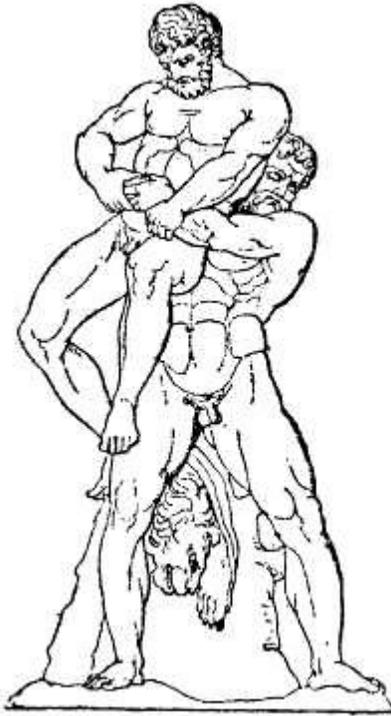


**The Antaeus Column*:
Digital information and the 'privatisation of knowledge'.**



Heracles defeating Antaeus.
Public domain image: from the old Swedish encyclopedia *Nordisk familjebok*.

* 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.

Digital information and the 'privatisation of knowledge'.

Abstract

<i>Purpose of this paper</i>	To point out that past models of information ownership may not carry over to the age of digital information. The fact that public ownership of information (for example, by means of national and public library collections) created social benefits in the past does not mean that a greater degree of private sector involvement in information provision in the knowledge society of today is synonymous with an abandonment of past ideals of social information provision.
<i>Design/methodology/approach</i>	A brief review of recent issues in digital preservation and national electronic heritage management, with an examination of the public/private sector characteristics of each issue.
<i>Findings</i>	Private companies and philanthropic endeavours focussing on the business of digital information provision have done some things - which in the past we have associated with the public domain - remarkably well. It is probably fair to say that this has occurred against the pattern of expectation of the library profession.
<i>Research limitations/Implications</i>	The premise of this paper is that LIS research aimed at predicting future patterns of problem solving in information work should avoid the narrow use of patterns of public-private relationships inherited from a previous, print-based information order.
<i>Practical implications</i>	This paper suggests practical ways in which the library and information profession can improve digital library services by looking to form creative partnerships with private sector problem solvers.
<i>What is original/value of the paper?</i>	This paper argues that the LIS profession should not take a doctrinaire approach to commercial company involvement in 'our' information world. Librarians should facilitate collaboration between all parties, both public and private, to create original solutions to contemporary information provision problems. In this way we can help create pragmatic, non-doctrinaire solutions that really do work for the citizens of our contemporary information society.

Paper type: General review

Keywords: Libraries; Digital Libraries; Library history; Preservation; Collection Management.

Introduction

Librarians often hark back to the golden age of Victorian public libraries. This was the time when it seemed all worthwhile knowledge resided fully in the public domain as a collectively owned asset. Any individual, rich or poor, could walk into one of the great municipal public library reference libraries and access – or, if not access, at least locate - all of the significant knowledge that was important to the citizen or scholar. This idea of the Victorian 'social' library has been summed up as:

"The idea of the library as an inclusive institution, playing its part in the construction of an organic, harmonious community through the creation of a common culture..." (Muddiman, 2000)

As Muddiman points out, this is an over-simplification of the reality of the Victorian library – and, we should add, of LIS professionals' nostalgic image of Victorian social libraries. But it is a clear statement of one view of what libraries should do: they should put knowledge into the public domain as a social benefit.

This 'social' view of the post-Industrial Revolution history of the library can be used to hold up a mirror to the new information order. How socially responsible is this new information order? Can we use the Victorian model to assess how well information is provided in a society that is going through a new revolution, the digital information revolution?

And is it true to say that the apparent annexation of digital technology provision by large multinational corporations has been a retrograde step, creating a digital divide which makes one long for the inclusiveness of print-based public libraries? Has digital technology effectively 'privatised knowledge', to the impoverishment of us all?

