Bibliographic Management 2.0

Frank Norman

There are now several online tools, many free of charge, for managing researchers' personal collections of scientific literature. Some of them are in the mould of traditional bibliographic management software, but others started out as Web 2.0 applications. These new tools have changed the face of bibliographic management.

The bibliographic reference is the foundation of scholarship (1). A reference is a surrogate for knowledge, a surrogate for research results that have been condensed into a journal article and then stored in the minds of scholars. Reference management therefore is all about knowledge, though we sometimes forget that.

In the early days the main function of personal bibliographic database programs were to provide a readily accessible and searchable store of knowledge, and possibly also to act as an index to a reprint collection (2). Until the early 1990s (3, 4) these remained the main reasons for wanting to use such programs. Annotating and subject tagging references were important functions, but bibliography production was relatively unsophisticated. This was the era before end-user online searching, so having access to your personal database on your desktop was an advantage. The tools at this time emphasised information storage and retrieval.

During the 1990s networked desktop access to major bibliographic databases became common, and personal word-processor packages were widely adopted. These two developments led to changes in personal bibliographic software usage, and the production of a bibliography in an appropriate format became the raison d'être of having a personal database. Why bother creating your own comprehensive personal database to search when the whole of PubMed is available to search from your Web browser? Hence, the knowledge component of reference management was pushed aside and we focused more on the notation of the reference – the arcane codes that we use to refer to documents – and on the mechanics of bibliography assembly and production, integrated with the process of writing a manuscript. A large number of programs competed in this market, but two programs became dominant: Endnote and Reference Manager, both owned by the same company. In UK Higher Education, site licences covering both products became available, and we thought we had the situation covered. But nothing stands still, and a new trend emerged towards online bibliographic management tools.

Endnote produced a web-based version called Endnoteweb. It was originally a cut-down version of the main program, and is still less functional than the desktop version of Endnote. A new product called Refworks was designed from scratch as a standalone online tool. EndnoteWeb and Refworks have been competing with each other in the last year or two, launching new versions at a frightening pace. Both are still primarily about bibliography assembly, I feel. A new generation of bibliographic social bookmarking tools have put knowledge back at centre-stage: Connotea, CiteuLike, Zotero and Mendeley. I think bibliographic social bookmarking is the next wave of bibliographic management.

Social bookmarking is a way for Internet users to share bookmarks to information resources on the Internet. It's an information science version of "You show me yours and I'll show you mine". The earliest tool (5) was Delicious but others quickly followed, including Connotea, which was aimed at the scientific community. The growth in e-journal availability means that now nearly all journal articles can be regarded as "information resources on the Internet" and therefore can be bookmarked by these tools. Connotea added functionality that lets it recognise different data elements in online bibliographic references and thus bibliographic social bookmarking came into being.

From the early days of Eugene Garfield's Institute for Scientific Information (ISI) and his pioneering work in citation indexing, it was realised that a reference, or a collection of references, can define a topic. Papers that have all cited the same group of references are likely to be closely related in topic. If I am very interested in a particular paper and another three papers have all cited that paper then I am probably interested in those three papers too.

Bibliographic social bookmarking takes this idea a step further. If a scientist has bookmarked twelve papers and I have also bookmarked eight of those papers, then there is a good chance that the other four papers will also be of interest to me. Papers bookmarked by several people who share interests with me (as defined by their previous bookmarks) are also very likely to be of interest to me. As the number of users of social bookmarking increases, so the chances of finding common interests increases. Eva Amsen (7) has described how the adoption of bibliographic social bookmarking could work to the benefit of scientists, comparing it to the way that Flickr has become a large repository of images.

Programs like CiteULike, Connotea, Mendeley and Zotero are at the heart of this trend. They vary in functionality but all are free, and all have grown into the bibliographic world from the Web world, whereas the established products have gone the other way. Connotea and CiteULike are both supported by major publishers; Zotero is an open-source product supported by a US university; Mendeley is a start-up company that has been successful in raising venture capital. These tools have many users now – Mendeley claims to have 8 million references shared, with 100,000 users (6). These new tools have also been marketed at scientists and researchers directly, bypassing library support. We now need to become familiar with these tools, embrace them if they are useful, and ensure that their developers are aware of the role of libraries in providing support for bibliographic management.

