

Digging for Victory: Finding Difficult Information

a review of the East of England ISG workshop held 26 April 2006

Candace Guite, Librarian, Christ's College, Cambridge

This was an all-day workshop held in Norwich's new flagship Archive Centre. The entire collection has recently received designated status from the MLA and, during the lunch-time break, delegates were able to view a selection of the Archive's treasures on display.

The day was aimed at all sectors of the ISG membership base, with librarians attending from public, academic, local government and special libraries, and it provided plenty of opportunities for networking in between four excellent talks and the Group's AGM. All speakers rose magnificently to the occasion when persistent technical glitches resulted in the failure of all internet connections, but fortunately none of the substance of the talks was affected. The theme of the day "Finding difficult information" focused particularly on the challenges of extracting the precise and the relevant from the plethora of information available on line. **Phil Purdy**, Development Officer of the Museum, Libraries and Archives Council, gave us a practical working insight into recent progress on the People's Network and demonstrated how the provision of a high-quality portal to web resources can bridge the concept of the traditional library and the library without walls. **Terry Kendrick**, currently a freelance business consultant, considered the 'deep' or invisible web and looked at ways users can access these hidden resources. **Alastair Allan**, Senior Liaison Officer at the University of Sheffield Library, looked at the complex area of e-Government, and in particular at the problems that can arise with diverse publishers and a bewildering variety of published formats. Finally **Jane Mortimer**, in her role as Academic Librarian: Resources and Services at De Montfort University, considered support for the virtual user, taking as an example the students on the De Montfort distance learning programme.

The People's Network Enquire service

Phil Purdy gave us a detailed look at 'Enquire', the online reference service for members of the public launched in October 2005. 'Enquire' is one strand within an overarching strategy (Enquire, Discover and Read) designed to ensure that members of the public are given the practical resources and the 'know-how' to navigate through and derive the best advantages from all e-Britain has to offer.

'Discover' www.peoplesnetwork.gov.uk/discover provides a single portal to news and trusted web sites, and gives access to collections of digitised images from museums, libraries and archives. 'Discover' uses open-source software, allowing users to personalise the portal to meet their particular requirements and interests, and facilitates communication and the sharing of

resources.

'Read' www.peoplesnetwork.gov.uk/read/index.html is underpinned by, but goes beyond, the basic literacy agenda, with a remit to encourage the exchange of ideas amongst readers. Users can contribute their own online reviews, and access reviews contributed by others. They can also link to online reading groups or contact a reading group local to where they live. All three interlinking strands of the People's Network are managed by the MLA, and are lottery-funded at present. Future costs may devolve down to local authorities via a subscription scheme, but the economies of scale should ensure that costs are kept low and that the service continues to provide good value for money. A new and refreshing element (and one vital to the success of the project) has been the close partnership achieved between UK and North American libraries in order to deliver Enquire's national 24/7 service.

The new service is not intended to replace existing traditional reference services where the user approaches the reference enquiry desk to ask a question. Rather it should be regarded as a complementary value-added service, and because it is delivered through libraries it has the public's traditional trust in the quality and professionalism of the service provided.

The value-added component includes more than just the 24/7 element. The powerful software OCLC-PICA (QuestionPoint) which underwrites 'Enquire' facilitates the pooling of expertise, a very useful feature when particularly difficult and complex enquiries arise. It also allows the librarians who staff the service to draw on a very wide range of regional, national and international resources, and then to link these, where appropriate, to local resources available to the user.

It appears, though, that the public sometimes labours under a misconception that 'Enquire' functions as a remote service, with a time-delayed response; but it is hoped that as users become more familiar with the service this will change. 'Enquire' does just what it says it does www.peoplesnetwork.gov.uk and provides a service which connects to a real librarian in real time. But your friendly helpful online librarian just might be in America or Canada if your question comes at 2am!

