

Reference Management

Martin Fenner, Kaja Scheliga and Sönke Bartling

*If I have seen further it is by standing on the shoulders of
Giants.*

— Isaac Newton

Abstract Citations of relevant works are an integral part of all scholarly papers. Collecting, reading, and integrating these references into a manuscript is a time-consuming process, and reference managers have facilitated this process for more than 25 years. In the past 5 years, we have seen the arrival of a large number of new tools with greatly expanded functionality. Most of the newer reference managers focus on the collaborative aspects of collecting references and writing manuscripts. A number of these newer tools are web-based in order to facilitate this collaboration, and some of them are also available for mobile devices. Many reference managers now have integrated PDF viewers (sometimes with annotation tools) for scholarly papers. Reference managers increasingly have to handle other forms of scholarly content, from presentation slides to blog posts and web links. Open source software and open standards play a growing role in reference management. This chapter gives an overview of important trends in reference management and describes the most popular tools.

M. Fenner (✉)
Public Library of Science, San Francisco, CA, USA
e-mail: mfenner@plos.org

K. Scheliga
Alexander von Humboldt Institute for Internet and Society, Berlin, Germany

S. Bartling
German Cancer Research Center, Heidelberg, Germany
e-mail: soenkebartling@gmx.de

S. Bartling
Institute for Clinical Radiology and Nuclear Medicine, Mannheim University Medical Center, Heidelberg University, Mannheim, Germany

Introduction

Reference management is perceived to be tedious and time consuming by many researchers, especially when it is done manually. In the past, references used to be written on index cards and stored in boxes. Now, reference management software allows for the digitalization of a personal collection of relevant scholarly publications. The earliest programs to manage the basic task of storing references and adding them to manuscripts have been around for over 25 years (including Endnote and BibTeX/LaTeX-based programs which are still popular today), but each individual entry had to be typed by hand. In the last 15 years we have seen a number of significant developments that have made reference management much easier for the researcher:

1. Retrieval of reference information from online bibliographic databases
2. DOIs and other persistent identifiers for bibliographic information
3. Automated management of PDF files
4. Open Access for easier access to full-text content
5. Web-based reference management for easier collaboration and use across multiple devices

In this chapter we describe what reference managers are and provide an overview of some reference management products. We do not make recommendations as to which reference manager may be the best as this is a personal choice and depends on the workflow of the individual researcher.

What is a Reference Manager?

A reference manager supports researchers in performing three basic research steps: searching, storing, and writing (Fenner 2010a). It helps researchers find relevant literature, allows them to store papers and their bibliographic metadata in a personal database for later retrieval, and allows researchers to insert citations and references in a chosen citation style when writing a text. To support those steps, a reference manager should have the following functionalities as identified by Gilmour and Cobus-Kuo (2011):

1. Import citations from bibliographic databases and websites
2. Gather metadata from PDF files
3. Allow organization of citations within the reference manager database
4. Allow annotation of citations
5. Allow sharing of the reference manager database or portions thereof with colleagues
6. Allow data interchange with other reference manager products through standard metadata formats (e.g. RIS, BibTeX)
7. Produce formatted citations in a variety of styles
8. Work with word processing software to facilitate in-text citation

A reference manager is a software package that allows scientific authors to collect, organize, and use bibliographic references or citations. The terms citation manager or bibliographic management software are used interchangeably. The software package usually consists of a database that stores references and citations. Once a citation is inserted into the database, it can be reused to create bibliographies which are typically found at the end of a scientific text.

Almost all reference managers allow direct importing from bibliographic databases through direct access from the reference manager and/or bookmarklets that import content from the web browser. Alternatively, references can be imported from other reference managers or from files in the BibTeX standard format with the help of import tools.

The reference database can then be searched, indexed, and labeled. Most reference managers offer tools for organizing the references into folders and sub-folders. Some reference managers allow the inclusion of full-text papers in PDF format. References can be shared via the Internet and organized into workgroups so that all members can use the same reference database.

Reference managers offer tools for exporting citations and references into word processing programs by selecting relevant items from the database. The citation style can be selected from a corresponding database which contains styles that aim to cover the requirements of a large number of scholarly publishers. Some reference managers allow for styles to be edited and saved.

