



This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Googlepository and the University Library

[Sue Manuel](#) and [Charles Oppenheim](#) discuss the concept of Google as a repository within the wider context of resource management and provision in Further and Higher Education.

Main Contents	Section Menu	Email Ariadne	Search Ariadne
-------------------------------	------------------------------	-------------------------------	--------------------------------

Introduction

The development of an increasing array of tools for storing, organising, managing, and searching electronic resources poses some interesting questions for those in the Higher Education sector, not least of which are: what role do repositories have in this new information environment? What effect is Google having on the information-seeking strategies of students, researchers and teachers? Where do libraries fit within the information continuum? And ultimately, what services should they look to provide for their users?

The concept of Google as a repository was introduced at a recent JISC conference [1]. Hitherto, it has been speculated that Google might be considered to be a digital library [2]. This viewpoint provided a catalyst for the process of considering the differences between Google and repositories. We have evaluated this in terms of an exploration of their features and the services they provide. This leads on to a discussion of the potential value of these services to content contributors and consumers. In the long term, this will be expressed by users' engagement with the services and tools available to them.

Background

In discussing repositories and Google, it may be useful to consider some of the reasons for their growing use. Traditionally, scholarly information was disseminated in print form; libraries played a vital part in the delivery of this printed material to a mass audience. The role of libraries is summarised in the following quotation from Mullins, "Books have always been a means of communication, a store of the history of the world passed down from generation to generation" [3].

The physicality of books and the effect they have on people remains evident even in the age of born digital material. For example, the artist Rachel Whiteread used the theme of the library and books for her Holocaust Memorial unveiled in Vienna in 2000. Whether this relationship with printed texts will continue into the future remains to be seen. Set against the tradition of the use of texts, the role played by libraries in serving the needs of their users in the Higher Education sector is clearly in a phase of transition. Libraries now have a dual role in maintaining and enhancing services for printed resources, and also in providing access to, and maintaining, services for a growing collection of digital resources.

Thus, libraries are seeing a shift in the nature of their collections (physical to virtual) and in the ownership of resources. As Muir states,

"Increasingly, information is not purchased but 'rented' through licensed access rather than physical ownership of an information artefact. Libraries are increasingly losing physical control over their digital collections" [4].

This has repercussions for collection maintenance and the preservation of digital assets. In addition, libraries are experiencing changes in readers' usage of resources and in attitudes towards information retrieval [5].

Librarians recognise these trends and are responding to the complex range of factors influencing their profession. This process of transformation brings with it the requirement for new services, skills and approaches to the curatorial role libraries fulfil. In the following table, we outline key considerations in the new knowledge-sharing environment.

Resource	Factor	Consequence
Format	<ul style="list-style-type: none"> • paper • digital 	<ul style="list-style-type: none"> • storage • maintenance • management • ownership • preservation • cost (production, storage, management) • staff skills
Physical location	<ul style="list-style-type: none"> • library building • digital • repository 	<ul style="list-style-type: none"> • space (physical, server) • maintenance • management • ownership • preservation • cost • staff quotas • staff skills
Search and retrieval	<ul style="list-style-type: none"> • Index cards • face-to-face enquire • electronic enquiries • database searching • Information portals • search engines (Google) 	<ul style="list-style-type: none"> • staff skills • liaison with external organisations • information literacy • ease of use • quality

Use of Google

The growing number of digital resources leads information seekers to look for easy-to-use and effective search and retrieval tools. In turn, this has led to a proliferation of electronic tools (databases, search engines, etc) to facilitate the location and retrieval of digital versions of books, journal papers and other items. However, users today often focus exclusively on the use of Google.

Digital resources are the preferred source of reference for many. Evidence of this can be seen from studies into the techniques researchers employ for resource identification and location. One

recent study found that 72% of researchers' first, and sometimes only, port of call for resource discovery was Google [6]. Furthermore, their preferred format was a digital version of the item, and one that was immediately available and accessible by them [5].

