



PubMed Central, Royal Society Publishing, British Library Direct Service

MY picks are two databases that offer free access to articles in highly regarded journals, PubMed Central (PMC) and Royal Society Publishing (RSP). Open access (OA) for scholarly journals and free full-text articles have grown in popularity over the past few years. The pan is the new version of the database that fuels The British Library (BL) Direct services. It levies steep charges for millions of OA articles through copyright and service fees that often have little to do with reality. This not-so-nobly exploits the ignorance of those who remain unaware of the OA revolution in scholarly publishing. I picked PMC and panned BL Direct 2 years ago in this column, but the former got even better, while the latter got even worse, so an update is due.



My picks are two databases that offer free access to articles in highly regarded journals, PubMed Central (PMC) and Royal Society Publishing (RSP).



the picks

PUBMED CENTRAL

Among the government institutes, the National Institutes of Health (NIH) has made, by far, the best use of taxpayers' money. PMC (www.pubmedcentral.nih.gov) stands out with 1.6 million full-text searchable articles from more than 400 journals, most of which are available free of charge in both HTML and PDF (unless you have the misfortune of finding them through BL Direct). Out of the 1,619,713 total articles, reviews, letters, and corrections, only 3,988 were not available free for immediate download as of Aug. 2, 2008. The full-text searchability of all the documents is a huge advantage in comparison with IngentaConnect, Scitation, Scitopia, and WorldWideScience.

The monthly rate of new additions to PMC is sharply increasing—2008 alone saw the addition of one-third of all the content in PMC. It certainly helped that at the end of last year the U.S. government mandated immediate deposit of NIH-funded research papers with full availability 12 months after publication.

Many of the best biomedical and life science journals are represented in PMC, but their retrospective ranges differ widely. Some are available from the first to the current volume immediately, such as the excellent *Bulletin of the Medical Library Association* and its successor, the *Journal of Medical Library Association*. For some journals, going back to the first volume may mean only a few years, as is the case for some of the newly launched titles from OA publishers BioMed

Central, Ltd. and Hindawi Publishing Corp. PMC lists some 470 distinct journals, but note that title changes are counted separately.

About 350 of the journals have no moratorium for OA and their publishers deposit articles automatically. The largest publisher in this group is BioMed Central, followed by Hindawi. Many of its journals are also covered by Web of Science, Scopus, Inspec, and ZentralBlatt MATH.

This does not mean that the other 100-plus journals have less than a 12-month moratorium before allowing OA. Some have a 24-month moratorium, while half of the more than 40 BMJ Publishing Group, Ltd. journals have 36 months of delay. Interestingly, the other half are immediately made OA.

PMC's software is good or excellent in every regard. It has the best browsing style available for author names and journal names, which spares you the pain of guessing how to search authors whose name includes van, von, de, della, or del as prefixes, which in many systems become suffixes, making them unfindable. Similarly, journal names are easy to look up because you don't need to know if *JAMIA* is the proper title or the acronym of the *Journal of the American Medical Informatics Association*. You don't need to know the fact that the *Bulletin of the Medical Library Association* started as the *Bulletin of the Association of Medical Librarians*, which was changed after the first volume, then changed again 100 years later to its current title. You just need to type in the string `med i c a l l i b` and the three variants (along with two unrelated journal names) pop up in the index.

It would be great to implement the same style of browsing for many of the other data elements where the entries may scatter due to different abbreviations and punctuation, such as the author affiliation index.

Searching is comprehensive—practically every metadata element is searchable; only the names of cited journals are missing. They are part of the Reference data element, but just as the Cited Author element is searchable in the distinct Reference Author index, the same should be true for Cited Journal/Source. Otherwise searching for journals such as *ONLINE* as a cited source is impossible, as the entry in the Reference Index may refer to words in the title of the referenced paper or its publication status. Searching for *Science* as a cited source is even more hopeless without a separate browseable index for cited journals' names.