There is a wide variety of reference management software, and the strengths and weaknesses of reference management software are perceived differently depending on the workflows of individual scientists. The deciding factor for a particular reference manager is often its popularity within a particular community, as collaboratively writing a manuscript is facilitated if all authors use the same reference manager (see chapter [How This Book was Created Using Collaborative Authoring and Cloud Tools](#)). Reference managers have been commercially available for a long time, but free solutions offer comparable functionalities and are increasingly gaining importance.

Some reference managers allow sharing, collaborative editing, and synchronization of reference databases across a private workgroup and/or publicly via the Internet. Public sharing of references is the focus of online-only social bookmarking tools such as CiteULike and Bibsonomy, but is also available with other reference managers. This functionality makes it possible to share Open Access papers online (see chapter [Open Access: A State of the Art](#)) and to generate usage statistics as a novel means of measuring scientific impact (see chapter [Altmetrics and Other Novel Measures for Scientific Impact](#)).

Getting References into the Reference Manager

All reference managers provide the functionality to manually enter bibliographic data. However, it is more convenient if the references are automatically extracted from an online bibliographic database such as Web of Science, Scopus, PubMed,

or Google Scholar. Most reference managers can also import references directly from a webpage, usually using information embedded via CoinS. All reference managers can import/export references in the BibTeX and/or RIS format; this is a convenient way to share reference lists with colleagues.

Bibliographic Databases

Some of the largest bibliographic databases (Web of Science, Scopus, and others) are only available via a subscription. In the last 10 years we have seen the emergence of an increasing number of openly available bibliographic databases. This trend started with PubMed in the late 1990s, includes Google Scholar, and, more recently, Microsoft Academic Search and the CrossRef Metadata Search, and now also includes bibliographic databases built by reference managers themselves (e.g. Mendeley or CiteULike). The availability of these databases increases the options for researchers to automatically import citation information, either via direct integration into the reference manager, or via a bookmarklet that captures the bibliographic content on the web page.

COinS: Hassle-Free Import of Bibliographic Data

ContextObjects in Spans (COinS) is a method that includes relevant bibliographic metadata of a scientific publication into the HTML code of a web page. If appropriate plugins are installed in a standard web browser, the bibliographic information of a reference can be easily retrieved by a reference manager, thus omitting tedious copy and paste processes. For example, if a reference is found in PubMed, a little symbol appears in the browser address line if the Zotero plugin is installed. At the click of a button, all important bibliographic information will be transferred into the Zotero database. Many scientific databases, scientific social networks, and journals support COinS (Fig. 1).

Digital Object Identifiers and Other Unique Identifiers

Most journal articles can now be uniquely identified by a digital object identifier (DOI). DOIs for journal articles are issued by CrossRef, a non-profit organization that has most scholarly publishers as its members. DOIs can also be used for other content, e.g. conference proceedings or book chapters. DataCite is another non-profit organization that can issue DOIs, focusing on DOIs for datasets. There are also other unique identifiers for scholarly content, e.g. the PubMed ID, PubMed Central ID, or the ArXiv ID. These identifiers make it much easier to handle

bibliographic information: reference managers can extract the DOI from imported PDFs, obtain more citation information using the DOI, store the DOI internally to help find duplicate records, etc. Authors only need to worry about the DOI (or other unique identifier), all the other information they need (authors, title, journal, link to the full-text) can be obtained from it.

Standardized Bibliographic Data Formats: BibTeX and RIS

BibTeX and RIS are the two most established file formats for storing bibliographic data, and one or both of these formats are supported by all reference managers. Exporting data in a standardized format is important because it allows users to backup their reference lists independently of the reference management software, to switch from one reference manager to another, or to use multiple reference managers in parallel.

- **BibTeX** has existed since the mid 1980s and was designed to be used in combination with the typesetting system LaTeX. The format is now widely supported by reference managers that work with Microsoft Word and other authoring tools, and by online bibliographic databases such as Google Scholar.
- **Research Information Systems (RIS)** is a standardized tag format originally invented by Research Information Systems (now part of Thomson Reuters). The format is widely supported and has been adapted over time, e.g. to include a field for digital object identifiers (DOIs).
- **Endnote XML and Citeproc JSON** are newer formats which are not yet as widely supported. BibTeX and RIS are plain text formats. XML and, more recently, JSON have evolved into the standard data exchange formats of the Web, and are easier to process automatically. They may therefore over time become the predominant formats for exchanging bibliographic information.