The dominance of Google continues to be the cause of much debate, especially in relation to the benefits and drawbacks for student use. University students often use Google as their primary tool for locating information for their coursework. Some of the perceived benefits in using Google are that it is quick and easy to use and it often returns many search results. The drawbacks include: the reliance on a single source for information; too many results may be returned especially if keywords are not carefully selected; duplication of hits; too many false drops; and inaccurate material may be presented alongside authoritative items. Information literacy encourages the use of a range of search tools, as well as the analysis of information retrieved to ascertain its quality, accuracy, currency and provenance.

University lecturers also routinely use Google to locate information [7], and have reported using Google to search for materials to supplement their teaching [8]. In the future, deposit of teaching and learning resources into repositories may help to make a wider range of materials available to teachers for use in the creation of courseware. No doubt these resources will be located by the individuals' preferred search method. In other words, repositories will attract and house content that is then discovered via a Google search.

Google as Repository

Google is a search engine, is it not? Perhaps not, as many additional services are now provided under the Google banner, including Web space and analytic tools. Therefore, we can say that Google is no longer simply a search engine, but a provider of a range of Web services. Franklin [9] has suggested that Google is his repository. This position is partially supported by the minimum requirements for repository services described by Heery and Anderson as, "put, get, search, access control" [10]. Some of these core components are provided by Google. For example, content can be shared through Google Base [11], information is retrieved via Web links, and Google's search capabilities are well known. Access control is not yet part of the Google package, but it can be provided at source if required.

A growing number of services can be viewed as repositories. These could also be said to have a limited number of the core repository characteristics. The content housed in these services, and in repositories, also varies. To illustrate, some contain metadata records only, others hold items focusing on research outputs while others contain a range of resources for use in teaching and research. There are also single subject repositories, regional repositories and those whose content is derived from a range of subjects and regions. Repositories can be managed locally, by institutions; regionally, by a consortia of institutions; nationally, by centrally funded organisations; or they may be personal repositories created by utilising services that are freely available.

However, the concept of Google as repository presents a number of potential problems. Google is not a managed store of resources; this means that standard metadata is not provided, items are not preserved, and they can be moved or deleted at will. For libraries in the Higher Education sector, two issues arise with Google as repository: information literacy; and the transparency of resource provision.

Information Literacy

Google, and other search engines, often return resources out of context, i.e., with no supporting information. This makes their provenance, accuracy and currency difficult to judge. Repositories can provide this contextual information in the metadata describing the item, as well as in any supporting information for its use. With this data, the origin of the item and its quality becomes easier to determine. Providing this additional information does require that repositories have procedures in place for recording, or capturing metadata, and that the quality of that metadata is good.

You could, perhaps, argue that Google generates its own metadata records dynamically on-the-fly. These are presented to the information seeker in the form of search results. But, the links provided may, or may not, lead to items with supporting descriptive text.

Resource Provision

Students may not be aware that they only have access to an item, located through Google, because their university's library has paid for it. Users are understandably not necessarily concerned with where a resource has come from, just that it is available to them. However, this may be of concern to libraries as they might want to brand their subscription resources. In addition, students might conclude that the library was not needed because information was located and retrieved through Google. Hence, users are not aware of the resource provider, or of the cost to their institution.

Within the existing landscape, there are no clear solutions to the issues we have touched on here. Repository services can help to provide solutions by providing contextual information alongside the link to a resource, and by branding items so that users are aware of the source.

Unique Elements to a Repository Service

There are a number of elements that set repositories apart from other systems and services including:

- persistent identifiers - facilitating reliable access to items
- open access - making repository content available to a global audience without charge
- mediated deposit - reducing workloads for content contributors
- cataloguing and metadata - describing items, their creation, conditions for reuse, asserting copyright, and providing a record of modification
- advocacy - promoting and encouraging contribution to a repository
- preservation - thereby guaranteeing future access to resources

Two of these elements, advocacy and preservation, are particularly pertinent to these discussions. Advocacy leverages content that would otherwise remain locked in personal stores, or behind university firewalls. Preservation activities ensure that data remains accessible in the long term.

