronic items such as:

- online continuations of print originals already established in library collections (especially ‘print to online’ migrated government publications), or
- paid for digital items (above all, electronic journals, and more recently e-books).

There are a number of gains from such opac-based internet cataloguing practice. By systematically adding urls for online resources into the opac, the opac becomes more than just a finding tool for what are - sadly - increasingly less attractive print materials. It is also a tool for retrieving exciting online materials that exploit the uniqueness of the internet.

In fact, by promoting the opac as an online information retrieval tool with integrated print resources, the information user with a preference for networked resources will encounter print resources alongside their favourite online resources and rediscover the print library via its catalogue. By as it were ‘ambushing’ them in this way, the reluctant library user may be seduced by the hyperlinked opac into finding print resources as good as or better than their electronic equivalents. They may indeed be converted back to the traditional print library.

**The lessons of experience**

The argument sounds seductively convincing. However, it is a good while since this form of cataloguing practice first became popular in the late 1990’s, so we should now be in a position to assess the value or otherwise of this well intentioned innovation. What does hard experience show us, some ten years or so further on?

Firstly, it should be said that there were many from the start who were sceptical about the value of cataloguing internet resources into library catalogues, especially resources that were not simply electronic facsimiles of familiar print formats, or were not deliberately purchased for the collection. Library catalogues, in print or electronic from, were never designed for this sort of task.

Thus, Antelman (1999) argued why we should “attempt to accommodate the[se] new resources in the old gateway?”. Given the limited functionality of the typical web-based opac, the ‘webpac’ is not ideally suited to capturing details about web objects and must be distorted to accommodate the new online world.

Above all, catalogues are built for the static world of print, where the print object, once purchased, has an unchanging location on the shelf and its content does not change as it sits in the library collection. Put more simply, URLs decay, so URLs in catalogue records must be maintained to offset this decay. This is real problem in a number of ways. Opacs are essentially databases, and databases, unlike hierarchies
of hyperlinked web pages in a standard browsable web site, do not make themselves available for link validation very readily.

As Tyler (1999) says, “Dealing with problem URLs in the library catalog is far more complex than in the more familiar ‘homepage’ environment.” Link checking mechanisms are not often available as standard functionality in many integrated library systems, and the alternative to integrated link validating software is to graft on an external package to your cataloguing system routines. This in turn involves exporting links into a discrete file for checking as a separate, regular task. This is onerous.

And even if link checking is possible, the fact that the link enshrined in a url may not change, but the content to which it refers might, again challenges the conventions within which catalogues were created. Domains can remain the same but be used for completely different sites – the more alarming results of such changes to educational web sites listed by educational institutions have been reported quite widely (Taylor, 2001). By contrast, print items do not alter their content once catalogued.

This infinitely extending content-validation task is nigh impossible for the human cataloguer managing a traditional opac. Metadata created by the human mind consist of a snapshot of an item’s content taken at one moment in time. Metadata generated by machines such as search engines can crawl over entire texts andanalyse it on a recurrent basis through time. The best that a librarian can do is keep re-reading the internet ‘stock’ in their catalogue to make sure it has not changed its nature. Again, this is onerous, but omitting to do so can, in the very worst cases, leave the institution hosting the opac highly embarrassed, if not worse - vulnerable to legal action (ibid.).

The evidence against
If accurate URLs in the opac cannot be maintained, then the percentage of links which fall into error accelerates rapidly. On the open web, this phenomenon is called ‘linkrot’ (Nielsen, 1998). A number of authoritative studies tell us that URLs break within only a few months (Markwell and Brooks, 2002), after which they must be repaired. Opac hyperlinks will therefore fall apart very soon at this rate.

So how have libraries dealt with this problem?

Burke, Germain and Van Ullen (2003) carried out a recent analysis of the issues of catalogue reliability and accuracy as affected by the practice of adding URLs into the opac. Their intelligent study of the members of the Association of Research Libraries in the USA showed that the fears of the early sceptics about URLs in opacs have in many ways been realised: “A large percentage of the researched library catalogs...had a significant number of errors.”