Citation Styles and Citation Style Language

Citations can be formatted in many different ways: what information to include (authors, title, journal, year, issue, pages), how to order and format this information, and how to reference these citations in the main text (e.g. by number or author/year). These so-called citation styles are important for printed documents, but are not really relevant for digital content (where citations are exchanged in BibTeX and other data formats). Unfortunately, most manuscript submission systems do not accept references in digital format, and authors are forced to format their references in the style requested by the publisher and include them as plain text at the end of the manuscript (and, in turn, publishers then spend time and money to get these references back into a bibliographic data format).

What publishers are really interested in are unique identifiers, such as the DOI, for all references. This allows them to double-check the reference information against

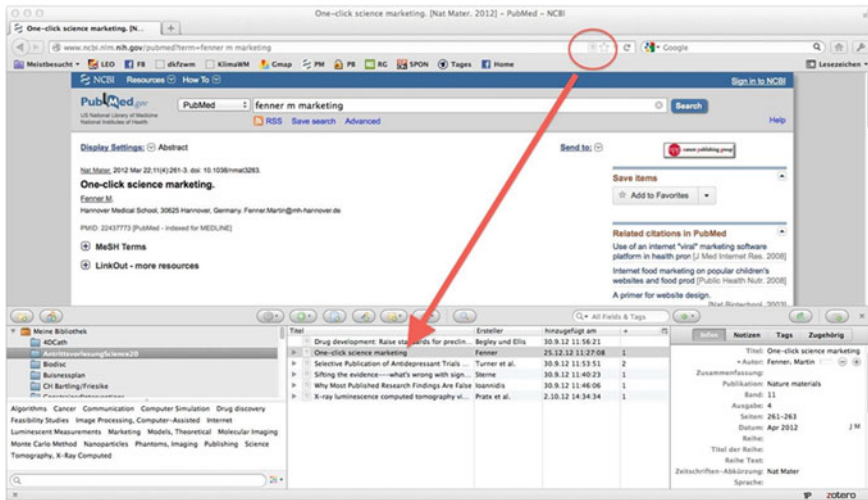


Fig. 1 Showing COinS in action. At the click of a button, a reference is included into the reference manager software (Zotero) from information that is contained in the COinS information in the displayed web page: No need to manually copy references

bibliographic databases (using tools such as eXtyle), and to format the citations into their preferred style. Citation information in free-text format can contain errors, and these errors are propagated if citations are entered manually (see Specht 2010).

Citation styles are needed not only to correctly identify all references (for which bibliographic formats and digital identifiers are better suited), but also to help the researcher while reading the text. Citations are an important part of all scholarly documents, and citation styles should facilitate that process. Even though a number of common styles exist (e.g. APA, MLA, Chicago, Vancouver), there is no standard style for citations in scholarly documents, and with the differences in citation practices between disciplines, it is not likely to ever happen. Some disciplines use simple reference lists at the end of the document, whereas other disciplines use footnotes at the bottom of pages and/or make heavy use of annotations.

Until it becomes standard practice to submit references in a bibliographic file format together with manuscripts (some publishers do this already), authors must resultingly deal with a large number of citation styles. This also often means changing the citation style when a paper has to be resubmitted to another journal. This is a time consuming endeavor, thus automating the process of adjusting to the various citation styles is an important feature of all reference managers.