Advocacy

Early research repository managers recognised the importance of advocacy efforts in raising awareness of repositories, and in encouraging contributions to them. Advocacy to encourage the sharing of resources for use in teaching is even more vital as it can help to overcome the reluctance to share. These resource creators can be unwilling to make their materials available to others for a number of reasons. Amongst those cited for this behaviour are: not realising the value of their materials to others; being uncertain that their resources are good enough to be made widely available; and not knowing of any suitable repositories in which to place their items [\[12\]](#). Repository advocates can provide information and assurances to allay these fears.

A range of advocacy efforts can be employed to engage repository content contributors and content consumers. Printed materials can be used to highlight the benefits of sharing research and teaching resources. The steps required to make items available (self- or mediated deposit) can also be outlined. One-to-one sessions and group events can be organised to publicise a repository. Advocacy may not always be as effective as we would like, but some institutions have been successful in their content recruitment campaigns.

Alongside an institution's internal advocacy activities, there are national and international initiatives to introduce mandates for the deposit of research outputs into repositories. The merits and effectiveness of mandates for research repositories are described in detail by Brody et al [\[13\]](#).

Mandates may be one way of driving content into repositories but they are not always popular with resource creators.

Preservation

Another element in the repository mix is the preservation of digital assets. Preservation is not an issue for Google, but it is for repositories. At the JISC Digital Deluge Conference [\[1\]](#), Franklin argued that, ideally, teaching materials over five years old should be regarded as out of date, and therefore have no value in reuse. Others would disagree with this standpoint, but in any event, preservation activities encompass a range of material categories. This includes teaching resources, published research papers, conference papers, raw data, images and the like.

The importance of the preservation of digital material has been highlighted by the MIDESS Project. Its participants describe the range of information required for the effective preservation of digital objects, stating that,

"Simply physically storing the digital data isn't enough for long term preservation. Details about the application used to create the digital file, rights to the digital file, ability to extract the physical attributes of the file (such as size, format, etc) and store these attributes as metadata is also important for the long term digital preservation of digital material" [\[14\]](#).

Repository managers need to make decisions about whether to preserve, what formats to preserve, and what preservation methods to adopt. The cost of this activity has to be balanced against any possible future gains. Preservation of the significant properties of an object may be viewed as a sensible route to take.

Repository policies need to take into account the preservation of digital assets and the metadata associated with an item. Key problems to address here are detailed in the DCC Digital Curation Manual [\[15\]](#). The specific issues associated with the preservation of metadata for learning objects are described by Campbell in the Instalment on Learning Object Metadata. In this Campbell describes the IEEE LOM standard; its application and uptake; some of the issues that implementers of this standard need to be aware of; and future developments. In addition, Campbell also points out that LOM "was not designed with long-term preservation of either resources or metadata in mind" [\[16\]](#). Digital curators, and others, are currently grappling with these issues.

Decisions in this area will be supported by the ongoing work of standards organisations, JISC CETIS [\[17\]](#), existing and start-up repositories, and other initiatives, such as those funded by JISC. To illustrate this, 'Retention of learning materials: a survey of institutional policies and practice' [\[18\]](#) looks at policies and practice in Further and Higher Education for retaining learning materials. It also considers short-term preservation plans with a view to highlighting good practice. Another useful example is the HATII toolkit which provides a useful audit tool for repository managers [\[19\]](#).

The benefits and drawbacks of the provision of good-quality metadata are also often discussed within library and repository communities. One consideration for the future might be the loss of cataloguing skills within these communities. As many repositories are established within the remit of the library, it is of some concern that a number of libraries are out-sourcing their cataloguing work. By this we refer to the practice of buying in shelf-ready books. Economies of scale take precedence where budgetary constraints are in operation, but loss of key skills may have long-term implications.