One warning: The list of cited references may not include all the items cited in the original paper. Papers scanned from print publications identify and link to only those cited references that have a matching master record in the PubMed file. The PDF versions of the articles, of course, include all the references, even if they are not digitally actionable.

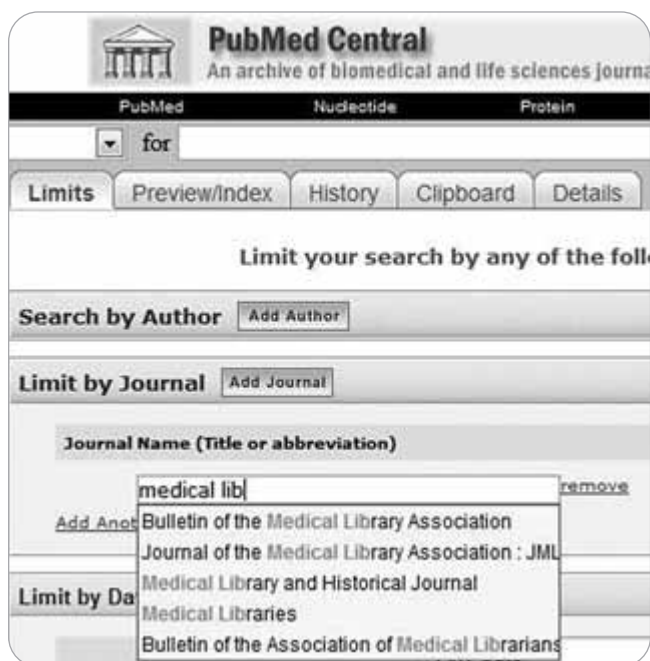
Except for the sort options, the output features are fine. The layout and design of the results list and the item display are ideal, showing the chapter/section titles in a sidebar and links to related articles in PubMed. This is based primarily on the similarity of words in the title, abstract, and MeSH terms of the source and related items, not on the ratio of common citations in them (as in Web of Science and Scopus).

The sort options include journal name and several dates (publication date, e-publication date, PMC entry date). It should include author names (as does PubMed, which even distinguishes the first author and last author as sort parameters), as well as citedness count (within PMC), which could be a very useful clue in trying to find the most-cited papers on a topic that yields hundreds of hits. Relative citedness by age of the article would be the best for this.

All in all, this excellent database should be the first resource when searching for a very new or esoteric topic in life sciences, where full-text searching is the most needed, and every time before making a delivery order for a source document.

ROYAL SOCIETY PUBLISHING

Royal Society Publishing (<http://publishing.royalsociety.org>) is the publishing arm of the Royal Society of London, the oldest national society for promoting science and the functional



PMC's drop-down list for journal name shows variant titles.



The excellent layout and design of the record display, with additional options in the left-hand sidebar

equivalent of the U.S. National Academy of Sciences. But the analogy stops here because while the latter (which is among the top cited multidisciplinary serial publications) has its backfiles dating to the first volume from 1915 accessible online for free, with only a 6-month delay on making the most current issue available, the journals of the Royal Society of London have a much smaller free backfile and a 1-year moratorium before becoming freely available.

In spite of these limitations, I chose it as a pick because the site was finally redesigned, getting a better interface and search engine (based on the MetaPress software). The site itself is one of the best designed publisher sites—it has a good layout, background information, and even citation metrics data about the journals (ISI impact factors, immediacy index, rank within category). With that said, the data is not consistent, and it is hard to understand why the h-index given for *Philosophical Transactions A of the Royal Society* from Scopus is a very low value of 30 (in the multidisciplinary field), when the Century of Science database produces a much more appropriate h-index of 141 to reflect the cumulative clout (as envisioned by Jorge E. Hirsch) accrued by the journal through more than a hundred years. True, it is not as high as *Proceedings A of the Royal Society*, which has an h-index of 271 in Century of Science versus 60 in Scopus for a variety of reasons that would require much more space to explain.