Looking specifically at those 24 ARL catalogues which linked to free URL resources as opposed to subscription materials, 67.50% of respondent libraries had an average error rate over 14.58%. The highest error rate was no less than 58.33%. Yet this is hardly surprising, given that only 37.5% of the respondents to the survey reported doing any type of automated link checking of the URLs in their catalogues. Burke et al. contrast this high error rate for broken URLs with what we know about the average incidence of missing print monographs listed in traditional well maintained catalogues of print collections— for example, this was calculated at just 4.30% in a representative US University Library (ibid.).
Most alarmingly – especially in view of Taylor, 2001 – not one library in the survey reported checking for content consistency. In the face of such error rates, the authors of the study ask:

“As librarians, do we accept this as a tolerable rate of failure in a resource valued for its authority? Since our mission is to keep the catalog relevant, do we [not] dilute its credibility with resources that are not stable? Are we not concerned with the possibility of catalog records quickly turning into erroneous information?”

Given the seniority of those who responded to the survey – collection development heads and technical services heads – one might conclude that the majority of these senior LIS professionals’ response to these despairing questions would be in the negative. Simply, for them, incorrect URLs in the catalogue may be better than no URLs, because at least then the catalogue shows that libraries indeed can ‘do’ the internet, albeit with results of variable quality.

The response
Subsequent to Burke et al’s sceptical study being published in 2003, there has been at least one direct response to their challenge, demonstrating that it is possible to add URLs to an opac, and to maintain their accuracy, while also improving the service to readers. Brown’s 2004 study of patterns of access to US federal online documents via opac URLs, which references Burke, Germain and Van Ullen’s critical paper, states that, at the University of Denver, their use of URLs in the opac has been successful and carried through to a high standard.

In 2004 Denver maintained no fewer than 182,329 URLs in its opac, which is a sizeable amount. The study tracked URL click throughs, compiling data about them via a database system and estimating the value of these click throughs by analysis of the collected results. Brown points out that a regular link checking routine is maintained for their Library’s URLs, carried out weekly by means of their integrated library system’s automated link checking facility. Even so, Brown admits “This large number of URLs creates a management headache for URL maintenance” – how big we aren’t told, because the author does not actually give us a percentage error rate as an indicator of ‘linkrot’ or URL accuracy. This is a failing in the study, in view of the importance of this measure as established by Burke et al.

On the positive side, the author shows that the total number of online accesses (3,809) is higher than the total print circulation figures (2,080) for this type of material over the one year period of the study. He concludes that “federal document URLs in the OPAC are worthwhile: users are finding documents they otherwise would not have found. Moreover, users are using the additional URLs we have added through our aggressive URL-adding projects.” And the evidence adduced does seem to bear this out.

However, Brown expresses the reservation that equally “[users] could do a Google search and land upon government documents.” This inevitably raises the question: why bother to create and then check some 200,000 opac URLs in the first place when Google will maintain the same links for your users for free?

Discussion
Brown (op. cit.) and Burke et al. (op. cit.) represent two different views of the value of URLs in the opac. Brown points to the increase in traffic through opac URLs in one library and argues that this proves their value. Burke takes a wider view of a larger community of libraries, and argues that the problem of ‘linkrot’ in opac URLs is a serious factor that promotes the ‘disintegration’ of the opac rather than ‘integrated’ searching for print and electronic items via the same library search tool.

In order to help shed more light on the matter, we will look at experience of this issue in the UK and offer some points of comparison. Rather than conducting original research, as did Burke et al. (op. cit.), we will use publicly available statistics.

Both the ARL in the US (ARL, 2006) and SCONUL (2006) and LISU in the UK (LISU, 2006) collect useful statistical data on library usage on behalf of their respective LIS communities. And broadly speaking, as noted in the introduction to this paper, these statistics shows that there is a long term trend in the US ARL library community, starting in 1991, whereby circulation figures (and by inference opac use, which is the means by which items are found prior to being circulated) have fallen year on year without respite. Yet this is the library community that has adopted the practice of adding URLs to the opac most extensively. If the integrated opac was meant to stave off the decline in catalogue and collection use, then, on the evidence of broad trends, the experiment has failed.

