

Privileging Information is Inevitable

James Currall, Michael Moss and Susan Stuart

James Currall is a statistician who has always worked in multi-disciplinary environments. His formal job at the University of Glasgow is in the strategic planning of ICT. He is also a Senior Research Fellow in the Faculty of Arts and over the last few years he has collaborated in, led projects and written on: information value, security and trust and how the digital world impacts on teaching and the way that we find and use information.

Michael Moss is Professor of Archival Studies at the University of Glasgow. He writes and researches on historical and information issues. His study of army doctors in the French wars with Laurence Brockliss of the University of Oxford, *Advancing with the Army*, will be published by Oxford University Press in the autumn of 2006. He is a member of the board of the National Trust for Scotland.

Susan Stuart is a Senior Lecturer in Philosophy in the Faculty of Arts at the University of Glasgow. Her principal research interests include the phenomenology of consciousness, enactive cognition, Kant's epistemology, questions of ontology, and idealism and technology. She is Regional Director of the European branch of Computing and Philosophy (E-CAP) and the international representative on the American Philosophical Association committee for Philosophy and Computing.

Libraries, archives and museums have long collected physical materials and other artefacts. In so doing they have established formal or informal policies defining what they will (and will not) collect. We argue that these activities by their very nature privilege some information over others

and that the appraisal that underlies this privileging is itself socially constructed. We do not cast this in a postmodernist or negative light, but regard a clear understanding of it as fact and its consequences as crucial to understanding what collections are and what the implications are for the digital world. We will argue that in the digital world it is much easier for users to construct their own collections from a combination of resources, some privileged and curated by information professionals and some privileged by criteria that include the frequency with which other people link to and access them. We conclude that developing these ideas is an important part of placing the concept of a digital or hybrid paper/digital library on a firm foundation and that information professionals need to learn from each other, adopting elements of a variety of different approaches to describing and exposing information. A failure to do this will serve to push information professional towards the margins of the information seekers perspective.

Introduction

The digital world is forcing people to ask some fundamental questions about the nature of collections, as they struggle to come to terms with what a collection might be when the 'objects' could be distributed synchronously across the globe rather than stored in one place at one time. Here we discuss the nature of collections, with particular reference to the archival world in Currall's work.¹ In the traditional library world, a variety of authors characterise a collection as being: 'physically constrained to a single space and ordering',² 'the total sum of library materials that makes up the holdings of a particular library',³ or simply 'libraries [in the past] were a collection'.⁴ We will argue that, not only should we be asking searching questions about the nature of a collection in order to make sense of what a virtual or digital collection might be, but we should also be undertaking a thorough re-evaluation of what a collection is in the traditional world of physical objects; because unless we understand the nature of a collection, irrespective of medium, we cannot hope to develop a robust model that is capable of dealing with 'objects' of whatever form – physical, digital or abstract. Without one, we will argue, there will be a failure to capitalise on the important skills that traditional information professionals have. They in turn will become marginalised by many who will increasingly turn to technology as a means to manage and find information.

Information providers

Hur-Li Lee sees the nature and scope of a collection as being simply a function of the collectors' understanding of what the collection is and Lee draws attention to the fact that different professions see the term 'collection' as having subtly different meanings.⁵ In the library world he sees a collection as 'intermediated' by professionals and intended for a user community or communities. This is not far removed from the appraisal of records to form an archival collection, except that in the library context the collection usually represents aggregation whereas in the archival context it is a process of reduction. There is an interesting contrast here with Clifford Lynch, who is 'starting to believe that collections don't really have natural communities around them' and particularly in the digital world may 'find their own unexpected user communities' which we might characterise as epistemological.⁶ Such a perspective resonates with Francis Miksa's claim that in the digital world collections can 'be tailored to the individual library', or we might say archive,⁷ which he goes on to describe as a 'personal space' as opposed to the 'public space' of the traditional library. He sees such personal space as postmodern, the antithesis of the well-ordered library with its structured finding aids. This we would argue is a false dichotomy which derives from a narrow definition of library and a failure to grasp that some people organise their personal collections as carefully as those held in public space.

A recent editorial in the *Financial Times* conjured up a perversion of personal space where information providers (librarians, if you will) deliver personalised information based on preferences and patterns of previous use.⁸ This is an extension of the 'Daily Me', postulated by Negroponte.⁹ Differentiating himself from the authors in the previous two sentences, Lynch seeks to explain this new paradigm by stating that libraries (we might add archives and museums) and collections are not one and the same, with collections characterised as sets of raw material, and libraries, archives and museums as the mechanism by which such collections are usefully made accessible. Ross Atkinson goes further to propose that in future libraries will simply be 'switching centres', which could, of course, equally apply to a personal collection of bookmarks for websites.¹⁰ This begs the question regarding selection and of what stock forms the collection. Kennedy describes this as a 'high profile' element of librarianship but he might also have added (which Wernick also

considers) is the greatest professional challenge and the most important area of archival activity.¹¹ While David Levy would agree, he rightly draws attention to the persistent ambiguity between the notion of the library as an institution and as mere collections.¹² What all these commentators have in common is the claim, in the public information world at least, that a user community of some sort (intended or otherwise) is crucial but whether or not this is a property of the collection, the library, or some other entity, is much less clear.

Little significant progress can be made unless we are clear about what is meant by the term 'digital collection'. Levy is quick to point out that the notion of the library as an institution has been qualified in the digital world to mean 'institutions that oversee digital collections'.¹³ In the archival domain neither Cook nor Duranti get even this far, simply skirting round the problem, or refusing to admit that the digital does 'not alter the system in any way'.¹⁴ Lee develops this thinking claiming that since collections in the digital world can no longer rely on such concepts as: tangibility, physical collocation, format and ownership, traditional thinking about collections in the information world has a lot of work to do to transform itself.¹⁵ It seems trivially true then to say that it must take on a much broader definition of what a collection is. However, both Lee and Levy fail to indicate that this re-thinking is as necessary in the physical world as it is in the digital. Determining the members of any set is far from easy, especially at the borders where ambiguity and multiple set-membership is the norm. This is what troubles Miksa, who is convinced that in the postmodern world 'there is no *one best* (our emphasis) classification of knowledge system – that is, best in the sense of being accurate in any absolute sense'.¹⁶ But you do not have to be postmodernist to accept such a position, and its recognition should not prevent the use of classification schemes which will allow users to define their own collections.