A repository could be described as a service that includes a tool for uploading resources which are then stored, organised, catalogued and preserved. In addition, they also offer internal search functionality, presentation of metadata for search by external agents, and the retrieval of stored items. All this is supported by unique advocacy activities. On the other hand, the primary role of Google could be defined as that of a search tool.

Value to the Higher Education Community

As the number of resources being commissioned, created and reused in the Higher Education sector grows, so the task of effective management becomes more complex. Developers aim to provide effective systems to store, organise and present resources to users. In their turn, resource creators and information seekers look for easy-to-use systems for sharing, locating and accessing the materials they need. Thus, one service of value to researchers lies in the provision of a mechanism for the dissemination of information. As Swan and Brown state, "Communicating their results to peers remains the primary reason for scholars publishing their work; in other words, they publish to have an impact on their field." [6]. Google, repositories and libraries all have a part to play in improving dissemination, and thus research impact.

Libraries can benefit from setting up a repository because it affords them the possibility of regaining a degree of control over the digital assets created within their institution. Preprints of published papers can be archived and preserved for the long term. This may be useful where the publisher's version is not available and where a digital copy is required for preservation purposes. Licensing terms and copyright issues in respect of preservation are complex, especially where the item is not held in a library's permanent collection. This is increasingly the case with subscriptions to digital resources that are held in a publisher's database [4]. Licences terms for repositories can be created to include provision for copies to be made for preservation purposes. In addition, copyright in the metadata associated with these items can be asserted.

With the co-operation of a number of publishers, the LOCKSS Project is currently testing the creation and storage of copies of journal papers on an institution's server [20]. These can be used as a backup when a publisher's version is not available, or when a publisher ceases operations.

Information seekers benefit from the improved findability of items as a result of the addition of good-quality metadata to describe the item and its conditions for reuse. Morvill discusses the issue of findability and identifies some of the potential problems in this area as, "Poor information architecture. Weak compliance with Web standards. No metadata. Content buried in databases that are invisible to search engines" [21]. As Morvill states, "Google loves metadata"; it is one of the three elements included in its algorithms, the other two being full text and popularity measures [21].

Community Engagement

There are increasing trends towards engagement with roles that have traditionally been undertaken by professional librarians. Many information seekers see no need to consult the services of professionals in their search for resources; they believe Google can find all the data they need. Another task, once the preserve of librarians, is that of cataloguing. We are all assuming this role when we tag items deposited in social software sites. Social software tools for organising, cataloguing, tagging and sharing resources are growing in popularity. Amongst the items being shared are texts, images, multimedia, bibliographic details, reference lists, reading lists or Web links. Services that allow users to share their personal library catalogue with others include LibraryThing [22] and Reader² [23]. With these services, items can be tagged, feeds can be used to pull data into other services (blogs, news readers or aggregators), and users with similar reading tastes can be discovered.

The Open Library Project [24] takes the activity of community tagging a step further. Their aim is to create a comprehensive online catalogue comprising the details of every book published. This will be seeded by libraries which donate their existing catalogue to the Open Library. The metadata entry for each book will then be edited and supplemented by public contribution.

Repositories also need to engage with content contributors to ensure that a critical mass of content is reached. Achieving this through seamless blending with existing workflows will help to make contributing to repositories more of a common-place activity.

Artists too can benefit from the greater exposure and availability of electronic resources. To illustrate this, we can look to the Archives Hub which features a digital artist in residence. The work

of Aileen Collis was shown in June 2007 [25]. The starting point for the creation of her work during the residency was a digital image of an archive document. One key to the reuse of materials is raising awareness of copyright issues to ensure compliance with the law.