There are 63,265 documents in the digital collection, but only a portion of them are freely available—papers published before 1997 and after Aug. 1, 2007, are exclusively for subscribers. Knowing that PMC has 4,000 papers from *Proceedings B*, and half as many from *Philosophical Transactions B* (along with 400 papers from *Biological Letters* and nearly 250 from *Interface*), I guesstimate that there may be an additional 7,000–8,000 papers from *Proceedings A* and *Transactions A* that are not available through PMC.

Although relatively small, it is still a very important, partially free service, considering the clout of the journals, especially the two flagship publications, the *Philosophical Transactions of the Royal Society* and the *Proceedings of the Royal Society* family. I call these family because both split into Section A and Section B to focus on mathematical, physical, and engineering sciences (A), and biological sciences (B), with recurring title changes.

These are the journals that published many of the landmark papers in the 18th, 19th, and 20th centuries, including the writings of Edmund Halley about the eponymous comet, Jeremiah Dixon and Charles Mason about the Mason-Dixon line, Watson and Crick about the double helix, and Isaac Newton about numerous discoveries. They are not free, but just looking at the records and reading the titles, which are often as lengthy as the entire abstracts in some indexing/abstracting databases, is fascinating. In the current free subset there may already be some papers that will be considered landmark in the next few decades—particularly as they will be more widely available and germinating than the ones I mentioned above, exactly by virtue of

their free access to the biologists, physicists, physicians, mathematicians, and engineers of today and tomorrow anywhere in the world.

The software is good, but it has shortcomings. It has no browseable indexes at all, so the many misspellings of Halley's first name (Edm is one version) are not apparent. Yes, you can search by last name alone and find 100 Dixons as authors, but I cut to the chase, and his four papers, by remembering Dixon's rather distinguishing first name—not because I am so educated, but simply because I often listen to *Sailing to Philadelphia*, one of the splendid albums from Mark Knopfler, where the title song opens with the line "I am Jeremiah Dixon, I am a Geordie boy." In fairness, the software has a nice clustering sidebar that shows the most prolific authors, creating a form of minibrowsing, but it does not help in the case of misspelled singletons and other small hits. Results cannot be sorted and there are no options for different results display formats, nor is there an indication about the citedness of papers included within the collection.

The software has some irritating features, such as suggesting weird alternatives for your search terms. The query `halley comet` prompts a suggestion for `hal le comet`, and brings up two hits, then twists my arm offering `hal l comet`. It offers `matsunami` when I search for `tsunami`. The annoying thing is that `matsunami` brings up four hits (for a cited author). However, I am more than consoled by finding 38 very promising hits on tsunamis, of which 29 are freely available and more than 30 will become freely available by the time you read this, as their moratorium will soon expire.

The MetaPress search software will be replaced in early 2009 by my favorite search software from HighWire Press, which together with the current site design will provide an even better platform for this high-value, partially free digital collection.



RSP search results showing the strange suggestion to use *matsunami* instead of *tsunami*



the pan

BRITISH LIBRARY DIRECT

In my early 20s, while working in Hungary, I was a frequent user of the Boston Spa services. For the sake of the younger generation of librarians, I may need to add that this did not mean massage services from a parlor in Massachusetts, but interlibrary loan services from the then Lending Division of The British Library in West Yorkshire. It was expensive but subsidized, and I benefited from accessing current documents, so I fancied BL. Lately, my like for this division has vanished after seeing inconsistent and unrealistic copyright fees that were much higher than what the publishers charged, as I illustrated in my earlier review in *ONLINE* (November/December 2006).

Although BL Direct (www.direct.bl.uk) corrected some of the absurd copyright fees and calculations, the rise in service charges well beyond the inflation rate makes my blood pressure rise. The latest 20% raise in cost of the 2-day electronic delivery (which can be 4 days on weekends, as when I wrote this piece) adds insult to injury. In addition, while the old software was replaced, it still does not search the full text, which keeps the system atavistic for resource discovery.