The privatisation of knowledge

Superficially, a social critique of the digital technology revolution does seem to have much evidence in its favour. The ownership of society's prime information assets seems to have passed from public into private ownership during the last half of the twentieth century. The municipal public libraries of the Victorian metropolis, or library of the nation as envisaged and created at the British Museum by Panizzi – all of them examples of disinterested and inclusive information provision - have been pushed aside by new owners of cutting edge information tools and information resources. These information owners are, by contrast, with their Victorian predecessors, profit-making corporations.

At the same time a 'digital divide' between information 'haves' and information 'have-nots' has appeared, seemingly in correlation with the emergence of these profit-making corporations. Could one even say that the privatisation of knowledge may have caused this digital divide?

Two of the most obvious examples of such corporations are Microsoft and Google. The social critics of these companies can point to their many failings. Microsoft's market dominance means that it can obliterate competitors such as Netscape, who produced the first genuinely original mass-market web browser, by cloning their information tools and forcing them down the throats of unwilling consumers. Microsoft's Internet Explorer (IE) browser was a shallow derivative of Netscape, but IE established market dominance by being bundled together with Microsoft's Windows operating system. If you wanted Windows, you got IE, like it or not.

Windows itself was also a shallow derivative of a sparkier competitor – the Apple-Mac graphic user interface – and lay on top of a command line operating system that was, unsurprisingly, a shallow derivative of a public domain product, the UNIX operating system. Arguably, Microsoft has never had an original idea in its entire company life – it is merely a capitalist juggernaut that eats up other people’s ideas just because it knows how to make money out of them better than their originators.

Google, by contrast, *has* had one original idea – the idea of a search engine algorithm that weights its ranking of results in favour of sites that are much linked to by other web sites. However, its success has led it to behave like any other corporate shark. The inner workings of the Google search algorithm are a closely guarded secret, as is the code behind Microsoft’s operating system (imagine the Anglo-American cataloguing rules and the principles of the card catalogue being company secrets that could not be taught openly in library schools!)

There is worse: as is well known, Google has gained access to the highly profitable network space of China’s internet market in return for censoring searches in line with Chinese government requirements: “a search on Google.cn for the banned Falun Gong spiritual movement directs users to a string of condemnatory articles.” (BBC news web site, 2006) The Google company motto is ‘Do no harm’ – unless, of course, you’re dealing with opponents of an authoritarian South East Asian regime, in which case it’s okay to do them quite a bit of harm in return for massively increased market share.

Unlike Google, the inclusive public library has always been a genuinely benign institution. A century ago, the main tool for information retrieval would have been a non-proprietary device in the public domain which everyone could reverse engineer and use for their own purposes: the library catalogue.

But, no matter how much we dislike it, the fact is that much, if not a majority of digital information today is retrieved via Google’s proprietary search engine, from a Microsoft corporation PC attached to the internet. Dare it be said, this paper was composed entirely on a Microsoft PC with much of the information in it gathered through Google. And, even worse, it was written on a home computer linked to the internet by Sky broadband – Rupert Murdoch’s internet service provider helped the author write these stinging criticisms of internet capitalism!

Public Good – Private Bad?

Having made the confession above, it’s clear that raucous social critics of the so-called privatisation of knowledge need to think carefully lest they disappear in a cloud of sanctimony! The social analysis of information provision is a bit more complex than facile moralists would have us believe.

In fact, any vision of the Victorian social library as inspired by unalloyed and uncompromised ‘good intentions’ is naïve. Muddiman (op. cit.) points out that many of the proponents of Victorian libraries sitting on library boards were themselves successful capitalist philanthropists - ‘the spirit of philanthropy motivated many library supporters very deeply’. Their investment in libraries was both disinterested and self-interested: they wished to promote an inclusive and organic community, but they frequently wanted to focus on the deserving poor, whose self-improvement via the resources of the public library created a bigger labour pool of well-educated, literate and knowledgeable workers.