By contrast, the UK HE community has been far more cautious in spreading URLs into their opacs. Yet their circulation figures over a comparable period have been holding up well and have increased in many cases. The resilience of UK circulation figures is remarkable, because it has been maintained in spite of the fact that there are fewer resources available to UK libraries for building up attractive collections of stock than in the USA.

There seems to be some sort of very general pattern whereby URLs in opacs are associated with libraries in decline, in contrast to library systems with a conservative attitude to URLs in the opac, whose collections and opacs seem to remain in higher demand.

If Burke et al. (op. cit.) are correct, then integrating URLs into the opac has certainly created a problem with ‘linkrot’. But we can add to that negative finding our perception that such ‘linkrot’ may even have accelerated the user’s disinclination to use libraries and their opacs, at least in national library systems which have adopted the practice on a large scale basis. However, this is a perception based on very general trends, and is a finding that could only be asserted confidently by examining specific libraries, their opac practice and their own specific institutional patterns in catalogue use and stock circulation.

Nevertheless, it is no more than common sense to hazard a few thoughts about the impact of ‘linkrot’ on opacs. Although broken links do not directly contaminate accompanying records with reliable hard copy locations in them, the cognitive overload of distinguishing incorrect internet resource records from correct print item records must surely exasperate the information user. Opac interfaces are thought to be intrinsically unattractive when compared to alternative search tools: "Electronic catalogs, wherever you go in the academic world, have become a horrible crazy-quilt assemblage of incompatible interfaces and vendor-constrained listings." (Burke, 2004). The last thing libraries should do is increase the difficulty of using them.
Above all, if our target user is an impatient internet junkie, with the attention span of a television advert, overexposure to broken links in the opac will fast exhaust their faith in the entire catalogue system. The experience of lots of ‘HTTP 404 - File not found’ messages has a very clear meaning to a surfer: ‘This site is no good!’ Truly therefore, as one writer has said, ‘Rotten links hamper learning.’ (Dean, 2002)

**URLS in the opac: UK practice, focussing on Scotland**

To explore this further, it is helpful to examine UK opac management practice in a little more detail. Specifically within the UK, recent analysis of circulation figures in the public domain (Joint, 2004) has shown that circulation trends in Scotland are in step with the general pattern of UK academic and libraries. Circulation figures have been steady or rising in the main, with an average increase in hardcopy issues throughout the decade following the advent of the first commercial web browsers in 1994 of some 15% overall. There are no directly available figures collected for opac use, but let us infer that the opac, as the gateway to the collection, has been well used in parallel with these circulation figures.

In terms of Scottish HE library policy on URLs in the opac, this is also broadly conservative, in tune with UK general practice. There are three general approaches, based on the categories described earlier in this paper:

- **The middle way**: to catalogue only those online continuations of print originals already established in library collections (especially ‘print to online’ migrated government publications), or paid for digital items (above all, electronic journals).
- **The path of no change**: To maintain a divide between all online documents and print items, using the opac purely as an inventory of print items, pushing lists of online electronic resources into the library’s separate web site.
- **The radical path**: To catalogue familiar library items (‘print to online’ migrations and paid for digital items), but also to include freely available web sites that would normally be the preserve of Google and other search engines.

As a general rule, Scottish libraries do not use link checking facilities in their integrated library system to validate links and avoid ‘linkrot’. By and large this is because Scottish HE libraries have been pursuing ‘the middle way’ in creating URLs in the opac, and have avoided including freely available web sites that would normally be the preserve of Google and other search engines, while creating URL links to stable items traditionally associated with the Library collection.

One great benefit of this approach is that link maintenance can be carried out via alternative means than whole scale checking of links. If a library has an Open URL resolver package or a good serials management package, these will effectively do the job of URL maintenance for you, without resorting to use of an integrated library system link management tool or an external utility such as Xenu (Hausherr, 2006). Open URL resolvers and serials management packages do not help libraries manage hyperlinked opac records for free-standing internet resources which only exist on the web. Only dedicated link checking software can do this.