Conclusion

Libraries have long been recognised as central stores of information. The increased availability of digital resources has provided new opportunities for both the acquisition of resources and the spread of knowledge through user communities. System developers, service providers, funding bodies, publishers, contributors to repositories and repository users all have their own drivers for facilitating resource exposure. Systems and services are increasingly overlapping, merging and interoperating to satisfy these diverse agendas. Today's information seekers want to be able to locate and access materials as quickly and easily as possible. Google is seen by many as the ideal tool to use in the quest for information.

There does not appear to be a definitive answer to the question, 'is Google a repository?' There are some similar functions and services provided by the two and the Google brand is constantly expanding to include a variety of services and tools. Perhaps we could view Google as a meta-repository. Digital materials reside in repositories or on servers; these items are exposed to Google; Google matches search terms to information contained within resources; a list of possible matches is then presented to information seekers; they then use their judgement to reduce the number of hits to that which best meets their requirements.

Google is a commonly used tool for locating resources for use in research and teaching. Can Google be used more effectively and selectively? Yes, so long as teachers and learners are willing to participate actively in improving information literacy skills. However, it is important to remember that Google is just one tool to help information seekers. The role of the library is to demonstrate how best to use this tool, and to expose users to a range of other sources of information. Libraries play a vital role in making resources accessible to a wide audience; Google and repositories serve a similar function in an online environment.

It is possible that in the long term, all information will be transferred electronically. The tactile quality of paper copies will be lost as well as its 'affordance' [26], but much will be gained. The quality of digital copies is as good as the source, items can be produced and transmitted speedily, searching for suitable information will be fast, and creators will have new sources of inspiration in both the digital materials themselves, and the new tools and technologies used in their creation, transmission and repurposing. This will involve a transformative process. What is required is a shift in both personal and technical perspectives.

Repositories, libraries and Google complement each other in helping to provide a broad range of services to information seekers. This union begins with an effective advocacy campaign to boost repository content; here it is described, stored and managed; search engines, like Google, can then locate and present items in response to a search request. Relying on Google to provide search and discovery of this hidden material misses out a valuable step, that of making it available in the first instance. That is why university libraries need Google and Google needs university libraries.

References

1. Digital repositories: dealing with the digital deluge, JISC Conference, 2007: http://www.jisc.ac.uk/events/2007/06/repositories_conference.aspx
2. Cloonan, M. and Dove, J. 2005. Ranganathan online: do digital libraries violate the third law? *Library Journal*, 130(6), 58-60.
3. Mullins, C. 2004. *Rachel Whiteread*. London: Tate.
4. Muir, A. 2004. Digital preservation: awareness, responsibility and rights issues. *Journal of Information Science*, 30(1), 73-92.
5. Brown, S. and Swan, A. 2007. Researchers' Use of Academic Libraries and their Services. Available online at: <http://eprints.ecs.soton.ac.uk/13868/>