The service is staffed on a rota system mostly by English public libraries (98 library authorities in England representing about 65% of public library authorities) between 9am and 7pm. Phil explained that the MLA hopes that this figure will go up to represent 100% and that in future academic libraries will be included too. Outside English office hours it is staffed by North American libraries. 'Enquire' currently deals with about 60 enquiries a day, 72% of which are answered by English public libraries and 28% by international partners. Most of the enquiries received are answered using online resources, and feedback from users has been very positive to date. Two authorities, Kent and Essex, have pioneered 'Enquire' as a local service sitting alongside the national resource. It is staffed by their librarians during office hours and then switched over to the national resource out of office hours.

Information on the invisible and deep web

Terry Kendrick gave an inspired presentation on the invisible, hidden or 'deep' web. Terry's background in public libraries (he clearly relished his earlier incarnation as a reference librarian!) has influenced his down-to-earth and practical approach to the problem. In Terry's estimate, only 6% of what is on the web is visible. It is an indication that you are entering the deep or invisible web when the www disappears from the top of your screen. Terry suggested that if we consider the capabilities of just one search engine, the ever-popular Google for example, it can, at the most, spider off 10% of the web.

Terry went on to pose a number of 'What if's?'. For example, what if the Google spider visits the site but goes away before searching the whole site? The trawl would then retrieve only some of the pages or - what is just as bad - only parts of pages. What if the author of the web pages scanned by the spider had used a "no robots" command? Terry again drew from his own experience to give us an example with some of his PhD research (the application of risk management tools and techniques to strategic marketing planning) which is posted on the web. There are benefits in sharing research and in stimulating debate, but clearly a web author would also want to exercise control over when and how that access happens. What if the material was put up last week or operates in "real time" and is dynamically generated? What if the spelling is wrong? Some search engines suggest corrections and offer alternatives but not all.

So for the user there is a real danger of bypassing just the kind of high quality resources likely to contain the answers to the questions posed, e.g. databases, library catalogues, private, and password-protected commercial and academic sites, picture collections and mashups, which comprise programmable technology mashed up with a database. Terry's example of a mash-up was <http://chicagocrime.org> which links the Chicago Police Department's crime records to Google maps, allowing the user to make very detailed and precise searches, e.g. homicides on 45th Street. The nature of much of this embedded and graphic material is simply not amenable to a conventional single engine search strategy approach. To balance this, new technologies capable of revealing parts of the hidden web are constantly evolving; for example, 'Bittorent' (this file-sharing software that facilitates the fast movement of files has all sorts of applications).

Terry gave us another example from PowerPoint. The notes from a PowerPoint presentation are embedded information, but by selecting 'edit' slides the user can gain access to the notes pages, and thereby retrieve potentially useful and possibly commercially sensitive information. The slides notes are there, of course, to act as a prompt for the speaker, but often what remains unsaid can be equally revealing.

Google is text-driven and does the best it can, but it might bypass an essential database. Terry gave us a scenario using Companies House and his own company as an example. For a typical business-related quick reference query, where the user wants to establish the date of incorporation of a

company, it is best to go directly to the single most authoritative source, i.e. the Companies House website. This is a fast, direct method that yields accurate information. But what happens if the user chooses to conduct a 'Terry Kendrick' Google search? An experienced user might well be aware that different search engines yield different results according to the search algorithms deployed. There can be an overlap as low as 30% between two different search engines' results, so information which appears to be hidden may be so purely owing to the limitations of the search engines used. Our user might know about the danger of collapsed sites (the phenomenon whereby a significant proportion of retrieved hits simply does not display), and is probably aware of the problems inherent in language, e.g. in variant spelling forms, or of the issues that arise *vis-à-vis* the use of natural language as against technical terminology. But assessing the accuracy, relevance and authority of the retrieved items can be more difficult. In the case of our Google 'Terry Kendrick' query, the user would be sadly led astray by one of the retrieved items, an interview with Terry Kendrick, in which the date given is wrong by a year or so: a mistake explained by a simple lapse of memory on Terry's part.