Most reference managers support a large number of citation styles: EndNote¹ supports over 5,000 bibliographic styles, and Mendeley, Zotero, and Papers all support 2,750 citation styles. Some reference managers include a style editor, in

¹ EndNote Output Styles: <http://endnote.com/downloads/styles>

Examples of citation styles**DOI**

<http://dx.doi.org/10.1126/science.1197258>

shortDOI

<http://doi.org/dc3dhn>

APA

Wolfe-Simon, F., Blum, J. S., Kulp, T. R., Gordon, G. W., Hoeft, S. E., Pett-Ridge, J., & Oremland, R. S. (2011). A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science*, 332(6034), 1163-1166. doi:[10.1126/science.1197258](https://doi.org/10.1126/science.1197258)

Vancouver

Wolfe-Simon F, Blum JS, Kulp TR, Gordon GW, Hoeft SE, Pett-Ridge J, et al. A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science* [Internet]. American Association for the Advancement of Science; 2011 Jun 2;332(6034):11636. Available from: <http://dx.doi.org/10.1126/science.1197258>

Nature

Wolfe-Simon, F. et al. A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus. *Science* 332, 1163-1166 (2011)

BibTeX

```
@article{Webb_Weber_Davies_et_al_2011, title = {A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus}, volume = {332}, url = {http://dx.doi.org/10.1126/science.1197258}, DOI = {10.1126/science.1197258}, number = {6034}, journal = {Science}, publisher = {American Association for the Advancement of Science}, author = {Wolfe-Simon, F. and Blum, J. S. and Kulp, T. R. and Gordon, G. W. and Hoeft, S. E. and Pett-Ridge, J. and Stolz, J. F. and Webb, S. M. and Weber, P. K. and Davies, P. C. W. and et al.}, year = {2011}, month = {Jun}, pages = {1163-1166}}
```

RIS

TY—JOUR

T2—Science

AU—Wolfe-Simon, F.

AU—Blum, J. S.

AU—Kulp, T. R.

AU—Gordon, G. W.

AU—Hoeft, S. E.

AU—Pett-Ridge, J.

AU—Stolz, J. F.

AU—Webb, S. M.

AU—Weber, P. K.

AU—Davies, P. C. W.

AU—Anbar, A. D.

AU—Oremland, R. S.

SN—0036-8075

TI—A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus

SP—1163

EP—1166

VL—332

PB—American Association for the Advancement of Science

(continued)

(continued)

DO—10.1126/science.1197258

PY—2011

UR—<http://dx.doi.org/10.1126/science.1197258>

ER—

Citeproc JSON

```
{ "volume": "332", "issue": "6034", "DOI": "10.1126/science.1197258", "URL": "http://
dx.doi.org/10.1126/science.1197258", "title": "A Bacterium That Can Grow by Using
Arsenic Instead of Phosphorus", "container-title": "Science", "publisher": "American Asso-
ciation for the Advancement of Science", "issued": { "date-
parts": [[2011,6,2]], "author": [ { "family": "Wolfe-Simon", "given": "F." }, { "fam-
ily": "Blum", "given": "J. S." }, { "family": "Kulp", "given": "T. R." }, { "family": "Gor-
don", "given": "G. W." }, { "family": "Hoeft", "given": "S. E." }, { "family": "Pett-
Ridge", "given": "J." }, { "family": "Stolz", "given": "J. F." }, { "family": "Webb", "given": "S.
M." }, { "family": "Weber", "given": "P. K." }, { "family": "Davies", "given": "P.
C. W." }, { "family": "Anbar", "given": "A. D." }, { "family": "Oremland", "given": "R.
S." } ], "editor": [], "page": "1163-1166", "type": "article-journal" }
```

case a particular style is not yet supported. Citation styles used to be in proprietary format and owned by the publisher of the reference manager, but the Citation Style Language² (CSL) has evolved as an open XML-based language to describe the formatting of citations and bibliographies. Originally written for Zotero, CSL is now also used by Mendeley, Papers, and many other tools and services. In 2012, a web-based editor³ to create and edit CSL styles was launched, facilitating the creation of additional styles.

Managing Full-Text Content

Reference management has traditionally been about managing information about scholarly content (authors, title, journal, and other metadata). With the switch to digital publication and the availability of content in PDF, as well as other formats, reference management increasingly dealt with managing this digital content: linking references to the full-text document on the computer, performing full-text search, making annotations in the PDF, managing the PDF files on the hard drive, etc. Papers was the first reference manager to focus on this aspect, but most reference managers now have functionality to manage PDF files.