As already noted, there is a difference here between the digital and the physical library, archive or museum. In the physical world an object can only reside in one place, whereas in the digital world this constraint does not exist as objects, and as we know, are stored arbitrarily.¹⁷ Users can allocate them to different sets by using a variety of discovery mechanisms from free text searching to the use of classification terms in supporting metadata. The key concepts that emerge from attempts to define collections of either kind, digital or physical are: selection

(appraisal) and intermediation, resource description and metadata, retrieval mechanisms, defined relationships, a user community and the management of resources. It is with these that we are concerned.

Collector and user

Lee and Miksa agree that there are, at least, two contrasting perspectives of what a collection might be.¹⁸ The developer perspective sees the collection in terms of selection and control, whilst the user perspective sees the collection in terms of resource discovery and access. In private collecting, these two perspectives are embodied in the one individual, the collector, but where collections are developed and maintained by one party for the benefit of others the roles of collector and user may become widely divergent. It might have been that in the past – and possibly even now – curators liked to form collections as an aid to efficient management of objects and this has serendipitously eased search and retrieval for users for a long time into the future. Geisler, concludes that:

Virtual collections encourage us to see a digital repository not as a unitary structure, but as a modular construction comprising many sets of resources, some small and others large, some separate and others overlapping, some stable and others transient, some defined by library managers and others established by library users.¹⁹

Users themselves have, in the past, employed catalogues and cross-references to construct their own 'collections' of information objects relating to topics of interest to them, and a good example might be the books and articles 'collected' to write this paper. The collection has a temporary fixity in that it is bounded by the writing of this article, but that is all. In this physical world, access to the collection is stored via 'bookmarks' in the form of a card index. Some of the objects are 'local' in that we have books, photocopies of articles and so on. Others are 'links' to articles and books stored in libraries, archives and museums. It is possible in the physical world, that links may become broken, in that items may no longer be available and might have to be sought somewhere other than their original location and occasionally it may not be possible to find the items at all. All the elements of this scenario have correspondences in the digital world, as indicated by the language used in describing them above.

As we have seen, the digital environment makes the construction of such transactional 'collections' of objects much easier. We do it unwittingly when we launch an Internet search which yields sets of hits mediated by search engine algorithms, that may or may not satisfy the intended subject of our enquiry. The search engine will, at most, index less than 50 percent of the material that is available on the Web and that proportion will not be a random selection geographically or culturally, as a result of intentional or unintentional aspects of the criteria used to decide what pages to index. The criteria used to match our query to candidate results, and then to order them by some relevance ranking, introduces further elements of privilege into the links that we actually follow as a result of the search. Introna and Nissenbaum discuss this phenomenon in the context of the politics of search engines and argue that there is bias in what we get from search engines, irrespective of whether or not there is deliberate intent in directing our attention to some sites as opposed to others.²⁰ Although the way such algorithms work is a commercial confidence, there is no doubt that they privilege information by, for example, ranking results by the popularity of sites or the number of links pointing to it.²¹

In some cases ranking can be improved by paying for the privilege. For these reasons there must be concerns that this sort of commercial mediation lacks objectivity.²² As the *Financial Times* put it, 'if commercial search engines one day write your shopping list for you or pick your news, you will never know how they made their choices'.²³ The editorial went on to quote the founders of Google, Sergey Brin and Larry Page, who have advocated that 'the world would always need at least one fully transparent search engine, preferably maintained in the disinterested academic realm'. But this is surely unrealistic, irrespective of intention, when commerce, culture, morality, politics or nationalism is involved, such 'objectivity' will always be a chimera. Instead we suggest the substitution of transparent protocols, such as controlled vocabularies that define naming conventions. These could be embedded in metadata, to be operated on by search engines to improve strike rates, but could also be used in other ways as better metadata has been a failure in the Web context.²⁴ We note that this, in itself is a form of privileging but note also that this cannot be avoided. Such utilities of their very nature should be transparent, but both in their construction and application,

are open to criticisms similar to those described by Introna and Nissenbaum above.

This sort of privileging is as much affected by demand as the commercial equivalent used by, for example, Amazon, the online booksellers. Controlled vocabularies, such as Library of Congress Subject Headings (LCSH), may claim objectivity but, in practice, the headings that are chosen for inclusion reflect popular usage. Thus, if most people in America call 'trousers' 'pants', then 'pants' it is. But it is also more than simply popularity for, amongst other things, it conforms to the political flavour of the day. As Sanford Berman, writing in the early seventies, states:

But in the realm of headings that deal with people and cultures – in short, with humanity – the LC list can only 'satisfy' parochial, jingoistic Europeans and north Americans, white-hued, at least nominally Christian (and preferably Protestant) in faith, comfortably situated in the middle- and higher-income brackets, largely domiciled in suburbia, fundamentally loyal to the Established Order, and heavily imbued with the transcendent, incomparable glory of Western civilization.²⁵

Moreover such headings must inevitably be dynamic, reflecting changes in public perception and in social conventions. Any attempt at standardisation of categories and usage assumes a homogeneity and stasis in the wider world that does not exist, is not possible, and is certainly not desirable. Endeavours to standardise terms now will look absurd to future generations for they will have different social conventions and political norms. Thus, it is with some surprise that we discover that Berman, even with his criticism of the LC list, never departs from his opening article of faith:

There can be no quarrel about the practical necessity for such a labor-saving, worry-reducing work, nor – abstractly – about its value as a global standardizing agent, a means for achieving some uniformity in an area that would otherwise be chaotic.²⁶

But we should not rush to adopt a nihilism towards sense and reference in the face of the inevitability of social construction, deconstruction and reconstruction, rather we should simply continue to remind ourselves that knowledge, and the way we handle and organise it, will always be

constructed and imbued with our own social, political and economic perspectives. Any search for global authorities is a chimera and, whether or not we accept this truth, the nonsenses that Berman indicates will simply multiply as the years progress. What is important is that authorities are themselves maintained as dynamic entities, making deletions, re-definitions and amplifications transparent. As we know, the more demand there is for information about a certain subject a greater degree of granularity can be expected from supporting controlled vocabularies as the quantity of resources rises to match demand. This can work in reverse and where demand subsides the granularity can disappear. Failure to record the process plays directly into the hands of a postmodern critique, as Bowker and Star assert, 'each strand of each category valorises some point of view and silences another'.²⁷ Cook makes much the same point when discussing the appraisal techniques employed by archivists in selecting records to form the components of collections:

The profession preaches the merits of accountability through good records to anyone who will listen; how accountable are archivists willing to be through keeping good records themselves about what *they* do and making these records readily available?²⁸

Piggott and McKemish go further and stress that record keepers and archivists select records and place them in a context that tells one particular story, but that different contexts would enable the same records to tell different stories.²⁹

Searching for information

Powerful search engines, such as Google and Yahoo, and the exposure of an increasing number of assets on the Web have arguably changed the relationship between users and information providers. Some would claim that this is a paradigm shift. When a search is executed the underlying algorithm delivers a set of results that logically satisfies the query arranged in order of relevance. Although logically constructed only some of the contents of the set or collection satisfy the intention of the enquirer, who selects those that are relevant to refine the set or collection and abandons the rest. This process is not far removed from what happens in the physical world, but the transactions take place much more rapidly than before and the potential for resource discovery is far greater and less

privileged. It is easy for information professionals to mistake an acceleration in the process for a paradigm shift, while overlooking the real change in the relationship between themselves and constituencies of users. Because of the very nature of the Web, relevant hits will include assets created for a whole variety of constituencies. School children can access resources designed for the scholarly community and vice versa. There is nothing to distinguish different types of asset, a publication from a manuscript or a museum object. In a physical library such assets would be segregated and defined by the space in which they are presented to the user, the children's room, the archive, the reference library and so on. Moreover, resources can be accessed across the Web which are created by providers whose products will rarely, if ever, be mediated by information professionals and is often intentionally the case.

This perception is reinforced by subsequent experience of revisiting resources held in a set. In the physical world there is an expectation that reference material (the books and manuscripts) will be available in the future. There may be different copies, because some books have been discarded by a particular library or repository and new books and manuscripts accessioned. Since a book has been published, declared in a formal process, the user and the discarding librarian can be fairly confident that missing items will be replicated elsewhere, for example in a copyright library. There can be no such confidence about resources discovered on the Web. Some may be unavailable because a server is temporarily down or because the algorithm has been changed or a site no longer ranks as highly. Some may disappear completely because the site has either been permanently turned off or because the resource has been deleted or archived. In some cases, it will be possible to discover missing objects in the collections of other institutions and people, but there can be no guarantee.

Declaring an object on the Web is not equivalent to the process involved in publishing a book or the transfer of an archive to a repository to guarantee permanence. In a sense, this has always been the case with information. What is different is that the Web enables anyone, who so wishes, to become with little effort both a user and supplier with none of the advantages and disadvantages of the privileging and mediation of formal information professionals, who include publishers. This is not to

say there is no longer any privileging, there is, as Brin and Page readily admit, because of the very way in which the algorithms in individual search engines work.³⁰ This can be demonstrated easily by executing the same search using different providers with often very different results. Although there is an analogy here with the different collecting and cataloguing policies of individual libraries and archives, many of the objects collected from the Web would never have come within their scope or certainly in the case of archives within a time frame that would ensure survival.

Collecting

When Mr Lovel first enters the retreat of Mr Jonathan Oldbuck, in Walter Scott's *The Antiquary*, what he sees is chaos and confusion or what we might call, in a digital environment, the arbitrary allocation of resources:

A large old-fashioned oaken table was covered with a profusion of papers, parchments, books, and nondescript trinkets and gewgaws, which seemed to have little to recommend them, besides rust and the antiquity which it indicates. In the midst of this wreck of ancient books and utensils, with a gravity equal to Marius among the ruins of Carthage, sat a large black cat, which to a superstitious eye, might have presented the *genius loci*, the tutelal demon of the apartment.³¹

When Oldbuck begins to describe the objects in his 'curious collection', Lovel comes to understand that there is a perverse logic to it:

Here were editions esteemed as being the first, and there stood those scarcely less regarded as being the last and best; here was a book valued because it had the author's final improvements, and there another which (strange to tell!!!) was in request because it had them not. One was precious because it was a folio, another because it was a duodecimo; some because they were tall, some because they were short; the merit of this lay in the title-page - of that in the arrangement of the letters in the word *Finis*.³²

Order and reason can be unearthed (with little effort) in what might initially appear to be chaotic. Anything that is designated a collection must have a mind of some sort at work; choices are made, either in the

decision to designate or in the decisions that have gone into bringing this particular set of things together rather than any another. Some intelligent design is evident and it is such that it necessitates the privileging, and thus retention, of some object or information over others.

