

## Raiders of the lost articles

Limitations in online journal archives and citation search tools, coupled with inaccessibility of older journals in libraries, threaten to disconnect us from our microbiological history.

In many scientific fields there seems to be an increasing lack of appreciation of the early literature around which the field has grown. As a consequence, junior scientists entering a field, who might be unaware of some of the work of the early pioneers, risk wasting a great deal of time and money asking scientific questions that are baseless or have already been answered.

The underlying reasons for this disconnect are varied, but one important aspect may be the way in which we access the scientific literature. Powerful online citation search tools such as [PubMed](#) and [Google Scholar](#) have revolutionized the way in which we interface with the published literature and are invaluable tools for anyone engaging in scientific research. Furthermore, free-access digital repositories such as the [PubMed Central](#) archive provide easy access to an increasingly large proportion of the published literature. However, these tools are only as useful as the contents of the databases around which they are built, and unfortunately much of the older literature is absent. In addition, the easy access to information using such tools may have tempered the desire for scientists to spend hours trawling through dusty volumes in darkened corners of the library looking for key pieces of information.

One area in which this is keenly felt is bacteriophage biology, a field that, according to Ry Young from Texas A&M University, College Station, USA, was “born prematurely” in the early part of the twentieth century. A search in PubMed using the term ‘bacteriophage’ brings up over 56,000 entries, the earliest of which is from 1925; by 1950, 475 entries are listed. However, this is almost certainly a substantial underestimate of the total number of papers that were published before 1950, and most of those that are listed have no abstract archived, let alone a digitized copy of the paper. Another bacteriophage researcher, Stephen Abedon from Ohio State University, Columbus, USA, maintains his own database of the early literature. Comparison of this personal database to PubMed provides a telling estimate of the amount of published literature that is effectively invisible to most people in the field; a ‘phage therapy’ PubMed search yields 171 hits, going back to 1946, whereas Abedon’s database goes back to 1926, with over 50 journal articles being ‘invisible’ even to a ‘bacteriophage therapy’ PubMed search. Some might question whether this discrepancy actually matters and

whether experiments in the older literature would stand up to the scrutiny of more modern techniques. However, both Young and Abedon reject this idea, and Young suggests that, in the early literature, “The data are more accurate than the current stuff, with more attention to the biology, more attention to controls”. Both scientists work at large universities and are still able to access many of the early papers through their institutional libraries, although the older, hard-bound journals are increasingly being put into storage, reducing access to papers in these volumes. However, at many institutes, physical libraries face closure, as more and more users want to access content only electronically. As Abedon puts it, “For many people, if they can’t put their hand on a reference in under five minutes, then the study doesn’t exist.”

The problem is by no means limited to the bacteriophage field; in the early part of the twentieth century, ground-breaking research was carried out in fields such as bacterial physiology, antibiotics and virology, much of which is not listed by citation search tools. So what can be done to address this issue? The simplest solution would be to ensure that all of the older literature is digitized and to make the abstract, as well as the title, visible to searches in PubMed. Many journals have done just this; for example, the online archive of *Science* contains PDFs of articles dating back to the first issue in July 1880. Other journals, including *Nature*, have digitized their entire back catalogue, although not all of the online archive is listed by PubMed. However, ensuring that the gaps are filled in the literature that is available electronically will not be sufficient in itself. Whether through writing grants, papers and theses, through teaching activities or simply through background reading on a research project, the next generation of scientists must be encouraged to engage more fully with the older literature. Not only will this provide individuals with a greater appreciation for the history of their given field, but it will also help to inform their current work and the scientific questions that they ask.

If we fail at this task, then we face the very real prospect that much hard-won knowledge will be lost.

### FURTHER INFORMATION

Google Scholar: <http://scholar.google.co.uk/schhp?lr=>  
PubMed: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>  
PubMed Central: <http://www.ncbi.nlm.nih.gov/pmc>  
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

“ the next generation of scientists must be encouraged to engage more fully with the older literature ”