Online

**Column Editor: Anita Vautier,
Information Services, University of Wales,
Aberystwyth**
e-mail: **aiv@aber.ac.uk**

Institute of Physics

The Institute of Physics have digitised their journal collection back to 1874. This means that over 110, 000 articles are available digitally as PDF documents that can be cross searched and contain linked references. The archive is available as a one off purchase or on a subscription basis. http://journals.iop.org/

JSTOR

A new search engine has been released for JSTOR. This new search engine has been developed to improve the speed ability. The basic search will now search across all the content which includes articles and reviews. Other features available in the basic search are Boolean operators, phrase searching and field searching. When a Boolean operator is not entered the automatic default is AND. http://www.jstor.org/help/search.html

Ulrich's Resources Linker

Bowker have recently announced their Resource Linker for serials in the library. There is an A-Z listing of journal titles and a search facility that searches the journal title and articles. It is also possible to subject browse. The Resource linker is hosted by Bowker, who will set-up, maintain and update the link server, which will update any catalogue changes made to the collection. http://www.resourcelinker.com

Google Scholar

This search engine is provided by Google and is the beta version. A search entered will retrieve books and abstracts for any articles found. If an open access journal appears in the results the full text will be available. The Google Scholar search engine searches scholary journals and any peer-reviewed literature. http://www.scholar.google.com

People's Network Online Enquiry Service

This service will be available to the public in March. It provides a free information service twenty four hours a day, seven days a week. The public can submit a question via the online form. The service is provided by MLA (Museums, Library and Archives Council) who are currently piloting the scheme. They are asking library and information staff to submit a question to help test the service. http://www.peoplesnetwork.gov.uk/

Internet News

**Column Editor: Susan Miles;
e-mail: susan_miles_2002@hotmail.com**

Googlizing the Library?

Google announced in December 2004 that "it is working with the libraries of Harvard, Stanford, the University of Michigan, and the University of Oxford as well as The New York Public Library to digitally scan books from their collections so that users worldwide can search them in Google" [1]. This precipitated newspaper articles both here and in the U.S. trumpeting the liberation of students and researchers from "musty library catalogues", along the lines of the Guardian's "endless journeys to the library could become a thing of the past" [2].

The Scope of the Project
So, what exactly is the scope of this project? At first glance one's initial reaction is that this new project is certainly a Herculean task. Yet when I tried to find out, actual numbers seem somewhat thin on the ground – whilst the University of Michigan is contributing seven million volumes and the Bodleian Library one million volumes [3], other participants are more guarded about the scope of their contributions. Stanford University has a collection of over eight million volumes but its FAQ on the new project says "The agreement with Google is open-ended; it neither targets specific collections nor specifies a minimum or maximum number of books to be digitised". [4]

Similarly, the New York Public Library's press release makes no mention of how many books will be scanned, it is participating on the basis of a "pilot program that will make a subset of its books available" [5]. There is another pilot project at Harvard University; initially 40,000 of its 15 million volumes will be digitised. Interestingly their selection critieria for materials to be included in the project is to randomly select entire shelves of materials at the Harvard Depository, which will provide a diverse range of items in terms of age, topic and language mainly because the books are shelved by size! [6]

How will it be done?
Google seems to be taking a two pronged approach to the digitising process, Stanford University will send books direct to Google's headquarters in California, whilst the other four libraries will digitise in-house. Google will establish local units which will carry out the digital scanning and processing using proprietary, high-speed scanning stations, which once fully operational will be able to produce as many as 10,000 electronic books per week. Once the book has been scanned, the resulting images will be sent to Google's central facility for quality control and conversion to text. The contributing library will receive a copy of the digital file for their own use. Whilst Google appears to be

providing the staffing for the scanning stations I wonder how much library staff time and effort will be taken up with delivering, keeping track of, and returning, such large quantities of books on an ongoing basis. Other observers have conjectured that an average scan rate of 3,200 volumes a day for 365 days a year will be needed to meet the 6-year timeframe for the University of Michigan's seven million volumes. As far as guessing how much this is going to cost Google, Gordon Macomber, president and CEO of Thomson Gale has indicated that $10 per book was below Gale's experienced cost. They have extensive experience with digitizing for their 18th and 19th Century Collections products. [7]

As anybody who has had anything to do with the digital capture of documents will tell you, these projects can (perhaps this should read often) prove to be more difficult than is initially anticipated, issues such as quality control, or costs, can easily escalate, and let's not forget the challenge of managing the workflow at the sort of work rates which will be needed.

Copyright Issues
Initially it is likely that out of copyright works will be the first books to be digitised. Users searching with Google will see links in their search results page when there are books relevant to their query. Clicking on a title will deliver a Google Print page where users can browse the full text of public domain works and bibliographic data for copyrighted works. It is intended that searchers will read these books online, and may not be able to print out the whole document. There may be plans to offer searchers tasters for copyrighted works, namely, three "snippets" of text from throughout the volume. A "snippet" would consist of approximately three lines of content. There would be pointers to direct searchers to where they can obtain full copies of copyrighted works. By the end of the project Google will have effectively created the largest out-of-print bookstore.

Industry Response
One day after the Google announcement, the Internet Archive announced details of their own collaboration with libraries of five countries [Canada, China, Egypt, India and the Netherlands] to add their digitised books to open-access archives. Over one million books have been committed, with 27,000 currently available, with an additional 50,000 expected in the first quarter of 2005. [8] It remains to be seen if, and how, companies like Amazon, Microsoft or Yahoo! will respond to Google's announcement. Equally, the vast collections in the Library of Congress and the British Library remain outside of these large-scale digitisation projects. They remain the jewels in the crown.