Good social information provision more often than not comes about as a result of a complex meshing of public-spirited intentions supported by private capital. The

public and the private have to work together to achieve the best results for society at large – and anyone who thinks different is probably marooned in a student union debate that last took place in the 1960's.

Privatisation of knowledge = preservation of knowledge?

To illustrate this point, we can look at one particular topic in digital librarianship that has greatly exercised the LIS world in recent years. We have been concerned for some time about the issue of digital preservation. In library terms, we have tried to understand this problem by extrapolating from the past to the present – that is, we have used the past model of public/private involvement in information preservation and applied it to the present and future state of digital information preservation.

In the past, private publishing companies have created the majority of print information resources, sold them into the community and to libraries, and then taken no interest in their long-term preservation. Libraries have fulfilled the (public) preservation role for resources originally manufactured in the private sector. Both partners have been happy with this situation.

Now that digital information resources – especially electronic journals – form such an important part of contemporary research library provision, librarians have become concerned about who is to fulfil the traditional role of the library in creating long term digital preservation facilities for electronic resources provided by commercial publishers. Print-based libraries have no tradition of digital preservation, although they have been able to offer good access tools for digital resources (webpacs, library web pages listing electronic journals and the like). However, without a long-term digital preservation strategy, this reduces digital information to the status of ephemera.

The underlying belief amongst the library profession is that creating long-term digital archives for the public good is a noble function that we cannot expect the private sector to fulfil. Indeed, as the privatisation of knowledge has accelerated apace in the digital age, it has seemed that we might be facing some sort of digital holocaust, as commercial companies turned their back on the datasets to which they supply only short-term access.

As recently as 2001, collection development librarians were lamenting about archival e-journal back-sets:

“Few electronic journal providers have an archiving policy in place to ensure access into the future. The ongoing availability of a journal if the publisher ceases to exist, or the title moves to a different publisher, is also unresolved.”

Australian Library Collections Task Force (2001)

However, in the intervening six years, commercial electronic journal providers have become more and more able to offer some sort of facility for providing back-sets of digital files. For example, Elsevier now provide retrospective e-journal backsets (ScienceDirect, 2007) – and some of the Science Direct files have start dates as far back as the nineteenth century, though most start off back in the mid twentieth.

This does not mean that the digital preservation problem is solved (“The ongoing availability of a journal if the publisher ceases to exist, or the title moves to a different publisher, is also unresolved” – this is indeed still an open question). But the fact that a firm such as Elsevier, who have been one of the most reviled companies in digital information provision over the last decade, can find some

interest in combining its corporate goal of making profits with helpfully taking over the information preservation role of libraries, should discourage us from thinking that the future is always a mechanical replication of the past. For 'privatisation' of knowledge, should we read 'preservation' of knowledge?

Digital preservation formats

Another part of the digital preservation debate which has a 'public versus private' dimension is the topic of digital preservation formats.

If we are to preserve the digital objects that comprise digital collections, it is best done by specifying a preferred digital format for those objects before the collection is even built, let alone preserved. Institutional repositories built by self-deposit of e-prints/preprints are particularly vulnerable on this issue: the ethos of self-deposit gives choice of format to the person who deposits. In terms of standardisation of digital preservation formats this is a bit of a nightmare. One way round this unhelpful diversity is to intervene at the point of self-deposit to convert the e-print/preprint to a standard preservation format. The problem is, which format?

Most e-journal collections are made up either of HTML (hypertext mark-up language) files or PDF (portable document format) files. How do these two file formats compare from the preservation point of view?

The HTML format (and its various derivatives) has many champions who view it as the best preservation format. A public domain entity owned by no-one but overseen by the World Wide Web Consortium (W3C, 2007), it is comprised of simple elements that are standardised and irreducible: plain text ascii characters, which tend to create very economical file sizes. Because it is non-proprietary it is a format that can be completely controlled by the owner of any collection of HTML content. So you have complete freedom of action over the 'preservability' of the format in which you have created your digital library – you aren't going to find your format is unusable because the company which supports it has gone bankrupt.