Our user might be less aware that when he performs a Google search the results are retrieved from a cache. How and when that cached information is indexed and updated also clearly affects the outcome of the search. So what can the user do to find the information he needs? To view the scope and range of resources available, Terry suggested the use of lists, e.g. 'Complete Planet', a directory of databases; 'Gary Price's Direct', links to hidden resources, and subject gateways. A good subject gateway will reveal the hidden web (e.g. databases) as well as the visible resources. Additionally the user with a good knowledge base of the subject, and one who can make an imaginative leap and visualise how the database layout attempts to map that subject, clearly has the advantage. A final plea from Terry: "Don't give your brain away to a search engine!"

Searching for information on government web-sites

Alastair Allan helped unravel the complex maze of e-Government information. E-Government publication is all about making that information available in some kind of electronic format. There are huge potential advantages in the use of an electronic medium, for example in widening access, increasing speeds of transmission, and in building a genuine two-way communication with citizens. But Alastair suggested that the government has been less ready to consider the associated information issues that arise. Some of these issues are technical, e.g. effective long-term archiving, and some are social, e.g. the digital divide and technological illiteracy.

The way that e-government is conducted varies but can include Freeview, e.g. BBC Parliament, Teachers' TV, digital televising of council meetings and SMS messaging as well as web publication. The advantage of these methods is that the public have direct access to all the information and can weigh up the debate for themselves as opposed to receiving it through a media filter.

Users need to be able to find information easily and to know which

department produced it and when. The user also needs to know how to track the document again, should it be required for future reference, so long-term archiving, bibliographic and version control are as important as good web design. But good website design for e-Government is another area in which Britain lags behind Canada and the USA. Alastair suggested that the key features of a good government website must be accessibility and ease of navigation, whether this is for the first-time user or the information specialist. Regrettably there is a dwindling number of government information specialists (they currently number only 28), and universities too no longer seem to be appointing or maintaining such specialists.

Alistair looked at the major obstacles to retrieval that face the user of government information. The first is related to the diversity of publishers and access to the published documents. The user cannot get hold of the information if he does not know who the publisher is. There have been fundamental changes in the way in which government information is published, so the search strategies and processes are as varied as the enquiries. This is manageable for the specialist dealing on a day-to-day basis with official publications but it presents a real challenge for the reference librarian who may only deal with such enquiries on an infrequent and irregular basis.

There used to be a single government publisher (HMSO) which operated with selected High Street outlets, but now citizen and information specialists have to deal with an increasing number of conventional publishers, as well as the growing trend in desktop publishing. Every single government agency can also act as a publisher, producing information in diverse data formats. There are, for example, approximately 300 thousand Home Office PDF pages. In this instance, unless a user knows exactly what he is looking for, or is lucky enough to have a web citation, his search is likely to be lengthy. It all makes for a compelling argument for a commitment, by all departments, to well-linked and structured archives. It also makes the case, as Alastair pointed out, that more information specialists are needed, not fewer. Our traditional taxonomic skills apply as much to web page and database construction as to catalogues. Alastair's suggested formula: "One good information manager equals two brilliant graphic designers".

Once the user has identified the document required, the next hurdle is access. As we saw earlier, High Street availability has gone with the demise of HMSO, and there is no longer any single central sales point. To make matters worse there is still no complete national bibliography and there are some gaps in the national library archive. To compensate, it is worth trying databases like COPAC, which contains 27 million subject links and is one of the best resources for government information.