In his fascinating study of the contemporary American map thief Gilbert Bland, Miles Harvey interviewed a serious collector who claims that 'the key thing is what the piece of paper represents. So if you don't know the historical and cultural elements that produced a map, I think you're missing most of the fun'.³³ He goes on to say that 'selecting a piece for a collection has nothing to do with the individual merits of the item. It's what builds a collection, the sum is of greater interest than each of the individual pieces'.³⁴ This may be a little extreme but it gets at what Sarah Tyache has called, in the case of archives, the 'recordness of the record' or in the case of collections their phenomenal characteristics that bind the individual items together.³⁵ These characteristics are mediated by the preferences of the collector whether it be an antiquarian such as Oldbuck, a thief such as Gilbert Bland, or an institution such as a library or an archive, and few would deny that the act of mediation privileges the contents.

Selection, rules and membership

Lagoze and Fielding see 'a collection as logically defined in a set of criteria for the selection of resources from the broader information space' and go on to suggest that in the digital world this process, once the criteria are established (and made explicit), could be carried out automatically, so long as appropriate 'standardised' metadata is available for the objects.³⁶ As Lee points out, this model relies, for successful implementation, on objects having complete standardised metadata available. However, the terms that are entered in the various metadata attributes are subject, as we have already argued, to social construction and cultural interpretation. All objects, moreover, that do not possess the correct metadata attributes, in the appropriate form will be automatically excluded from such a collection, irrespective of what characteristics they possess. On the other hand an object that has the required metadata entered in error will, quite erroneously, be taken to be a member of the collection.

Although Lee sees automatic processes taking over some of the work of collection development, he believes subjective elements such as the quality

of the object will continue to need human intervention.³⁷ The issue of the socially constructed nature of metadata is an obstacle that has not been given the attention that it deserves and needs to be clearly in the frame when such automation processes are being devised, as its consequences may be felt 'at some distance' from the time, place or context of the original automation. So, what, if any, are the rules that govern the subjective elements of collecting, or is collecting, even in the professional library world, subject to the moods and fancies of the collection developer or budget holder?

As a start we might say that a collection is defined by a set of rules, explicit or implicit. The rules may be formally defined in collection policies or remain unspecified for others (users of public collections or later discoverers of private collections) to infer from what the collection contains. Some of these rules may be simply unarticulated, but justified preferences as in Oldbuck's case. Even in instances where the rules are explicit it is unlikely that they are entirely static, changing with circumstances or perhaps with the whim of the collector. Rules may be based on well-defined 'strong' attributes but there are also likely to be more poorly-defined factors with varying degrees of specification, particularly where they are assembled in Miksa's personal space.

Just as with controlled vocabularies, an important part of the description of a collection should be the rule-base for membership and a record, as Cook argues, of how this changes through time will be of fundamental importance to future users of the material. For example, if we consider that at any time a collection only represents part of what the collector is attempting to achieve, then collections are rarely complete. If at each point at which collection policy changed, the collection was complete, it might be possible to infer what the policy had been, but to attempt to infer the policy of a collection that only has some of the intended members present sets an intractable problem for the user. The degree of incompleteness of the collecting task and the regions of the set that represents the complete collection that are empty will depend on a range of factors, which cannot be known to the user, such as available finance, the knowledge and expertise of the collector and availability of suitable items. Additionally (or perhaps subtractively), items may be lost from the collection for reasons that are beyond the control of the collector, such as fire, flood, burglary or more probably unreturned loans and these affect

what is actually in the collection, but are not themselves part of collection policy.

In the past, collection policies have been set by institutions such as libraries and museums, and have been implemented by curators or collection developers who have adopted the mediation role of the individual collector. What seems to worry authors such as Lee is the fact that users may get used to accessing information that has not been given this treatment explicitly, an idea that he calls 'dis-intermediation'. If the idea that collections have a strong user dimension has any credence, why should those rules not be determined heuristically by user behaviour in much the same way that Amazon dynamically offers its customers 'collections' based on the behaviour of other users when it attempts to draw you into a further purchase by saying: 'Other people who bought this book also bought these ones'. This is already happening in many public libraries where shelves are packed with books that it is predicted customers want to borrow rather than with books that might be thought to be 'good' for them.³⁸ This is undoubtedly a level of dis-intermediation that would give Lee further cause for concern, but it has a strong connection with the way that we make choices in other areas of our lives for example in where we go on holiday or the sort of food we eat.

This raises the important moral question, which concerns the editor of the *Financial Times*, of the role of authority in instrumentalising society through the mediation of information,³⁹ whether it be the State itself acting explicitly or the State acting implicitly through its agents or trans-national corporations such as Google and Yahoo. Society is divided on this issue between those who believe that there should be no mediation at all and those who want greater restrictions, such as popular assertions that children should not be exposed to advertisements for fatty foods and pornography. Bowker and Star are right to warn that 'algorithms for codification do not resolve moral questions'.⁴⁰ Miksa is concerned that, from such a perspective, classification schemes can be condemned 'as not so much exercises in asserting what is in the world as they are exercises in 'losing information' so as to build a construction of reality'.⁴¹

We could argue that this is not really dis-intermediation at all, rather, a shifting of the boundaries of privilege. In the case of the Amazon, purchase recommendations, the knowledge, experience, prejudices and explicit rules employed by the information professional are replaced by users 'voting with their feet', by fashion or the will of the majority. What

Amazon is doing is offering another mediated navigation route through its stock, by adopting another set of criteria for defining a collection, which is no less subjective than the decision by a librarian to allocate a title to a specific place in a classification scheme. In the case of Google searches, the ordering of the hits is determined algorithmically by the number of other pages pointing at the page in question, resulting from the activities of other webpage creators/maintainers, and the frequency with which the sites are accessed via Google searches, resulting from the activities of Web users. For all the claims that Google exploits the 'uniquely democratic nature' of the Web in both these examples, the information that users find most easily is just as much mediated – perhaps even to a greater degree – as that in physical libraries or museums.⁴² There is simply a different set of 'selection' mechanisms at work which are determined to a greater or lesser extent by subjective rules. The only way to solve the problem of lack of objectivity that concerns the editor of the *Financial Times* may be to look to the market to generate an increasing number of alternative mechanisms for set allocation and leave it to users to adopt the mechanism (and thus rule set) that gives them what they want. This allows them to satisfy their prejudices and cultural outlook. The key point here is the information seekers will have a much greater choice in the new order than hitherto.

