

Burning books without fire

To most people who have enjoyed, or at least been grateful for, an academic education, book-burning is an especially egregious crime, because it is done to limit the minds of a culture by erasing the recorded thoughts of those who went before. The annulment of the intellectual achievement of the authors is an irreplaceable loss, if all copies are destroyed and, to many people, obliteration of an author's writing or art or mathematics betokens all too clearly a wish to obliterate the author. And anyone else inconvenient. As Heinrich Heine warned, *Dort wo man Bücher verbrennt, verbrennt man auch am Ende Menschen*.

Burning of books is usually a very public affair, intended to excite one set of people, to intimidate others, or both. And in recent times, deliberate burning of books in the West has usually been harmless to the survival of the actual literature: the loss of the copies of Harry Potter books recently burned by Greg Locke and his congregation in Tennessee, for example, will not make an appreciable dent in the availability of the texts. I appreciate that the same is not true in other parts of the world, and that irreplaceable texts and artworks have been lost in Iraq, Syria and Afghanistan. It is all public because it is deliberate, and done to make a point.

But hard-won literature and knowledge can also be lost not because someone chooses to destroy it, but because nobody will keep it safe. This kind of loss has none of the drama of a bonfire of thought, and perhaps because of that, nobody much seems to care.

The phenomenon of lost knowledge stems mostly from the 21st Century trend towards electronic-only publication, and is one of the two reasons I have never liked this trend (I'll mention the other one briefly at the end). In the days of paper books and research journals, even if a journal folded or a publisher went bankrupt, the libraries that had subscribed to the journal or bought the books would still have the copies on their shelves. Many journals I have read exist no longer, and it has not mattered beyond the need to cross a windswept George Square and enter the library, nodding to its famous cat in passing. But this is not true of electronic books and journals. Access to most of those depends on the publisher maintaining a web server or some similar equipment. This does not just mean keeping a box plugged into a wall socket: internet standards and database standards are constantly changing. Websites that do not keep up with the new ways of working effectively become unreachable (think of all that material made for Adobe Flash, which is now very hard to view for anyone who has not made a point of keeping a machine non-updated). So if a publisher

goes bankrupt or just gets fed up of spending money maintaining access to the archive of a failed journal, the journal effectively vanishes.

This is not gloomy prognostication on a dystopian future: it is already happening. Earlier this month, Mikael Laakso, Lisa Matthias and Najko Jahn published a preprint of an article in which they showed that 176 digital journals, not papers or individual volumes, but complete journals containing a great deal of hard-won research, have disappeared from the web, apparently completely. There were not, by and large, predatory 'journals' that are really money-making scams that print any old nonsense for money. The fact that 88 of them were associated with learned societies makes that clear. Further more, there are at least 900 digital journals that have closed, but that are still available online, for now. Nobody is burning them: they are just pulling the plug, but the loss is the same.

This is very worrying. There have been various attempts to address it. Some libraries cooperate with publishers in keeping copies of content, to act as a backup for publishers, though typically publishers have to pay for this. One suspects that the ones who have the resources and common sense to pay are probably not the ones likely to fold in the first place. There is also the Sci-Hub project, which makes copies of academic papers available free (whether they are copyright or not). Because Sci-Hub holds copies, it is at least an effective backup, though its coverage is by no means complete and it tends to focus on non-open-access papers from large commercial publishers. The publishers object, of course, and keep trying to shut it down or to get national domain name servers or web services to cut off access (if you can't get access, use a VPN to connect to the Web in India, where the law currently allows Sci-Hub to operate openly, and then you can reach it with ease).

At the very least, very strong scientific nations ought to build a system to which failing publishers can send their journals *free*, so that they can be made available *free* forever from the serves of national libraries of those nations.

The problem is even bigger for databases. Scientific funders are often relatively happy to fund the creation of a new database, but not to fund its continued maintenance, which has to go on forever, both in the sense of updating with new knowledge and in the sense of keeping the computational gears nicely oiled. This is a real problem, and some very high-profile databases, used by tens of thousands of people and also by AI systems, are at real risk. When approached for help, research

fundes reply, correctly really, that the maintenance of a database is not research but infrastructure (like keeping a library open or the internet fibres intact), and therefore not their remit. But there seems to be nobody whose remit it is. We (many databases) go on, sailing perilously close to the wind, scratching money from all kinds of odd sources, with key staff, highly qualified, faced with extension after extension to short-term contracts. And sometimes databases just die, taking with the vast amounts of important knowledge. Unlike journal articles, databases are not simple 'documents' that can be passed to a library, but are active, functional things that depend on the right kind of computational environment. Providing even this, from a national resource, would be possible, as big computers can always emulate smaller and older ones. But again it takes some country's research leaders to commit money to a very boring project that will not excite the public one little bit, but that they know is actually important.

In the Guide to Pharmacology, one of our responses is to produce a paper (and pdf) version of the essential core of the database, the Concise Guide to Pharmacology, and to publish it every two years. It is not a database but, if the database is lost forever, at least that much information will remain.

The other reason I distrust e-journals? It's a boot-strapping problem. Reading paper journals and books requires no special technology,. If things go very, very wrong with our civilization (think of a global catastrophe scenario of your choice) then as long as libraries exist physically, people wanting to build again can at least have access to all the knowledge that went before. But if we end up depending on 21st-century or later technology to be able to read instructions about how even to build the technologies of earlier times, then people trying to make it our of new dark ages would have to work everything out from scratch again instead of being able to look it up. This strikes me as more than a little stupid.

Links:

The Laakso paper: preprint at <https://arxiv.org/abs/2008.11933>; stable version at <https://doi.org/10.1002/asi.24460>