One of the down sides of using dedicated link checking software is that checks may be carried out on all the e-journal links from a single commercial supplier’s server, causing the supplier server overload and considerable inconvenience – the copious accesses may even be misinterpreted as a denial of access attack!
It is important, therefore, when thinking about link management in the opac, to make a distinction between two types of URL:

- **Type 1) COMMON STEM URLs** (different ones derived from one common source server location/domain). These are primarily culled from a single source - e.g. NetLibrary, single commercial ejournal suppliers and so forth. These all tend to have the same stem between the http:// identifier and the first slash, and a different file name thereafter.
  Examples (all are URLs taken from a Scottish webpac and were valid at the time of writing):
  

- **Type 2) DISPARATE URLS** (different ones from many sources or from many server locations or domains): - e.g. self-contained web sites. These all tend to have different stems between the http:// identifier and the first slash, while the file name following, if there is one, will vary as in 1).
  Examples (all are ‘linkrotted’ URLs taken from a Scottish webpac and were broken at the time of writing):
  
  - http://jobs.psbappointments.co.uk/
  - http://turpion.keldysh.ru/
  - http://www.rsc.org/is/journals/current/russian/rcrcon.htm

Using Type 1) URLs in your opac has the advantage of link stability, and conform to the patterns of user demand found in studies of hyperlinked opacs by recent writers investigating Burke et al.'s criticism of opac URL management practice. They have found that readers use hyperlinked opacs “for connecting to the electronic resources subscribed to by the library” (Ortiz-Repiso et al., op. cit.), rather than for free web sites offered on the open internet.

In particular, the studious avoidance of Type 2) URLs avoids the previously mentioned ‘domain change danger’ whereby content alters without the URL breaking, as noted in the popular press (e.g. Taylor, op. cit., describing the fate of Ernst & Young's lapsed moneyopolis.org domain; and Dean, op. cit., who notes a study reporting that “...a handful of links changed into porn links, which could be a real concern...”). There are examples of not dissimilar problems in unmaintained Scottish webpac URLs:

International tax systems and planning techniques bulletin.
Imprint: London : Thomson Tax Ltd.
http://www.thomtax.co.uk/

The link above is clearly no longer to a website that corresponds to the opac metadata describing it.

A final advantage of using only Type 1) URLs in your opac, derives from the impact of opac URLs on the integrity of the data within you library systems file. The more risks you take with opac URLs, the more likely it is that you will slowly corrupt the data in your systems file. To some extent there have always been errors and inconsistencies in catalogues, and readers have simply taken these in their stride.
Indeed, in the web environment there are those who say that broken URLs are expected by readers who will happily ignore them (Dean, op. cit.):

"'Typically we don’t react to a missing link,’ said Craig Clawar, assistant director for technical operations at the Professional and Distance Education program at Rensselaer Polytechnic Institute. ‘Most of our audience is sophisticated enough to know how to react to that.’"

This is all very well in open web sites or in VLEs. But library system files have at some point to be migrated from one version of a system to another, or from one library system to another system entirely, if the host library is changing system supplier.

When this happens, the integrity of the data held in a system must be checked. ‘Under-managed’ links in the opac detected at this stage will raise the costs of system file migration, because, like corrupt ISBNs, they are easy for commercial suppliers to detect, and they present evidence that the systems file is badly maintained and even corrupt. The supplier will be suspicious. They will quite justifiably want compensation for dealing with an unreliable set of data. So a proliferation of messy Type 2) URLs will raise the cost for (or even jeopardise) this whole process of file transfer.

Finally, a few words about the path of no change. Again, it is largely a matter of common sense to realise that users who find hybrid print/electronic journals catalogued in an opac as purely hard copy entities will find such information incomplete. It is frustrating for them to have to look for details about parallel electronic versions in a separate set of web pages maintained outside the opac by their host library.

Thus both the radical path and the path of no change in opac development are fraught with difficulties. In the anonymised 2004 study of declining circulation statistics in Scotland (Joint, op. cit.), the two libraries who report decline in circulation figures and, by inference, opac use, have respectively pursued one of these two opac management policies. Unlike this minority of Scottish HE libraries and their US counterparts, the majority of Scottish HE libraries have pursued the middle way, and certainly this had had no negative effect on their statistics of library use.