Boundedness

An issue that troubles a number of authors⁴³ is the nature of the boundaries of a digital collection. The problem can be pithily described as follows: if one digital object *A* references another *B* through a hyperlink and *A* is part of a collection, does *B* then have any status as part of that collection? This is much less troublesome in the physical world where one object may make reference to another but not provide direct access to it. This exercise is both a logical question of the extension of set membership and is a legal one. For example, in the case of objects made available as part of a collection over which the collector has legitimate authority and is yet 'recommending' other referenced objects over which they have no legitimate authority. Determination of set membership here is considered in terms of content and management – and therefore longevity – of the referenced object. Whilst these very practical aspects of the issue are discussed by these authors, the fundamentals of the boundedness of digital collections and the objects contained therein

receive scant attention. We might express this problem in the following ways:

- If I have a circle of friends, what status do their other friends have in relation to me?
- If I have a collection of letters from the Bishop of Bath and Wells to Reginald Smythe, do the letters from Reginald Smythe to the Bishop of Bath and Wells form a part of the same collection?

Hypertext is an obvious way in which the bounds of digital objects become eroded, however it is not a concept that originates in the digital world. Wittgenstein's *Tractatus Logico Philosophicus*, written in the trenches and first published in 1922, uses a nesting of levels of argument that provide a linking pattern that can be traced by the reader.⁴⁴ Footnotes operate in a similar manner. Whilst neither of these examples link beyond the bounds of the individual work, citation contained either within the main body or via a footnote does. The lack of immediacy in being unable to follow the references cited in a paper document that are not immediately to hand, may introduce a disjunction in the reader's experience of the material, but this is replicated in broken links, servers unavailable, and the other imperfections of the digital world. A move to URNs (Uniform Resource Names), document handles, and name resolvers will improve the situation, but we see the digital world as a fluid development of the paper world rather than a step change as suggested by Lynch.⁴⁵ It is worth noting that via citations and link, one information object privileges others and a chain of such links establishes a 'collection' that is not independent of the starting point. This is true in information seeking in both the physical and digital worlds.

Lynch takes the problems of boundedness in the digital world further with the concept of 'objects talking to each other'.⁴⁶ This is based on objects having rich markup of their intellectual and semantic structure and this structure being available to enable automatic linking between objects, for example, linking place names in one object to appropriate map objects.

The boundedness issue, when whittled down, is much more fundamental than sustainability or legal liability, but it isn't new; we see elements of the problem in the dis-aggregation of books into chapters, pages and quoted extracts, in references to other works (which may no longer be

readily available)⁴⁷ and in the separation of a document (whether written, photographic or artefactual) from its context. The digital world has potential to blur the boundaries of objects further as it is so easy to reuse, dissect, aggregate and transform objects.⁴⁸

Intermediation and value

Lee is troubled by the possibility that the digital world permits a much greater degree of dis-intermediation than is possible in a physical library or archive.⁴⁹ As an individual I may get access to information without the need to choose items that have been carefully selected for me by information professionals but as anyone who has made stuttering attempts to find information with Web search engines can confirm, there is often more to finding resources than simply typing a few keywords into Google. If, as it seems, we have all become much more explicitly 'private' collectors of information (selecting from a mix of material some of which has not been professionally intermediated), the very act of bookmarking sites is a form of mediation, as information seekers we will have to discover a new set of skills, perhaps shifting our pedagogical emphasis away from the didactic back to the German ideal of *lehre und bildung*.⁵⁰ This raises a whole set of other issues about the acquisition of the necessary critical skills to mediate our own information which cannot be dealt with in depth here.

Lynch gets to the heart of a much broader problem related to intermediation: the packaging of 'raw materials' in a variety of ways such as learning experiences, curated exhibitions or interpretations.⁵¹ He shows that, in its very nature, such packaging can rarely be interpretation-neutral, is socially constructed and also relatively short-lived. We only have to think of school history textbooks which change depending on whose perspective is fashionable, or the world globes with most of the land mass coloured pink to represent the British Empire that appeared in every school in the United Kingdom until their hurried withdrawal in the 1970s.

What troubles Lynch is the problem of sustainability rather than the privileging discussed by Buckland.⁵² Lynch is excited by the prospect of an aggregation of digital materials as being more than the sum of its parts; but this is not a new concept as Harvey's collector and, of course, the whole of Gestalt psychology are concerned.⁵³ In addition he sees

computations across objects as leading to them being more than the sum of their parts. We contend that this happens in the paper world, in that cataloguing is a 'computation' across multiple objects (whether paper, digital or hybrid) and that 'computation' results in a greater value than the sum of the individual objects. Even if Lynch is right, that there is something special over and above the sum of the parts in the digital world, how, if it all, does it differ from a collection in the physical world?

