Editorial: Digital library futures: collection development or collection preservation?
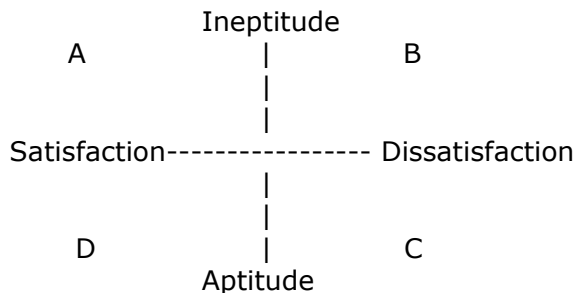
Abstract

Purpose of this paper	To argue that theoretical models from non-LIS disciplines can be of practical benefit to practitioner LIS research. In the area of digitisation collection development policy, such models highlight the importance of digital library preservation issues.
Design/methodology/approach	An application of formal models from cultural theories derived from structuralism and semiotics to LIS problems.
Findings	That theoretical models from non-LIS disciplines help illustrate and understand problems such as developing information literacy in the digital library environment or creating a balance between the need to develop new digital collections and preserving the digital collections which we have already created.
Research limitations/ Implications	This is a theoretical argument that could be tested by practical case study investigation.
Practical implications	Suggests that resourcing should be applied to digital preservation activity rather than a fresh round of digitisation of print originals, the preservation implications of which are uncertain.
What is original/value of the paper?	This paper gives some original perspectives on practical LIS challenges by applying abstract ideas from the area of cultural theory and applied linguistics.

Paper type: Viewpoint

Keywords: Digital libraries; preservation; theory.

One of the most widely read and well crafted pieces of LIS writing of recent years is the paper by Plutchak (1989), entitled 'On the satisfied yet inept user'. It is fair to say that there are good reasons for its popularity. On the one hand Plutchak can be used as an elegant and witty exposé of the dangers of letting the unsophisticated information seeker loose on "end user-friendly" information resources (which for us today means the open internet). Their thirst for knowledge may be quenched by misleading yet apparently convincing data, while traditional library tools (be they quality-controlled print collections or complex mediated online searches) offer better resources which are ignored due to the 'inept satisfaction' of the new internet generation.

But more than this, Plutchak's article is very nicely structured around a neat pair of simple oppositions: aptitude versus ineptitude, satisfaction versus dissatisfaction. When superimposed on each other, these two so-called 'binary antitheses' create a neat four-part matrix, which one can render thus:

```
                 Ineptitude
        A            |            B
                     |
                     |
    Satisfaction----------------- Dissatisfaction
                     |
                     |
        D            |            C
                 Aptitude
```

Fig. 1 The "satisfied inept" grid

What does this matrix illustrate? In the first place it shows that the traditional library user of the well managed print collection was offered a carefully filtered, evaluated information resource which presented a certain amount of intrinsic difficulty to use. Given this intrinsic but deeply educative difficulty, the inept user tended to reach a position of dissatisfied ineptitude fairly quickly in the print library world (Point B on the matrix above represents this state). Their well-intentioned but unskilled minds were easily defeated by the mysteries of traditionally classified library shelves.