Most scholarly journal articles are currently distributed via subscription journals. This makes it important to store a copy on the local hard drive for easier access, but it can also create problems when these PDF files are shared with collaborators (which most publishers do not allow, even within the same institution). Reference management software therefore has to make decisions as to what

² Citation Style Language: <http://citationstyles.org/>

³ Find and edit citation styles: <http://editor.citationstyles.org/about/>.

is technically possible and convenient for researchers vs. what is possible under copyright law (see chapter [Intellectual Property and Computational Science](#)).

Content published as Open Access does not have these limitations. This not only makes it much easier to share relevant full-text articles with collaborators, but it also means that we often do not need to store a copy of the full-text on the local hard drive, as the content is readily available.

Reference Management Tools

From the large number of available reference managers, we have chosen seven popular products that are described in more detail below. We have included a table that gives an overview of their basic features. A feature list is not the only criterion in picking a reference manager though; ease of use, stability, price, and available support in case of questions are equally important factors.

EndNote

EndNote is a commercial reference management software package produced by Thomson Reuters. Endnote is one of the most popular reference managers and has been around for more than 20 years. It allows collecting references from online resources and PDFs. References from bibliographic databases can be imported into EndNote libraries. Full-text can be imported too. EndNote provides plugins for Microsoft Word and OpenOffice. References can be exported to BibTeX. While EndNote does not include any collaborative features, EndNote Web provides the functionality for collaboration with other users. Users can give group members read/write access to their references and import references from other people's libraries. Endnote also integrates with other bibliographic tools produced by Thomson Reuters, including Web of Science and ResearcherID.

Mendeley

Mendeley is a reference manager developed by a London based startup, but has been bought by Elsevier earlier this year. Its strength lies in its networking and collaborative features, and also in providing facilities for easily managing PDF files. It offers both a desktop and a web version with synchronized bibliographic information, allowing access from several computers and collaboration with other users. PDF files can be imported into Mendeley desktop and metadata such as authors, title, and journal are automatically extracted. It is possible to do a full-text search, highlight text in PDFs, and add sticky notes.

The web version recommends papers to users based on their profiles and the content in their libraries. Users can create both private and public groups and share

papers and annotations. Mendeley is free to use, but costs a monthly fee if the number of documents in Mendeley web or the number of private groups exceeds a limit.

Zotero

Zotero is a popular open source reference manager, originally developed as a plugin for the Firefox browser. The newer Zotero Standalone offers the same functionality but runs as a separate program and works with Firefox, Chrome and Safari. Zotero also includes a hosted version in order to synchronize references across devices and share them in private or public groups.

Zotero allows users to collect and to organize a variety of web sources such as citations, full-texts, web pages, images and audio files directly in the browser. Citations from Zotero can be integrated into Microsoft Word and OpenOffice.

RefWorks

RefWorks is a commercial web-based reference manager by ProQuest. The Write N Cite utility enables the integration of references into Microsoft Word where in-text citations and reference lists can be formatted into various styles. RefWorks makes it easy to collaborate with others as all references are stored in the web-based version. The Write N Cite utility can also work offline, but RefWorks is not the right tool for researchers with intermittent or poor Internet connectivity.

Papers

Papers is a commercial reference management software, now part of Springer Science+Business media. Initially Papers was only available for Mac, but now there are also versions for iPad and PC. Its main strength is its excellent handling of PDF documents (including metadata extraction) and its polished user interface, whereas the collaborative features are less developed than in some of the other products. Papers uses the Citation Style Language and provides a word processor plugin.

JabRef

JabRef is an open source bibliography reference manager popular with LaTeX users. It runs on Java and is thus compatible with Windows, Linux, and Mac. The native file format is BibTeX which is the standard LaTeX bibliography format. The strength of JabRef is that references can be formatted directly in LaTeX, thus