6. Swan, A. and Brown, S. 2005. Open access self-archiving: An author study. Available online at:
http://www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf
7. McKay, D. 2007. Institutional repositories and their 'other' users: usability beyond authors. *Ariadne* issue 52. Available online at:
<http://www.ariadne.ac.uk/issue52/mckay/>
8. Loddington, S. Bates, M. Manuel, S. and Oppenheim, C. 2006. Workflow Mapping and Stakeholder Analysis: Final Report. Available online at:
<http://hdl.handle.net/2134/2124>
9. Franklin, T. 2007. Paper presented at [1].
10. Heery, R. and Anderson, S. 2005. Digital repositories review. Available online at:
http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf
11. Google Base: <http://base.google.com/support/>
12. Bates, M. Loddington, S. Manuel, S. and Oppenheim, C. 2007. Attitudes to the rights and rewards for author contributions to repositories for teaching and learning. *ALT-J, Research in Learning Technology*, 15(1), 67-82.
13. Brody, T. Carr, L. Gingras, Y. Hajjem, C. Harnad, S. and Swan, A. 2007. Incentivizing the Open Access Research Web. *CTWatch Quarterly*, 3(3). Available online at:
<http://www.ctwatch.org/quarterly/articles/2007/08/incentivizing-the-open-access-research-web/>
14. MIDESS digital preservation requirements specification. Available online at:
<http://www.leeds.ac.uk/library/midess/documents.html> See: " MIDESS workpackage 5 - Digital Preservation (PDF)"
15. DCC Digital Curation Manual: <http://www.dcc.ac.uk/resource/curation-manual/>
16. Campbell, L. 2007. DCC Digital Curation Manual: Instalment on Learning Object Metadata. Available online at: <http://www.dcc.ac.uk/resource/curation-manual/chapters/learning-object-metadata/>
17. JISC CETIS: <http://www.cetis.ac.uk/>
18. Retention of learning materials: a survey of institutional policies and practice: JISC
http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories/retentionlearningmaterials.aspx
19. Humanities Advanced Technology and Information Institute: HATII toolkit:
<http://www.hatii.arts.gla.ac.uk/news/toolkit.html>
20. LOCKSS: <http://www.lockss.org/>
21. Morvill, P. 2005. *Ambient findability*. Beijing: O'Reilly.
22. LibraryThing: <http://www.librarything.com/>
23. Reader²: <http://reader2.com/>
24. Open Library: <http://demo.openlibrary.org>
25. Collis, A. 2007. Archives Hub Collection of the Month, June 2007: Pick 'n' Mix. Available online at: <http://www.archiveshub.ac.uk/jun07.shtml>
26. Sellen, A.J. and Harper, R.H.R., 2002. *The myth of the paperless office*. Cambridge: MIT Press.
The authors use the term 'affordance' to mean the properties inherent in an object and the opportunities people perceive in using it.

Further Reading

Ayre, C. and Muir, A. 2004. Right to preserve? Copyright and licensing for digital preservation project. Available online at:
<https://dspace.lboro.ac.uk/dspace/bitstream/2134/343/1/Final%20report.pdf>

Barker, P. 2007. Digital deluge eLearning strand. Available online at:
<http://blogs.cetis.ac.uk/philb/2007/06/19/digital-deluge-elearning-strand/>

Beagrie, N. 2005. Plenty of Room at the Bottom? Personal Digital Libraries and Collections. *D-Lib Magazine*, 11(6). Available online at: <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>

Costigan, A. 2007. Google: the Good, the Bad and the Ugly. Presentation from the July 2007 Getting to Grips with the Google Generation workshop. Available from ALISS Website: <http://www.alissnet.org.uk>

Head, A. 2007. Beyond Google: how do students conduct academic research? First Monday, 12(8). Available online at: http://www.firstmonday.org/issues/issue12_8/head/index.html

Author Details

Sue Manuel

IT Support Officer
University Library
Loughborough University

Email: s.manuel@lboro.ac.uk

Charles Oppenheim

Head of Department of Information Science,
Loughborough University

Email: c.oppenheim@lboro.ac.uk

Web site: <http://www.lboro.ac.uk/departments/dis/people/coppenheim.html>

[Return to top](#)

Article Title: "Googlepository and the University Library"

Author: Sue Manuel and Charles Oppenheim

Publication Date: 30-October-2007 Publication: Ariadne Issue 53

Originating URL: <http://www.ariadne.ac.uk/issue53/manuel-oppenheim/>

[Copyright and citation information](#) File last modified: Thursday, 15-Nov-2007 08:58:42 UTC

Main Contents	Section Menu	Email Ariadne	Search Ariadne
-------------------------------	------------------------------	-------------------------------	--------------------------------

Ariadne is published every three months by [UKOLN](#). UKOLN is funded by [MLA](#) the Museums, Libraries and Archives Council, the [Joint Information Systems Committee \(JISC\)](#) of the Higher Education Funding Councils, as well as by project funding from the JISC and the [European Union](#). UKOLN also receives support from the [University of Bath](#) where it is based. Material referred to on this page is [copyright Ariadne \(University of Bath\) and original authors](#).