Thus, the dissatisfied inept would turn to the reference librarian, who would either take the easy way out and give the reader what they wanted, creating a satisfied inept user with the right information (located at Point A on our graph). Or, with some deftly applied user education support, the librarian might succeed in creating an "apt" user with the skills to satisfy their present and future information needs. In such a case the library user will have navigated under the librarian's protection from dissatisfied ineptitude, to point D, satisfied "aptitude".
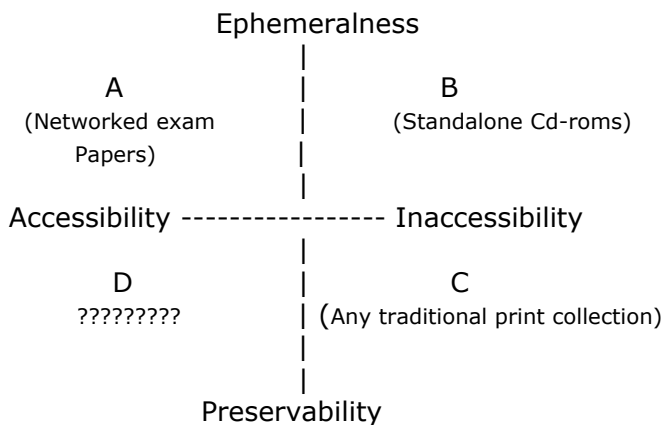
Today, the easy to access and easy to use world of WWW/Internet resources presents itself as authoritative, transparent and satisfactory, despite being full of unfiltered and misleading information. In consequence, the inept user is immediately lulled into satisfaction while remaining unaware of their ineptitude. The satisfied and inept move repeatedly on an axis of ignorance between Point B and Point C, repeatedly slaking their thirst for valid information by ingesting virtual rubbish. By contrast, the virtuous pattern of traditional library use has always been to move on a journey from dissatisfaction to satisfaction by means of the acquisition of knowledge.

It has to be said that the paper is a little patronising to readers and a bit smug about librarianship, but it is very neat and illustrative, which, together with the use of the satisfaction/aptitude grid I think accounts for the level of interest in Plutchak's paper. There are, after all, many other papers which both lament the impact of IT on library users while applauding aspects of traditional library use. None has achieved the popularity of Plutchak.

However, the strength of Plutchak's paper lies not just in its content, but in its combination of content with formal structure. A matrix of paired binary antitheses may on the one hand be a grim bit of jargon, but it is also a well-established intellectual tool the origins of which lie in attempts to understand how meaning originates in language. These semantic models were then adapted from linguistics and used by exotic continental European theorists such as structuralist Claude Levi-Strauss to understand the mythology of non-Western cultures. After that, the models were introduced to more pragmatic Anglo-Saxon readers some years ago by proselytisers such as Hawkes (1977), and more recently, at the end of a long line of interpreters, by Chandler (2001).

The structuralist vision of such paradigms is that they underlie any successful attempt at conveying meaning. Once realised, the use of these paradigms can inform any other attempt to illuminate the meaning of a 'cultural activity', thus conveying insight and enhancing understanding. It is true that post-structuralism and deconstruction have long ago challenged these structuralist notions as naïve and even 'repressive' to the 'play of writing', but let us put that aside. It is worth asking whether this type of paradigm can be applied in any useful sense to Library and Information Science.

So, to illustrate the value of this paradigm to LIS, Figure 2 shows another pair of binary antitheses, here applied to the issues of digital libraries and digital preservation. These can be mapped onto the same 'semiotic' grid as above, but using different values:

```
                  Ephemeralness
                       |
        A              |           B
  (Networked exam      |      (Standalone Cd-roms)
     Papers)           |
                       |
Accessibility ---------------- Inaccessibility
                       |
        D              |           C
    ?????????          | (Any traditional print collection)
                       |
                       |
                  Preservability
```

Fig. 2 The "accessible and well-preserved digital library" grid

Just as with information literacy, where there is pursuit of the goal of creating the 'satisfied apt', in digital preservation there is also a particular goal in mind. It is to achieve the 'accessible and yet well preserved' digital library (and by accessible I mean here not only the quality of 'being available to all users, disabled and

otherwise', but the quality of accessibility that distinguishes networked digital libraries –available any time, any place, anywhere - from less accessible print collections).

The similarity with the grid in Figure 1 is that it shows how easy it is to reach one of these goals, but not both: end user tools create satisfaction but not necessarily 'aptitude', and similarly, the digitisation of a collection undoubtedly makes that collection much more accessible, but not necessarily any more preservable. Indeed, just as the habits of virtual information usage can in fact make the user more 'inept' than traditional information usage, so can the accessibility of digital information make one blind to the difficulty of achieving the qualities of accessibility and preservability in the same collection (represented by Point D in Fig. 2).

