Ten Simple Rules for Searching and Organizing the Scientific Literature

Denis C. Bauer[1]

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Qld. 4072, Australia

d.bauer@uq.edu.au

The exponentially increasing number of published papers (1.4 million per year by one estimate) makes it more and more difficult for us to manage the flood of scientific information. Each of us has acquired some protocol to find and organize journal articles and other references over the course of our careers. Most of those protocols are likely to have been formed by old routines or idleness rather than a structured approach to save time and frustration over the long run. Furthermore, with the *Web 2.0* revolution, new ways of handling information are emerging (O'Reilly 2005). For example, traditional standalone tools for reference management like *EndNote*[1] are being supplemented by centralized resources like *RefWorks*[2] and social bookmarking sites as described subsequently. This fusion of personal and public information offers the promise of efficiency through better organization, which in turn leads to better science.

How can seasoned scientists do better using these tools and those newer to the field start off in the right way? To start to answer that question, I present ten simple rules to master the search and organization of new literature. This is not meant to be comprehensive. It represents the experiences of a few and I welcome your thoughts, through comments to this article, on what you do to keep your references organized.

Rule 1: Facilitate an in-depth knowledge of your field.

A necessary prerequisite to scientific success is to have an in-depth knowledge of your field. That knowledge is contained in an ever growing network of scientific papers. Having identified seminal papers, typically by the number of times they are cited,

[1] http://www.endnote.com/ (commercial; Windows, Mac)

[2] http://www.refworks.com/(commercial, web service)

enables to explore the network for other frequently cited papers. Free tools are available to search for papers with related content, such as *Google Scholar*[3], *CiteSeer*[4] or *PubMed*[5]. A better user interface is provided by *ISI Web of Science*[6], for which most universities have a licence. There is the danger of a self-fulfilling prophecy, namely more cited papers get yet more citations. Useful papers can become orphans, while less relevant papers by the same authors get cited, so the tools should be used with some thought as to the content of the papers being identified by the tools. Which takes us to rule 2.

Rule 2: Identify papers, which are the most important to your research.

Categorize your papers into broadly related papers and "papers of extraordinary interest" to your research. One way to do this is using features offered by the reference manager. Papers of extraordinary interest will form the seeds from which you can mine for new literature. For you to find already published papers, most journals offer a list of records, which have cited your papers of extraordinary interest. Staying informed of new literature citing these papers can be done by creating a citation alert. Many journals, as well as article search engines (*PubMed, ISI Web of Science*), offer customisable alerts.

Rule 3: Stay on-top of the literature.

Creating citation alerts for the papers of extraordinary interest is a good start. However, relevant papers will likely still be missed. The *Faculty of 1000*[7] offers a service, where experts rank and comment on current research papers across journals in the field of medicine and biology in order to identify the most valuable publications. Though you can tailor the service to your interests, contribution to your research may not exclusively come from the cutting-edge papers reviewed by the Faculty of 1000. Checking for new literature manually is therefore recommended. This can be done by subscribing to the news feeds of your journals of interest. Using feed aggregators, such as *Google Reader*[8] or *FeedReader*[9], you are posted the papers in the latest issue automatically and can scan manually or filter automatically the titles for keywords to find interesting papers. Decide then and there if the paper is valuable to you. Staying organized means following through on rule 4 and 5 for every paper.

---

[3] http://scholar.google.com/ (free; web service)

[4] http://citeseer.ist.psu.edu/ (free; web service)

[5] http://www.ncbi.nlm.nih.gov/pubmed/ (free; web service)

[6] http://isiknowledge.com/ (commercial; web service)

[7] http://www.facultyof1000.com/ (commercial; web service)

[8] http://www.google.com/reader (free; web service)

[9] http://www.feedreader.com/ (free; Windows)

Rule 4: Archive digital copies.

Though it is more convenient to read a hard copy, digital copies are easier to organize, maintain and transport. Save the PDF version of the articles within a hierarchically organized folder structure and give the files meaningful names. Having a digital copy also enables you to apply search tools such as *Google Desktop*[10] to search the contents of your papers. This might give insight into cross links or simply identify the paper containing the phrase you want to quote. Keeping the PDF also enables cross linking with your citation program for offline work.

Rule 5: Archive references in a citation program immediately.

The usual routine, after finding a new record, is to print the paper and deposit it in the "to read" pile on the desk. While this might be a useful action leading you to actually read the paper, it is not sufficient in terms of organization. It is important to archive the reference to the paper immediately. Frequently, the source, from which the paper was downloaded, provides reference records formatted for programs such as *EndNote* or *Jabref* [11], supports online reference databases such as *RefWorks*[12] or *citeUlike*[13], as well as broader services like *del.icio.us*[14] or *Connotea*[15].

Rule 6: Label unread papers with what you expect from it.

If you are not reading the paper straight away you might forget why you thought it was important in the first place. Labelling the paper with what you expect to gain from it, or how you found the paper will make it easier for you to read it later on. All this can be done in your reference manager along with a flag highlighting unread papers.

Rule 7: Use tags and groupings.

Depending on what your citation program supports, tags or groupings are the main method to organize your reference bibliography. The challenge is to find a good trade-off between capturing the subgroups in the necessary detail and the depth of the organizing tree. The goal is not to summarize the content of the paper with a name but to group similar papers. For example, when searching for the citation for the PDB database you might find it more useful to open the group "resources and databases" and scan the content than to search by keyword for the authors or the exact title. If your citation program supports user-networks, tags are best used in an ontology-based form,

---

[10] http://desktop.google.com/ (free; Windows, Mac and Linux)

[11] http://jabref.sourceforge.net/ (free; platform independent)

[12] http://www.refworks.com/ (commercial; web service)

[13] http://www.citeulike.org/ (free; web service)

[14] http://del.icio.us/ (free; web service)

[15] http://www.connotea.org/ (free; web service)

where fixed vocabularies, e.g. GO or MESH, are utilized to cross link articles, comments and information from different users.

Rule 8: Write a literature review and summarize what was important to you.

Though time consuming, summarizing the content of a paper and noting what was important to you has two benefits. First, though the content appears easy to remember you will quickly forget about the details. Reading your summary of the paper will save you time later on. The second benefit is that summarizing and rephrasing the content forces you to understand the core message and to put it into the context of your own research.

Rule 9: Transfer all your notes to the digital copy and dispose the hardcopy.

Disposing (preferably recycle) the hardcopy will force you to archive everything properly. PDF-editing tools, such as *Acrobat Professional* or the free tools *skim*[16] and *PDFedit*[17], allow you to highlight sections and to add digital notes, which then can even be found by the search tools mentioned in Rule 4.

Rule 10: Create a group network.

These networks are not only a great way to find related papers that you may have missed, but also to build an online community to discuss papers, collaborations and ideas. Communities such as *citeUlike* and *nature network*[18] support the construction of such networks.

References

O'Reilly, T. (2005). "What Is Web 2.0." O'Reilly Network.

Acknowledgements

The author would like to acknowledge Philip E. Bourne, Stefan Maetschke and Fabian A. Buske for their insightful comments on these rules and their help in preparing the document.

[16] http://skim-app.sourceforge.net/ (free; Mac)

[17] http://pdfedit.petricek.net/index_e.html (free; Unix-like operating systems including Cygwin on top of Windows)

[18] http://network.nature.com/ (free; webservice)