**Conclusion**

The conclusion of this paper is that ‘the middle way’ in managing URLs placed into the opac is the most sensible way forward. Of course, it remains the case that the 1990’s vision of an ‘integrated’ opac, which can search both library resources and the open internet, is a powerful one. However, there are now better ways available to us which can achieve this goal of integrated searching. And distorting the opac by overloading it with inappropriate URLs may damage the value of the catalogue to users rather than enhance it (leading to the ‘disintegrated’ opac described by Burke et al.)

A recent study notes how the increasing Open Source movement will create a “revolution” in library systems and software (Dorman, 2005). Commenting on this, Macgregor et al. (2005) note that

“Amazon Web Services (AWS), for example, can now easily be invoked by a library OPAC during a query to provide added value to a user’s result set.
Book jacket images, reviews, contents pages, etc. can all be integrated within an OPAC results page to enrich individual bibliographic records."

It seems that the goal of integrated searching will be achieved by creating meta-search facilities which combine opacs together with other tools in an interoperable portal relationship, rather than by hardwiring URLs into the ‘guts’ of opac records via the MARC 856 field. This is certainly the vision of those such as Campbell & Fast (2004), whose idea of opacs and the semantic web informs Macgregor et al’s recent description of the future catalogue:

“The ability of OPACs to offer access to a wide range of library and information services supplied by multiple organisations, or "shared services", will reduce the need for the single local gateway currently exemplified by the OPAC. There will be many ways into the information environment, ranging from personalised domestic portals to predetermined sets of services offered by specific organisations for specific purposes; libraries will be a subset of the latter...

“And we can expect multiple layers of service sharing, from local to regional to national to international, with different aggregations in each layer serving different user groups. It is highly unlikely that a single central catalogue of all the world’s resources will ever be feasible, so at some level this must involve a distributed approach using some of the “hyper-clumping” ideas explored in the CC-interop project (CC-interop, 2005)."

Libraries may resent becoming ‘a subset’ of portal services offered by specific organisations for specific purposes. But if this is what the future asks of us, we must answer that call and take our place as an important provider (if not a monopoly provider) in the information universe.

The future of the Library and its familiar information retrieval tools thus appears extremely bright. However, we must preserve the best of our traditional values (such as a commitment to high standards of accuracy) rather than chasing after novel but ultimately unproven practices in cataloguing and metadata innovation. In this way we will maintain a role that is recognisable as part of the unbroken tradition of library and information practice that has underpinned scholarly research and teaching throughout previous millennia.

However, if we do not maintain a path of intelligent moderation and sceptical innovation, we will endanger that continuing role. The future of information provision will look more like a break with tradition than a continuation of traditional LIS work. Such an outcome would be bad for information users as well as bad for information workers.

To this end, the investigations above are offered as an attempt to sketch out an idea of what this middle way looks like in practice, with reference to one particular aspect of contemporary LIS work. It is up to intelligent reflective practitioners to take such investigations and decide on a way forward.

But we should note that, although our future is in our own hands, the user is the final arbiter of our fate. Ultimately we must offer to them tools that are fit for purpose. If we do not, we face a future of decline and desuetude, a future that will be etched in simple statistics showing plummeting levels of library use. We need to
look at what works and what does not work, and make decisions based on what the numbers tell us: this is the only basis on which we can look to the future with confidence.

Nicholas Joint
Centre for Digital Library Research/
Andersonian Library
University of Strathclyde.

References


[See also accompanying Press Release:]


Murray, P. “..in my experience, there has been a decline in the use of library catalogs.” Disruptive Library Technology Jester. Posted by the jester on Tuesday, June 13th, 2006, at 9:50 pm, and filed under Blue Sky, Disruption in Libraries, Library SOA. < http://dltj.org/2006/06/dis-ils/ et seq. > (accessed 13/12/2006)

Nielsen, J. (1998) [Nielsen refers to two authoritative surveys:] ‘6% of the links on the Web are broken according to a recent survey by Terry Sullivan’s All Things Web. Even worse, linkrot in May 1998 was double that found by a similar survey in August 1997’.


Received 15/12/06, Reviewed 20/12/06, Revised 21/12/06, Accepted 25/12/06