Two of the examples that Lynch gives link very different types of resources or 'mine' information in new and, as he sees it, exciting ways, whilst the third involves the resources 'communicating' with each other, and to external bodies and organisations. The implication is that there is great potential for creating collections in the digital world on the fly by adopting 'data-mining' techniques. Lynch's example is of astro-physicists downloading observational data held in digital libraries to create their own virtual sky databases. But another good example would be that, as more library catalogues go online (as online public access catalogues or OPACs) and as the granularity of cataloguing is improved, it is possible to discover more texts attributed to an author and for the first time to attribute a collection of texts to a given publisher. Although the digital facilitates such transactions, it is arguably doing nothing more than a card index would do, just a great deal faster and much more cost-effectively. Even where resulting 'collections' contain digitised content in the humanities, at least, this will invariably need to be supplemented by physical content, making it hard to argue for some special Gestalt conclusion in the digital domain. However, Lynch does have a case in regard to digital collections where, as he says, they are often unconstrained by copyright. It is straightforward to repurpose content, notoriously in the United Kingdom in the so-called 'dodgy' dossier on the case for war against Iraq.⁵⁴

Lynch regards collection level descriptions and other finding aids to be tools for management rather than access. Since there seems to be a very fine line between privileging and censorship there is some truth in this assertion, and there is certainly a need for a user perspective evaluation of the large sums invested in the United Kingdom in collection level descriptions by the Research Libraries Support Programme, and much will depend on what the majority – the largest market – of users want.⁵⁵ If, as seems likely, they are after precise bits of information, letters written by X or images of Y, there is every reason to doubt the utility of collection

level descriptions unless they are metaphors for curators saying 'we would like to catalogue all these objects in greater depth but do not have the resources'. In the world of books this is rarely the case; librarians have always started with the individual bounded object, the books, pamphlets and even the individual sheets, rather than their aggregation in collections. Archives and museums differ, but in a perfect world with unlimited resource, they would almost certainly have adopted similar strategies. Archivists make a great deal of the integrity of a collection, even when it has been artificially constructed, but they need to reflect on whether this is intellectually justified and ask if there are other ways of unambiguously identifying ownership and provenance, for example, by adding robust metadata to a document at the time of creation in much the same way as is already done in a printed book. Museum curators are more candid and willingly confess there is no point in cataloguing a bottle full of flies collected in the same place on the same day, in which case the bottle remains, arguably, an object and not a collection of flies with a glass boundary. In Lynch's world, bits of data would reside as 'granules' which could be assembled at will into virtual collections. The collections would not cease to exist as management tools, they would just be accessed at a granular level.

Lynch is excited by the potential for the granules to interact with one another by employing 'really good deep mark-up'.⁵⁶ While not wishing to pour cold water on his vision, this looks very much like trading the advantages in speed and power of the Web, and its viewing devices, for labour intensive handicrafts. Marking-up texts is, of its very nature, time-consuming and rarely can be automated because cataloguers, however hard they try, are not standard and cross-cultural consistency is impossible. There will be a place for creating such expensive resources, in just the same way that there is a place for publishing expensive critical editions of texts and reference taxonomies in the physical world. There are dangers in postulating a Web entirely populated with such texts because the mark-up will, inevitably, involve strong intermediation and, it is just conceivable that the mediators might be, at worst, wrong and, at best, not impartial.⁵⁷ The Web would become deterministic rather than probabilistic and lose part of its attraction. Perhaps a logical consequence of this would be that the desire of information providers for 'order' is misplaced and that the randomness of Mr Oldbuck's oak table, cat and all, is what we should strive towards.

Crossing boundaries

A collection in the library world is built up by aggregation of individual objects, separately described. In the museum and archive world, a collection is a set of related objects which are frequently not individually described. Information in the digital world tends to be individual objects, which are identified, not through being in a collection, but by a search that yields a set of items that logically satisfy the criteria used for the search, subsequently refined by user appraisal. This is true whether the search is on a library/archive/museum catalogue of 'selected' material or on the indices of a search engine. The result may be given some fixity through book-marking or saving the search. There are elements of this process that have more in common with the archives 'top-down' collection building than the library 'bottom-up' approach. On the other hand, the building takes place on individual objects as in the library world rather than on series or aggregate objects as in the archive world. The disaggregation is however more marked than in even the library world with chapters or sections of a work frequently having a separate existence in a way that they do not, for instance, in a book. Additionally there is often little information that indicates the relationship between objects that naturally go together, even those such as chapters of a book or illustrations of what is described in text. Librarians, archivists and museum curators need to understand how others manage information and its description and cataloguing if they are to be able to deal effectively with elusive digital objects in ways that will satisfy the way that information seekers go about building up their 'private' or personal collections. This is part of a wider agenda where, increasingly, experts need to work across the boundaries of their field and learn to operate in collaboration with experts in other fields. This results from a reduction in the discrete nature of disciplines, the trends towards trans-disciplinarity and a need for experts to be accountable in their exercise of expertise in a broader social context than has previously been the case. This is discussed in a set of articles by Strathearn and perhaps more succinctly (but specifically in the context of scientific expertise) by Nowotny.⁵⁸

The information professions should also be mindful of the fact that they are not the only route to high quality information and, increasingly, the information that individual information providers make available and

information seekers look for and find, may never be either 'published' or selected for preservation by information professionals, or ever come on to their radar. There is an increasing quantity of important information about a range of subjects and the state of the world that is available via the websites of private individuals and organisations that will never come into the custody of librarians, archivists or museum collections. The metadata describing these resources is not something carefully selected by a thoughtful process of privileging, but simply the full text of the information indexed by the likes of Google.⁵⁹ The Web offers new opportunities for custodians, to expose not just a carefully selected and described set of information objects via a carefully presented portal, but their entire catalogues; to be searched along with the rest of the Web, rather than in a separate space. The Online Computer Library Center (OCLC) and major research libraries are already grasping this possibility by working with Google rather than reinventing the wheel or fighting against it. These institutions will assist in presenting the information seeker with a far richer set of material from which to establish their private collections. If custodians refuse to engage with this agenda, they will be presiding over the marginalisation of their professions and by extension the resources they are responsible for.