There are portals designed to direct the user to the correct source, e.g. [Directgov](#) but in Alastair's view British e-government documentation and provision compares unfavourably with that in the USA, where over 1081 federal publishers operate. A major difference is that the major American universities have very good collections which include Britain, and Alastair explained that it is often helpful to go through an American portal for hard-to-

find British official publications. GODORT (hosted by Northwestern University Library) and the University of Michigan Library are particularly recommended. Alistair recommended the use of Google Scholar, or indeed any of the 'advanced' features provided by the search engine to limit searches to the gov.uk domain name. The second obstacle in attempting to access government sites is that it is difficult for users to formulate questions correctly, and they may not be able to confirm that the information found is what they require.

There is a good case for saying that we all need information skills training, and more investment is needed in librarians' training for the future. But we also need to ensure a wide distribution of leaflets that list the key local and national government web sites. Bearing in mind that some users have no keyboard skills or familiarity with web resources there is still a need for conventional publications to give, at the least, background support information.

Information provision in academic institutions

Jane Mortimer's paper looked at the needs of one very specific group of users, distance learners at De Montfort University, and considered the ways in which e-resources, combined with physical access, can support their study. The increasing range of electronic resources available combined with greater collaboration between HE institutions (e.g. UK Libraries Plus scheme) have helped equalise the gap between the full-time campus student and the distance learner.

All distance learners at De Montfort fall into a clearly designated group to ensure that they have full access to course resources: for example, there are some services which are offered exclusively to distance learners like the postal loans and photocopying services. There are numerous reasons why students may opt for distance learning but they include work, study, family commitments, geographical distance as well as personal choice. There are 16,222 full-time students (68%) but a sizeable proportion (24.2%) of the student population is part-time, with distance learners accounting for 7.5%.

Some courses at De Montfort are offered only as distance learning courses, and there are also part-time courses which suit students on placement, involved in research or working part-time. These are typically offered through partnerships with franchise colleges (there are 1,810 partner colleges).

De Montfort has taken steps to widen physical access to the collections by providing 24/7 access to the building, inter-library loans, links to external resources, and an online catalogue, thus ensuring that all students, both campus-based and part-time and distance, have maximum opportunities. Online reading lists link up with the catalogue.

Electronic resources at De Montfort comprise electronic journals, subject databases, e-books and reference materials, e.g. dictionaries, directories and standards. Electronic journals are accessed via OPAC and database links and from A-Z listings but, as with their print counterparts, their provision gives rise to significant cost issues. Subject databases cover both subject-specific

and multi-disciplinary resources and can be full-text or selected abstracts. Jane spoke about the issues of equity and equivalence, using the provision of subject resources for business and management by way of example. These provide comprehensive full-text services, but students of art and design are less well served. The e-books collection is small, consisting primarily of background material rather than primary texts, and does not appear to be heavily used by students. Jane referred to a current research project at UCL to examine use of e-books by academic staff and students. It is a project that will be watched with interest by HE institutions which are waiting to take decisions on investment in remote resources. De Montfort also provides a number of other information resources, e.g. web portals, digitised collections of examination papers, electronic course packs and so on. Students are also offered an off-campus library account, and have remote access to files created first on campus.

Offering such a wide range of services, and ensuring that they are correctly targeted to the part-time and the distance learner, presents considerable technical and other challenges. Athens authentication provides the standard entrance route, but students have different logins for email and VLE, and a separate login and PIN for library use. User education and support is seen as a priority. Face-to-face inductions are given where possible but are backed up by CD, video, web-based tutorials, and online library guides and publications. Other services offered include telephone and email support for users, and virtual reference and enquiry facilities are provided via a web-based enquiry form. Jane concluded her paper by reminding us all that library services need to be marketed, hence De Montfort's @access anywhere campaign: at the very minimum, students need to know that the library web page is their gateway and that for all students, not just distance learners, there is much more on offer. However, ensuring that distance learners in particular take full advantage of the support offered presents a real challenge for the future.

All four speakers took joint questions at the end of the afternoon and there followed some stimulating debate on the whole range of topics covered. If you missed this East of England event, make sure you register for the next one! It is a good way of keeping up to speed with fast changing and complex areas, and of networking with fellow practitioners.