You can also separate out the formatting elements from the content elements in your HTML-based digital library, which enables you to perform preservation interventions on the evolving formatting elements while keeping the stable content intact.

However, PDF has its own champions as well. PDF is a proprietary format, owned by Adobe Systems Inc. (Adobe, 2007) – so, by using PDF, you do risk finding your format is unusable because the company which supports it has disappeared or lost interest in supporting that format. However, this inevitable lack of control that is the result of using a proprietary format as a preservation standard does have attractions.

Because HTML is locally variable, it can be quite an eccentric standard, even within the parameters set by W3C validation (W3C, op. cit.). Because of its ability to be controlled locally, preservation solutions may also have to be created locally, where your local HTML 'dialect' is spoken. And quite frankly, fairly dismal, non-standard HTML can be created locally, outside the parameters set by W3C validation, HTML which is still intelligible to a browser. The ability of browser software to resolve dodgy HTML is a mixed blessing: it encourages low mark-up standards in casual web site authors. This does make the web a more open place to set up your HTML web site. It also guarantee a short life-span for such sites.

By contrast, Adobe PDF files are a 'closed box' owned by someone else. In preservation terms, this feels quite nice: you can't corrupt the standard because of

the closed nature of the files, and the company that owns the software has some sort of commitment to the lifespan of its legacy PDF files if it is to survive in the market.

But the W3C has no financial motivation to maintain the lifespan of HTML files. It is never going to produce a magic bullet to cure your preservation problems, although W3C may advise you how to go about it. Most of us don't want advice – we want a fix by means of which our files can be 'treated' so they last more than ten years before corrupting.

To sum up: there is no obvious solution to the dilemma of choosing an ideal preservation format. However, just as commercial e-journal companies have become quite interested in creating back-files of their own material, so may file format owners prove to be equally interested in solving the preservation problems of their own file formats. If you own a problem, you will try and solve it. Because public domain file formats are owned by no-one, the person who owns these formats' preservation problems is the person who uses the file format – the digital librarian, for example. Librarians may therefore prefer to use file formats whose preservation problems are owned by someone else – PDF sounds good from that point of view.

Archiving the internet

Finally, and quite summarily, perhaps the biggest challenge of all – archiving the internet itself!

In the UK we now have some sort of a statute permitting electronic legal deposit (Legal Deposit Libraries Act, 2003). It is a quite conservative statute, using the past model of information preservation as the model for future information preservation: the national library is likely to play a lead role in creating the future archival repository of the nation's 'digital objects'.

One simple question should be asked: if one compares the selective approach of this conservative model of national electronic collection building based on past models (for example, as exemplified by UKWAC, the UK Web Archiving Consortium, 2007) with the approach of an internet entrepreneur who made his money in commercial enterprise (Brewster Kahle, and the Internet Archive, 2007), which looks more promising?

UK legal deposit approaches to preserving the national electronic heritage have tended to focus on sampling a representative amount of the national electronic output. This is fine as far as it goes, but it does ignore the nature of the web: the World Wide Web is hypertextual. Because each hypertext document invokes other hypertext documents which in turn invoke other hypertext documents, and so on ad infinitum, a sampling approach seems conceptually self-defeating. The web is in a sense a collectively authored single document and preserving just part of it means that you are left with a sequence of bleeding chunks rather than self-contained entities. Even if one created a complete national archive of the national web space, it in itself would also be a vast bleeding chunk. A national electronic deposit collection can only leave dead hyperlinks that point beyond the national web space, looking out hopefully out into the international blank space of web domains. The Victorian national library deposit collection philosophy simply doesn't work so well with the Internet.