Conclusion

Resolution of the issues we have reviewed in this paper is vital if the digital environment is to be exploited to its full potential, but there are no simple solutions. We should perhaps not limit ourselves to looking for solutions, but look instead for realistic processes that lead us away from the problems towards a new environment where the current problems are rather less of an issue. In attempting to do this, progress will be hampered if the various stakeholders in information provision in both the physical and digital domains fail to enter into meaningful dialogue, not just to quibble about semantics but to debate the harder theoretical, technical and philosophical problems that we have raised and attempted to address. This presents new opportunities to us all, but threatens the carefully cherished boundaries between professions in the established order. The value of the experience and perspectives of librarians, records managers, archivists, statisticians, accountants, information technologists, and so on is considerably more than simply the sum of its

parts, but are valuable only if the different groups do not seek reinvention of (possibly square) wheels.

The authors of this article come from the diverse backgrounds of archives/history, philosophy and statistics with a common interest in the digital. We have concluded that information, that which has been chosen to be conveyed, is privileged and socially constructed, and not something objectively determined by a set of easily articulated criteria. Consequently privileging is inevitably dynamic, reflecting contemporary circumstances and preoccupations. Although we don't see this as negative *per se* there are moral and political implications that must not be disregarded. This raises serious questions about mechanisms for resource discovery and assembly, leading us to argue for diversity in devices, to caution against monolithic control vocabularies, and to urge for openness in the criteria employed in privileging information; all of this is possible but we must progress beyond humdrum defensive collecting policies. We see the new information landscape as presenting information professions with new opportunities in relation to information availability and access, although it is clear that we will all have to be prepared to embrace new ways of working. This will allow people from different cultures and perspectives to find and 'collect' information assemblages that are relevant and useful to them. Dis-intermediation might not give us sleepless nights, but we must be concerned about the loose use of heuristics⁶⁰ by information providers to populate 'private' collections.

Endnotes

1 JEP Currall, SM Moss & S Stuart, 'What is a collection?', *Archivaria*, 58, 2005, pp. 131-46.

2 G Geisler, S Giersch, D McArthur & M McClelland, 'Creating virtual collections in digital libraries: benefits and implementation issues', *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, 2002, pp. 210-18.

3 A Kent & H Lancour, *Encyclopedia of library and information science*. New York, Marcel Dekker, 1971.

4 H-L Lee, 'What is a Collection?', *Journal of the American Society for Information Science*, 51 (12), 2000.

5 Lee, 2000.

6 C Lynch, 'Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Information', *First Monday*, 7(5), 2002.

7 FL Miksa, *The DDC, the universe of knowledge, and the post-modern library*, Albany, 1998, p. 84.

8 *Financial Times*, 'Online searching who's feeling lucky?', London (21-22 February 2004), p. 10.

9 N Negroponte, *Being Digital*, Boston., 1995, p. 164.

10 R Atkinson, 'Text mutability and collection administration', *Library Acquisitions: Practice and Theory*, 14, 1990, pp. 355-8. See also G Benoit, 'Toward a Critical Theoretic Perspective in Information Systems', *Library Quarterly*, 72(4), 2002, pp. 441-71.

11 J Kennedy, 2002, *Collection Management, A concise introduction*, Melbourne, and also T Cook, 'Archival Appraisal and Collection: Issues, Challenges, New Approaches', Special Lecture Series, University of Maryland and to NARA Staff, 1999, electronic text available at: <<http://www.mybestdocs.com/cook-nara-990421-2.htm>>.

12 D Levy, *Scrolling Forward: Making Sense of Documents in the Digital Age*, Arcade Publishing, 2001, p. 135.

13 Levy, 2001, p. 135.

14 Cook 1999, L Duranti, 2000, 'The Impact of Technological Change on Archival Theory', p.7, electronic text available at: <http://www.interpares.org/documents/ld_sevilla_2000.pdf>.

15 Lee, 2000.

16 Miksa, 1998, p. 86.

17 There are examples of physical libraries where volumes are stored by both size and subject to save shelf space, for instance in the stacks of Oxford's Bodleian Library.

18 Lee, 2000, Miksa, 1998.

19 Geisler et al, 2002, p. 217.

20 LD Introna & H Nissenbaum, 'Shaping the Web: Why the Politics of Search Engines Matters', *The Information Society*, 16, 2000.

21 An outline of the Google algorithm as first described by Brin and Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 1998, which is available electronically at: <<http://www-db.stanford.edu/~backrub/google.html>>, although many aspects of the algorithm have subsequently been modified.

22 That is not to say that this subjectivity is in any way analogous to other human selection mechanisms.

23 *Financial Times*, 2004, p. 10.

24 See for example: C Doctorow. 'Metacrap: Putting the torch to seven strawmen of the meta-utopia', 2001, electronic text available at: <<http://www.well.com/~doctorow/metacrap.htm>>.

25 S Berman, *Prejudices and antipathies: a tract on the LC subject heads concerning people*, Metuchen, NJ, Scarecrow Press, 1971, p.ix.

26 *ibid.*

27 G Bowker & SL Star, *Sorting things out: classification and its consequences*, Cambridge, Mass, MIT Press, 1999, p.5.

28 T Cook, 'Postmodernism and the Practice of Archives', *Archivaria*, 51, 2001, p. 35.

29 M Piggott & S McKemmish, 'Recordkeeping, Reconciliation and Political Reality', Australian Society of Archivists Annual Conference, Sydney, 2002, electronic text available at: <<http://www.sims.monash.edu.au/research/rcrg/publications/piggottmckemmish2002.pdf>>.

30 See Brin and Page, 1998, full reference at note 21.

31 Scott, W. *The Antiquary*, chapter 3, electronic text available at: <<http://www2.arts.gla.ac.uk/SESL/STELLA/STARN/prose/WSCOTT/ANTIQUAR/contents.htm>>.

32 *ibid.*

33 M Harvey, *The Island of Lost Maps - A True Story of Cartographic Crime*, London, Weidenfeld & Nicholson, 2001, p. 252.

34 Harvey, 2001, p. 247.

35 S Tyache, 'Archives in a wider world: the culture and politics of archives', *Archivaria* 52, Fall 2001, *The Association of Canadian Archivists*, p. 2.