Libraries have started to take note of these tools. In 2008 in the USA the Northwestern University Library organised a workshop called CiteFest (8) to compare established tools like EndNote and RefWorks with newcomers such as Zotero, CiteULike and Connotea. Attendees worked through a series of exercises to test the functions of each product. Citefest declared Zotero the winner of their challenge. In 2009 I helped to organise a workshop held at University College London, "Bibliographic Management meets Web 2.0". It focused on the needs of researchers, who are heavy users of these tools. Representatives from six of online bibliographic management tools put their products through their paces and attendees had a chance to try some exercises (different from but inspired by the CiteFest exercises). Martin Fenner, one of the co-organisers and a speaker at the event, summarised the day on his blog (9) suggesting that since "all the reference managers demonstrated were up to the challenge (though Connotea and CiteULike couldn't put references into a Word document) it would be wrong to declare a winner". He also pointed out that "the market is developing so fast, that a feature comparison will look very different 12 months from today". Martin has produced a comparison chart (10) that shows the features of all the products (and others too) in a simple visual way.

Early in 2010 "Innovations in Reference Management" (11) was held. This event, IRM10 for short, was organised by the TELSTAR project "to showcase and discuss innovative ideas and developments in the use of Reference/Bibliographic Management software." This featured several of the same products again, but also included some usage case studies. The TELSTAR project is subtitled "Integrating References and Citations into Learning Environments" but the event considered the needs of researchers as well as students. The organisers of IRM10 suggested that there is "a lack of ‘community’ to discuss and collaborate around the practice of Reference Management" adding that many of the products lack a user group. The TELSTAR project blog (12) captured the main speakers' talks. Also in 2010, Martin Fenner co-organised another workshop, this time for a group of German librarians, "Reference Management in Times of Web 2.0" (13). A similar list of products was reviewed, but this workshop focused more specifically on Web 2.0 aspects of the tools, once more observing that there was no clear leader – it was more a matter of personal choice.

Portability is now a key requirement for bibliographic management tools, so users are looking for online tools that are flexible and easy-to-use. Users also want tools that support interaction with research collaborators. Easy sharing of references and the ability to tap into community knowledge is an interesting development but it has not yet become a 'must-have' feature. It may not be easy for the new generation of tools to wean users away from their existing tools, but since these new tools are mostly free of charge, they come with a low barrier to adoption. Although it is true that in general scientists are not rushing to adopt Web 2.0 tools, they will use tools that save them time and money. The ability to see what is this week's most-read paper in your field of interest in Mendeley or CiteuLike, or to see what papers are popular in your network of contacts, is a small thing but could become important especially as the user

base grows. It is an interesting time for bibliographic management.

1. Tim Birkhead, *Times Higher*, 4 Jan 2008. http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=400029

2. Eugene Garfield, (1983). In *The Application of Mini- and Micro-Computers in Information, Documentation and Libraries*. pp 79-84. Sci-MateTM: A User-Friendly Information Storage and Retrieval System for Microcomputers http://garfield.library.upenn.edu/papers/316.pdf

3. Thomas E. Wolff, (April 1992), *Database*, pp 34-39. Personal Bibliographic Databases: An Industrial Scientist's Perspective.

4. JHRD Correia, (1990) *CABIOS*, 6(2), 126-7. Management of personal bibliographic reference using a simple database program

5. http://en.wikipedia.org/wiki/Social_bookmarking

6. Techcrunch (2009) Mendeley, the-Last.fm-of-research, could be world's largest online research paper database by early 2010. http://bit.ly/3i0ypP

7. Eva Amsen (2008) How to get scientists to adopt Web 2.0 technologies. http://blogs.nature.com/eva/2008/08/19/how-to-get-scientists-to-adopt-web-2-0-technologies

8. CiteFest citation tool Face-Off. http://citefest.pbworks.com/

9. Martin Fenner (2009) Bibliographic Management meets Web 2.0. http://blogs.nature.com/mfenner/2009/08/01/bibliographic-management-meets-web-2-0

10. Martin Fenner (2010) Reference manager overview. http://blogs.nature.com/mfenner/2009/03/15/reference-manager-overview

11. Innovations in Reference Management (2010) http://www.open.ac.uk/telstar/event

12. TELSTAR project blog. http://www.open.ac.uk/blogs/telstar/tag/irm10/

13. Martin Fenner (2010) There is still so much to learn in reference management. http://blogs.nature.com/mfenner/2010/02/22/there-is-still-so-much-to-learn-in-reference-management

Frank Norman is Librarian of the MRC National Institute for Medical Research in London.