Kahle's approach with the Internet Archive is different from this public domain model of national digital preservation based on the print legal deposit model. His aim is

pretty much hubristic in its intent: 'Brewster Kahle's modest mission: Archiving everything' (Mills, 2006). This may be impossible but the results, when actually used, do seem more complete than anything that the narrowly derivative public domain model has to offer. This is not least because, like Andrew Carnegie over a century before, Kahle is an entrepreneur, whose success in the private sector gives him the wherewithal to try something breath-taking in its ambition:

"Kahle is not opposed to companies turning a profit--he pocketed millions in 1995 when AOL bought his first company, WAIS, one of the first Internet search systems. Much of that windfall went to fund the Internet Archive, which has an annual budget of about \$5 million." (Mills, op. cit.)

Here is a devilish suggestion: why not extend the legal protection of the 2003 Legal Deposit bill to the Internet Archive, and redirect public funding that would have supported the UK library community's selective electronic deposit activity to Brewster Kahle, specifically to archive everything held in the UK's internet domain? The synergies of such an arrangement would make for an immensely powerful public/private synergy.

This suggestion is partly made to be provocative, of course- but it would be deeply intelligent and highly constructive to extend the protection of legal deposit status to online libraries like the Internet Archive, since the actual legal status of their archiving activity is not protected and supported in the same way as a statutory national deposit library.

Conclusion

The overall thrust of this paper is to point out that past models of information ownership may not carry over to the age of digital information. And this may not necessarily be a bad thing – the fact that public ownership of information (e.g. by means of national and public library collections) created social benefits in the past does not mean that a greater degree of private sector involvement in information provision in the knowledge society of today is synonymous with an abandonment of past ideals of social information provision.

In fact, private companies involved in the business of digital information provision have done some things - which in the past we have associated with the public domain - remarkably well. We in the LIS profession should not take a doctrinaire approach to commercial company involvement in 'our' information world. Rather we need to act as honest observers of progress. We must point out failures and deteriorations in evolving patterns of information service, offering harsh words where appropriate, but we must also work together openly and constructively with all parties, both public and private. In this way we can help create pragmatic, non-doctrinaire solutions that really do work for all who are citizens of our contemporary information society.

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

References

- Adobe (2007) < <http://www.adobe.com/aboutadobe/> > (accessed 2/6/2007)
- Australian Library Collections Task Force (2001) Access to Electronic Resources in Australia, Agenda item 9 of the Meeting held 9 February 2001
< <http://www.nla.gov.au/initiatives/alctf/meeting6/item9.html> >
(accessed 2/6/2007)
- BBC web site (2006). Google censors itself for China.
< <http://news.bbc.co.uk/1/hi/technology/4645596.stm> > (accessed 2/6/2007)
- The Internet Archive (2007)
< <http://www.archive.org/index.php> > (accessed 2/6/2007)
- Legal Deposit Libraries Act (2003). Ch. 28. London, HMSO.
< <http://www.opsi.gov.uk/acts/acts2003/20030028.htm> >
(accessed 2/6/07)
- Mills, E. (2006) Brewster Kahle's modest mission: Archiving everything. Story last modified Fri Jun 23 10:37:26 PDT 2006
< http://news.com.com/Brewster+Kahles+modest+mission+Archiving+everything/2100-1025_3-6087167.html > (accessed 2/6/2007)
- Muddiman, Dave (2000) Public libraries and social exclusion: the historical legacy, in Muddiman, Dave, Eds. Open to All? : the Public Library and Social Exclusion, chapter 2, Vol.3, pp. 16-25. London: Resource: The Council for Museums, Archives and Libraries.
Available via E-LIS, the open archive for Library and Information Science
< <http://eprints.rclis.org/archive/00005469/> > (accessed 2/6/2007)
- ScienceDirect (2007) ScienceDirect Info Backfiles - Content: Journals
< <http://www.info.sciencedirect.com/content/journals/backfiles/collections/> >
(accessed 2/6/2007)
- UKWAC, UK Web Archiving Consortium (2007)
< <http://www.webarchive.org.uk/> > (accessed 2/6/07)
- W3C web site and validation service (2007) < <http://validator.w3.org/> > (accessed 2/6/07)

Received 25/5/07, reviewed 26/5/07, revised 1/6/07, accepted 2/6/07.