Although The British Library proudly announced that the number of journals in digital formats increased to 8,500, users barely see any advantage. It is certainly beneficial to The British Library, as even the lowest-waged employee can be trained in a few minutes how to call up an article and send an email to the customer at minimal cost, while the service charge for the 2-day delivery went from £21 to £25 (roughly \$44.50), on top of the copyright fee and the VAT beginning Aug. 1, 2008. [*The immediate download fee is £8.75, but this applies only to articles BL Direct already has in electronic form.*—Ed.]

As for customers, they should use BL Direct only when all other alternatives fail. Their savings on document delivery will be significant just by checking the largest and best (partially or entirely) free depositories, such as PMC, HighWire Press' free subset of 2 million articles, and the millions of high-quality and free scholarly papers available through MetaPress, arXiv.org, CiteSeer Plus, the Astrophysics Data System, RePEc, and other OA repositories, which grew in number from 100 in 2005 to 1,200 as of this writing. They can be located through the superb *OpenDOAR* (<http://opendoar.org>) now maintained by the University of Nottingham. Add to these the free articles directly available from hundreds of publishers—either because of their goodwill, obligation, or because the author pays (typically from their spon-

Your basket

You have successfully deleted items from your basket

Date added	Article title	Delivery	Order number	Total	Remove
02-Aug-2008 22:11:31	Tsunami: scientific frontiers, mitigation, forecasting and policy implications	2 hr delivery	822014000028	£37.02	<input type="checkbox"/>
02-Aug-2008 22:10:58	Tsunami science before and beyond Boxing Day 2004	2 hr delivery	822014000027	£37.02	<input type="checkbox"/>
02-Aug-2008 22:10:21	Focused tsunami waves	2 hr delivery	822014000026	£37.02	<input type="checkbox"/>
02-Aug-2008 22:09:58	Tsunami: a history of the term and of scientific understanding of the phenomenon in Japanese and Western culture	2 hr delivery	822014000025	£37.02	<input type="checkbox"/>

Total excluding VAT £126.00
 VAT (17.5%) £22.08
 Total price including VAT £148.08

The cost for 2-hour delivery of four articles that BL Direct could find

sor's funds) to make the article OA through the Open Choice program (and similar alternatives that publishers initiated). BL Direct offsets all these in case of the less-informed customers.

It is distressing to see how much (close to \$300) the user would pay for the four articles from the Royal Society journals that BL Direct could find for my tsunami search (29 out of 38 are free through the publisher's website).

Even many publishers without free services offer far better prices than BL Direct. For example, an article was published in *Tectonics* a year ago and predicted the likelihood of a catastrophic earthquake in Sichuan, which happened in May 2008. While I understand the authors' bitterness that this journal may not have reached the most-affected target audience, which could have mitigated the death toll, it is not fair to say, that "the article was effectively locked in an academic journal." Through the publisher, the American Geophysical Union, the paper is available for instant download for \$9—the price of a ticket for a lousy movie. The cost through BL Direct is inflated, with an absurd price of £36.43 (about \$70) for 2-hour delivery—that is unacceptable for most researchers.

I have no doubt that The British Library has noble services, and there were good reasons that Lynne Brindley, the chief executive of The British Library, was made a Dame of the British Empire—but BL Direct is definitely not a good reason. It must be difficult to keep charging such abusive fees for articles that are under the nose of the BL for free in the UKPMC version of the PCM database. Do they know about it? I trust they do—after all, it is hosted by The British Library. This must require a Janus-faced policy for promoting science. Customers pay the price in more ways than one.

Péter Jacsó (jacso@hawaii.edu) is professor of library & information science at the University of Hawaii's department of information and computer sciences.

Comments? Send email to the editor (marydee@xmission.com).