The prime causes of problems in digital preservation are familiar: digital technological platforms are wonderful at making information more accessible but not so good at preserving that information for posterity. Whereas archive quality print paper can last unaltered for 500 years, a pdf-based institutional repository or a web-based set of html pages needs updating on a regular basis - for example, every time there is a new html standard and new browser software with limited backward compatibility.

However, many librarians will remember the 1980s with the noteworthy emergence of concerns over print preservation and (for example) the requirement to deacidify books made from low quality mass-produced paper printed after the explosion in Victorian publishing from the 1860's onwards. In my own case in 1984 I was working with a group of trainee librarians on placement in a Rare Books collection (the late Victorian parts of which had required no conservation work since first publication over a century before), and I remember our ill-informed comments: 'Don't worry, we won't have to do this once we're well established professionals. All these old books will be electronic, networked and permanent. No more brittle pages covered in dust to clean and deacidify.'

By contrast, I was recently supervising a member of library staff who was engaged full-time in upgrading our extensive and complex set of Library web pages from an older html standard to xhtml 1.0. Much of this can be done automatically, but managing the preservation of a set of html pages involves motivating someone to care about (for example) the difference between the superseded equivalent pair of html tags <br>, and
 </br>, and the new single xhtml version of this break tag <br />. Having upgraded our entire web site, our member of staff commented 'This is boring, like deacidifying a rare books collection but repeating it every ten years! What happens when we have to move to xhtml 1.1? And then 1.2? It's easier working with a print Special Collection.' There is no convincing answer to this.

At the everyday level of ordinary Library practice, this is the reality of digital library preservation. Of course, library web pages are rarely the same as a library collection – they generally serve as a portal to electronic library full text services (such as Emerald), and any text that they present will probably consist of library guides, i.e. descriptive texts helping users exploit full text collections, print or digital. They are something like metadata rather than primary text. However, as libraries digitise more and more material (either from print originals or by electronically warehousing 'born' digital institutional intellectual property), then more and more of a library's web site and server space will consist of full text electronic material, probably in either pdf or (x)html format. As libraries collect more and more digital full text, there

will be more and more library staff caring about digital preservation issues such as the difference between
 (old) and
 (new). We are creating a Sisyphean task for ourselves[1].

The perpetual models of structuralism could have helped us see this problem over the horizon. Advances in library technology can generally be described as a process whereby one format supersedes another. The justification for the change is the loss of a bad feature of the old format in exchange for the good feature of the new format. Print collections are inaccessible because they are static and un-networkable. Digital collections are better because they are dynamic, changeable, constantly updatable and available any time, any place, anywhere. Print is bad, digital is good. This is a single binary antithesis.

Semiotics and structuralism maintain that in any area of cultural activity, binary antitheses are not necessarily found in isolation. Any one binary antithesis can be twinned with another in quaternal grids. The trick is finding out what the second binary antithesis is which matches the first, so that you can complete the quaternal grid. Anyone can see history as the movement from the bad old to the new good. The clever thing is to see the down side of the new, and the good things that are lost in the old.

From the birth of the first online databases in the late 1960's, through to the creation of the first library systems in the 1970's and 1980's and the spread of commercial web browsers from 1994 onwards, the birth of digital library services has more often than not been portrayed as the progress from low levels of accessibility to higher levels of accessibility ('Don't visit the Library any more, information can now be accessed over your desk top'). Digital good, print bad. Few if any pointed out that the static, ponderous, and inaccessible nature of print collections meant that they tended to stay put, that is, they were easily preserved. Whereas the dynamic, virtual and accessible nature of digital collections means that they don't stay put, they are difficult to preserve.

To return to the practical implications of Figure 2, the accessible/preservable grid, scrutiny of the grid shows that three of the four points on the grid have a practical example of a library format. The logic of the grid tells us that we will find four types of bibliographic format in any hybrid library, A) accessible-unpreservable or ephemeral (networked exam papers)/ B) inaccessible-unpreservable or ephemeral (standalone cd-roms)/C) inaccessible-preservable (print collections in general)/D) accessible-preservable (the unachieved ideal digital library!).