36 C Lagoze & D Fielding, 'Defining Collections in Distributed Digital Libraries', *D-Lib Magazine*, Corporation for National Research Initiatives, 1998.

37 Lee, 2000.

38 The University of Swansea has decided to close down its Departments of Philosophy, Sociology and Anthropology to channel more money into what 'the students want' despite protests from current students and staff. 'The benefits to students are the highest quality, up-to-date, career-enhancing courses,' said the Principal Professor Davies. 'All universities have from time to time to adjust their courses to changing patterns of student demand'. <http://news.bbc.co.uk/1/hi/wales/south_west/3556865.stm> accessed 23 March, 2004, 08:03 GMT.

39 Searching the IFLA website for references to 'privileging' produces a large number of references for 'censorship' and we find that it is used metaphorically, exactly as 'friendly fire' and 'downsizing' are, to remove any unpleasant connotation of there being constraints on our freedom and our

choices being determined for us by a body who 'know better' than us what is in our interests. See for instance: IFLA Conference in Glasgow, August 2002 IFLA/FAIFE, *Glasgow Declaration on Libraries, Information Services and Intellectual Freedom and Censorship*.

40 Bowker and Star 1999, p. 24.

41 Miksa 1998, p. 87.

42 See <<http://america.google.com/technology/index.html>> for a thumbnail sketch of how the search engine in question manages this problem.

43 Lagoze & Fielding 1998; Lee, 2000.

44 One example of the numerous discussions of Wittgenstein's text as a hypertext can be found in L Wittgenstein, 1922, *Tractatus Logico Philosophicus*, Introduction by Bertrand Russell, London: Kegan Paul, Trench, Trubner or electronically at <http://www2.uiuc.edu/unit/reec/wittgenstein/intent_Tractuatus.htm> accessed 23 July 2004 at 13.20 GMT.

45 Lynch, 2002.

46 Lynch, 2002.

47 A foretaste of the now commonplace 404 error produced by web servers when the page requested is no longer at the location specified.

48 For a more detailed discussion of this area see, for instance: A Allison, JEP Currall, MS Moss and S Stuart. 'Digital Identity Matters'. *Journal of the American Society for Information Science and Technology*, 2005, 56(4).

49 Lee, 2000.

50 Literally 'teaching and formation'. As expounded by Wilhelm von Humboldt (1767-1835), emphasising the importance of formation of the individual rather than the role of the teacher. See for instance Marianna Wertz's article in *The American Almanac* available at <http://members.tripod.com/~american_almanac/humboldt.htm> (accessed 17 June 2005 at 9:10 GMT) or Christoph Wulf's paper 'Perfecting the Individual: Wilhelm von Humboldt's concept of anthropology, *Bildung* and mimesis' in *Educational Philosophy and Theory*, 35(2), 241-249, April 2003, (accessed at <<http://www.blackwell-synergy.com/doi/full/10.1111/1469-5812.t01-1-00022>> on 17 June 2005 at 9:10 GMT). Humboldt understood *Bildung* to a large degree as mimetic, that is, as non-teleological, undetermined and uncertain. *Bildung* is aimed at the reconciliation between outer historico-social and inner individual conditions. The success of this process requires individual freedom and a variety of socially created education opportunities.

51 Lynch, 2002.

52 M Buckland, 'What will collecting developers do?' *Information Technology and Libraries*, 14 (3), 1995.

53 See, for example, FS Perls, *Ego, hunger and aggression: the beginning of Gestalt therapy*, New York, Vintage Books, 1969.

54 'The Dodgy Dossier' was a briefing paper issued to journalists by the British Prime Minister's press secretary, Alastair Campbell, on 3 February 2003 about Iraqi production and use of weapons of mass destruction. The paper, entitled *Iraq - its infrastructure of concealment, deception and intimidation* was a follow-up to the previously issued *September Dossier*, intended to make a persuasive argument for the decision to go to war against Iraq. The term 'Dodgy Dossier' was coined by journalists after Channel 4 News learnt that much of the work had been plagiarised from various uncredited sources, most notably from a postgraduate thesis published on the Internet. See <http://en.wikipedia.org/wiki/Dodgy_Dossier> accessed 26 March 2004.

55 A Powell, M Heaney, and L Dempsey, RSLP Collection Description. *D-Lib Magazine* 6 No. 9, 2000, online: <<http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/september00/powell/09powell.html>>.

56 Lynch, 2002.

57 M Buckland, 'Five Grand Challenges for Library Research', *Library Trends*, vol. 51 (4), 2003.

58 AM Strathern, 'Commons and Borderlands: Working Papers on Interdisciplinarity. Accountability and the Flow of Knowledge', Sean Kingston Publishing, 2004; H Nowotny, 2003, 'Dilemma of expertise: democratising expertise and socially robust knowledge', *Science and Public Policy*, 30(3).

59 In order to illustrate the changed world of information, we have tended to contrast the controlled world of traditional libraries and librarianship with the 'wild west' of arbitrary pages on the World Wide Web, as the ends of a spectrum. There are however many areas accessible through the Web where there is information that has been carefully selected and prepared and where useful metadata and better management practices give the user more confidence that it is what it says it is and gives the information itself a better chance of long-term survival - an example is the Glasgow Story, available at: <<http://www.theGlasgowStory.com>>. Communities of interest often have well organised websites - an example is the Comprehensive R Archive Network (CRAN) which supports a cooperative project to develop statistical software, where mirror copies of the site are hosted in many parts of the world, available at: <<http://cran.r-project.org/>>. These examples are only two of a vast array of collections of information on the spectrum of carefully catalogued and described through to completely free-standing.

60 In this case, unspecified techniques, rules and discovery methodologies that implicitly privilege the material placed in collections without exposing the basis for such privileging to scrutiny.