Unanswered questions
What about duplicates? There appears to be no plans at present to prevent duplicate copies across contributing libraries being submitted to Google. Also, it will not be possible to select items by contributing

library as no separate identifier will be created. Since the whole book is being digitised, if a library has barcoded and added its own identifying marks to say, the front or rear pages, then these will be captured and will be readable.

What about the impact on the wider library community? Carol Brey-Casino, current President of the ALA said in a Wall Street Journal interview, "We had this conversation when Internet began to get popular, and what's happened is that library visits have doubled in the last decade". Many commentators see no threat but a further push for librarians to become more active in empowering users to find, evaluate, use and access information in a variety of formats.

Some final thoughts:
I wonder if information professionals will be recruited into these digitising projects, or will existing experts in this field be clamouring to be involved in these projects?

One remaining question to ponder – it is likely that if you belong to a public library then you already have access to online electronic books, whether they be reference works or technical books – so, how often have you, or do you make use of them and have you ever read one cover-to-cover onscreen?

Further Reading

Googlizers vs. Resistors. Brian Kenney. *Library Journal*. 12/15/2004.
http://www.libraryjournal.com/article/CA485756

Google Partners with Oxford, Harvard & Others to Digitize Libraries. Gary Price. *Search Engine Watch*.
http://www.searchenginewatch.com/searchday/article.php/3447411

Google's Library Project: Questions, Questions, Questions. Barbara Quint. *NewsBreaks*. December 27, 2004.
http://www.infotoday.com/newsbreaks/nb041227-2.shtml

1. Google Checks Out Library Books. Google Press Center.
http://www.google.com/intl/en/press/pressrel/print_library.html

2. Top Libraries to go online in deal with Google. Bobbie Johnson. *The Guardian*, Wednesday December 15, 2004.
http://www.guardian.co.uk/print/0,3858,5085842-103690,00.html

3. Oxford-Google digitisation agreement. Press Release. http://www.bodley.ox.ac.uk/news/news58.htm

4. Stanford University and Google Mass Digitization Project: FAQ.
http://library.stanford.edu/about_sulair/special_projects/google_sulair_project_faq.html

5. NYPL Partners with Google to Make Books Available Online. Press Release.
http://www.nypl.org/press/google.cfm

6. FAQ: The University's Pilot Project with Google.
http://hul.harvard.edu/publications/041213faq.html

7. Google's Library Project: Questions, Questions, Questions. Barbara Quint. *NewsBreaks.* December 27, 2004.
http://www.infotoday.com/newsbreaks/nb041227-2.shtml

8. International Libraries and the Internet Archive collaborate to build Open-Access Text Archives.
http://www.archive.org/iathreads/post-view.php?id=25361

Intranets and Content Management

Column Editor: Martin White
Intranet Focus Ltd;
e-mail: martin.white@intranetfocus.com

Another ten intranet success factors

A couple of years ago I developed a set of ten intranet success factors that ended up being used as a give-away on the UKOLUG stand at Online Information 2003. In rebuilding the Intranet Focus web site (which should look different by the time you read this) I thought it was time to review and revise the success factors, so here is the 2005 version. The main difference is really that the initial set were for start-up situations, and this set are for more mature (in relative terms) intranets. Comments would be much appreciated.

1. Document the intranet strategy
   Develop and document an intranet strategy, based on a consideration of the requirements and balance of information/content, technology and governance. This strategy has to be aligned with the overall strategy of the organisation and other relevant applications, such as records and document management, client/contact management and digital asset management. An important element of the strategy is defining acceptable measures of performance. A marketing strategy also needs to be included.

2. Understand and support content contribution
   Information quality in an intranet is essential, and yet most organisations implement a distributed approach to content contribution without really understanding the issues around content management. Invariably intranet content contribution is not included in job descriptions or valued by managers, and as a result gets low priority. The old adage 'garbage in – garbage out' certainly applies to intranet content!

3. Develop and enforce standards, and encourage the adoption of good practice
   There should be standards for intranets, especially where the organisation has a number of different intranets within some form of common architecture. These might typically include the basic layout of pages, default arrangements for lists, and conformance to accessibility standards. There have to be enforceable sanctions for non-conformance. A standard without such enforcement is not a standard, just a dream. In addition there will be good practice that may vary from department to department and from time to time. This good practice should be collated and reviewed.

4. Manage metadata
   Content authors may not have the skills and expertise needed to add metadata. Metadata is not just about 'keywords'. Structural metadata describes the information architecture of the document; content metadata provides a way of identifying documents that may contain relevant subject information; descriptive metadata enables the type of document to be identified. And finally administrative metadata deals information such as the person and department owing the document, the date when the document would be checked for relevancy, and the language of the document. The effort involved in developing metadata schemes is always underestimated.

5. Provide effective search functionality
   Users have to trust the search functionality so that they feel totally confident that either they have found all relevant information, or that information is not on the intranet. Not being sure one way or the other is not an acceptable situation. Implementing a good search solution will depend on a robust metadata environment, understanding how users will search for specific types of content, and considering the extent to which the intranet search engine should also search other repositories, such as records and document management applications.

6. Provide access to the business environment
   Make sure that the intranet provides access to information from external sources, such as business and market information. It is easy to be so focused on the provision of internal information that access to external information is overlooked. In today's competitive business environment monitoring the external business environment is of the highest importance.

7. Undertake regular usability testing
   Usability testing should be carried out at all stages of the life of an intranet, and there should be a range of feedback channels to ensure that the content and the information architecture continue to meet the expectations of contributors, users and stakeholders. Carry out regular surveys to obtain success stories that can be used to market the benefits of the intranet, and also identify examples