The most valuable lesson of this grid is its illustration of the fact that, if we accept that digital collections are hard to preserve, the best collections to select for presentation in digital format are those which need to be accessed swiftly and temporarily by users, and then discarded. In a University context, the obvious material type that falls into this category is a collection of exam papers. Exam requirements change, older past papers become irrelevant, but the most recent few years of exam papers are vital to panic-stricken students who may need access from home at midnight a few days before their final exams. Digitise and network these, then delete them from your web servers as they become out of date. Preservation problems are deleted with them.

On the other hand, the grid also shows us that the tendency to digitise out of copyright heritage materials (because they present fewer problems of digital rights

management) is fraught with difficulty. Heritage materials tend to need to be preserved forever because they are by definition, 'heritage' materials. Digitising low profile, high value printed heritage materials does make them more 'accessible', does raise their profile. But the digitised versions are not to be thought of as 'the collection'. They are more like an ephemeral sort of document delivery format, available free of charge to all for a limited period. The collection remains the print masters.

Currently, JISC is performing a consultation exercise on how to spend four million pounds worth of money on digitising collections. Effectively, they are in pursuit of digital library collection development principles. One of the criteria for selection for digitisation is 'accessibility'. The implication is that there are worthwhile heritage print collections that are little known because they are inaccessible and deserve to be digitised, because they will then be more accessible.

There is a strong argument that we should not for now spend further amounts of money on digitising any more print collections. By contrast, we could consider investing the money and using the interest as a way of meeting the recurrent expenditure incurred by having to regularly update digital library collections from old html to new xhtml (among other ephemeral formats). There are a number of existing digital library collections that will need conversion from one format to another in the future, and we should invest in the future of these collections, not create new digital preservation problems by digitising afresh (unless we are creating digital ephemera such as exam papers).

Library and information science has thus probably been rather blinkered in its approach to digitisation as a way of generating accessible digital collections. One of the reasons for this may be its narrow intellectual base – rather than seeing itself as a broad subject with many features in common with other philosophies and intellectual disciplines, it tends to be limited in vision to certain pragmatic outcomes. However, when a writer such as Plutchak comes across a theoretical principle that allies LIS with broader intellectual horizons, the effect is startling. Apart from being rather clever, it is also very useful at the level of practice – the lessons of the two grids are simple and practical: a happy reader may not be a well-informed reader; don't start digitising until you know you can keep what you've digitised.

Of course, many such lessons could be learnt without recourse to theory. It is sobering to think that while many libraries aim increasingly to digitise collections (changing print to digital), most of global monograph publishing activity consists of the reverse process, i.e. taking author files delivered electronically to the publisher and printing out the text as hard copy for sale in book shops. This should give us a clue to the true nature of digital or non-digital library services: publishers tend to be good at knowing what customers want because they have to sell to them to survive.

However, to be truly useful as a practitioner discipline, LIS research needs to keep its intellectual vision broad and theoretically sophisticated. In this way LIS research can help us answer fundamental, practical professional questions, such as "whither digital libraries - collection development or collection preservation"?

Nicholas Joint
Editor,
'Library Review'

Note

1. Wikipedia entry on Sisyphus: "interminable activities are often described as Sisyphean." < http://en.wikipedia.org/wiki/Sisyphus > Accessed March 3rd 2006.

References

Chandler, D. (2001) Semiotics: the Basics. London: Routledge. Online version: Chandler, D. Semiotics for Beginners. Available at: < http://www.aber.ac.uk/media/Documents/S4B/sem05.html > Accessed March 3rd 2006.

Hawkes, T. (1977): Structuralism and Semiotics. London: Routledge

Plutchak TS. On the satisfied and inept end user. *Medical Reference Services Quarterly*. 1989 Spring; 8(1):45–8.