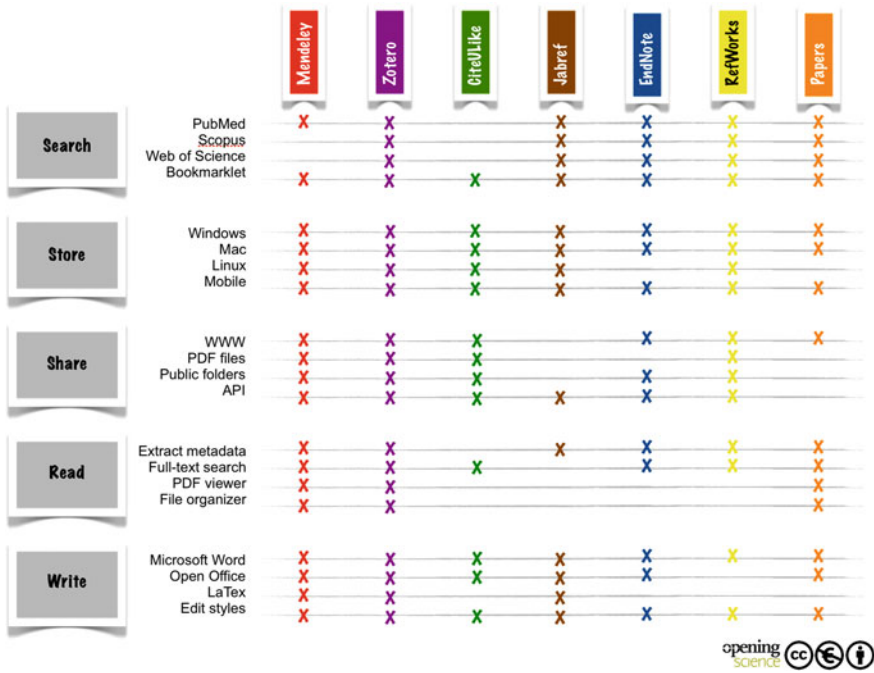


Fig. 2 Feature comparison of popular reference managers (see also Fenner 2010b)

providing access and control over a wide range of citation styles. JabRef provides direct search and downloads from PubMed and IEEEXplore. There are plugins for word processing programs and also other Java based plugins, which expand the general functionality.

CiteULike

CiteULike is a free online reference manager and social bookmarking tool. References are primarily entered via a bookmarklet that captures bibliographic content in web pages. New entries are public by default and are added to the common library, but entries can be also made private. Users can assign tags to entries which make it easier to organize and search through content. References can be exported in BibTeX and RIS formats. The social networking features are the strength of CiteULike. Users can create profiles, connect with other researchers, and create and join groups where they can collaborate on library content.

Other Reference Management Products

Many other reference managers are available, including Citavi which is popular in some disciplines and also helps with knowledge management, and ReadCube which has a very nice user interface and a good PDF viewer. An extensive list and comparison of available reference management software can be found on Wikipedia (Fig. 2).⁴

Outlook

Reference management has become easier, cheaper, and more social in the past few years, and this trend will continue. We will see the integration of unique author identifiers (ORCID, etc.) into bibliographic databases and reference management tools (see case in chapter [Unique Identifiers for Researchers](#)), and this will facilitate the discovery of relevant literature and the automatic updating of publication lists. We will increasingly see citations of datasets and other non-text content (see chapter [Open Research Data: From Vision to Practice](#)). Digital identifiers for content and support for the open Citation Style Language will also increase, as will the availability of open bibliographic information. Three areas still need improvement. Firstly, the automatic importing of the references of a particular publication, and the integration of reference managers into authoring tools. Secondly, the word processor plugins for reference managers still do not work together, and some of the newer online authoring tools (Google Docs, etc.) need to be better integrated with reference managers. Finally, instead of having references in plain text, which makes it difficult to get to the full-text and reformat it into a different citation style, publishers, institutions, and funders should start to ask for reference lists in standard bibliographic formats using digital identifiers.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Fenner, M. (2010a). Reference management meets Web 2.0. *Cellular Therapy and Transplantation*, 2(6), 1–13.
- Fenner, M. (2010b). Reference manager overview. *Gobbledygook*. Available at <http://blogs.plos.org/mfenner/reference-manager-overview/>.

⁴ Comparison of reference management software: http://en.wikipedia.org/wiki/Comparison_of_reference_management_software

- Gilmour, R. & Cobus-Kuo, L. (2011). Reference management software: A comparative analysis of four products. In *Issues in Science and Technology Librarianship*. Available at <http://www.isfl.org/11-summer/refereed2.html>.
- Specht, C. G. (2010). Opinion: Mutations of citations. Available at <http://www.the-scientist.com/?articles.view/articleNo/29252/title/Opinion-Mutations-of-citations/>.