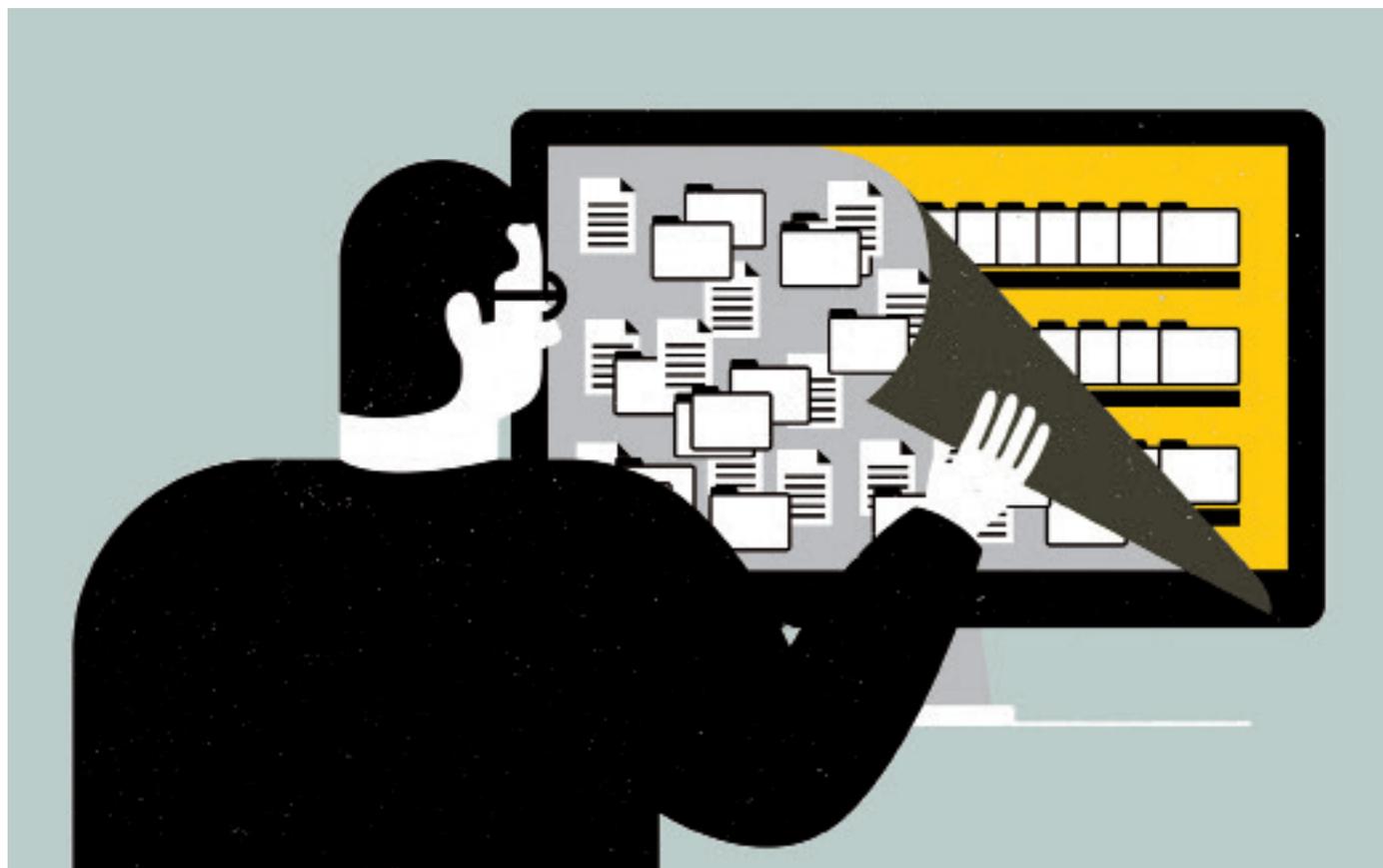


EIGHT WAYS TO CLEAN A DIGITAL LIBRARY

Scientists have a surfeit of options to choose from in the competitive market of reference-management software.

ILLUSTRATION BY THE PROJECT TWINS



BY JEFFREY M. PERKEL

Adam Rocker didn't expect the software that managed his digital reference library to flag up better ways he could be doing his research. But his electronic filing system of choice, ReadCube, periodically scans his library and suggests related papers, rather as some music-file-management programs highlight recommended tunes. And that feature, he says, has brought up some unexpected gems.

As a graduate student, Rocker, who is now studying medicine at the University of Ottawa, was researching bacterial infections in zebrafish. ReadCube highlighted a paper that described a way to entrap the fish using

microfluidics — a field whose literature he would not normally read — that was much easier than his own method. Being alerted to the research was “really rewarding”, Rocker says, although he was ultimately too invested in his own project to adopt the alternative approach.

As Rocker discovered, today's reference-management tools go above and beyond simple electronic filing. Rather like a Swiss-army knife, each tool now appeals to customers by offering an ever-evolving set of extra features.

This article focuses on eight tools — colwiz, EndNote, F1000Workspace, Mendeley, Papers, ReadCube, RefME and Zotero — all competing in the reference-management market (see ‘Reference-management software’). Some

excel at streamlining the process of browsing and building literature libraries, whereas others focus on creating bibliographies, aiding collaboration through the use of shared workspaces or recommending papers. (One, ReadCube, is owned by Digital Science, a firm operated by the Holtzbrinck Publishing Group, which also has a share in *Nature's* publisher.)

Each tool exists to help researchers to tame the digital flotsam and jetsam of scattered, downloaded PDFs. Most scientists can relate to that problem: as they grab PDFs from journal websites — where they are often assigned impenetrable alphanumeric codes as filenames — and dump them into any convenient folder, chaos can quickly take hold, with multiple ►

► copies of files spread across hard disks.

“In science, or at least in my experience, we tend to end up with a folder in the desktop with 3,000 really weirdly named PDF files, which we can never find when we need them,” says Raúl Delgado-Morales, a neuroscientist at the Bellvitge Biomedical Research Institute in Barcelona, Spain.

Reference-management tools address that confusion by indexing a hard disk. Typically, the process of dragging and dropping a PDF into an application window triggers the software to try to identify it using the DOI or title, and to retrieve relevant metadata (such as title, keyword and author names) from online servers.

Researchers can also assign software to monitor specific folders into which they drop their files. They can then find PDFs through a simple search for author name, keyword or, in some cases, their own notes. Delgado-Morales solved his problem, for example, by organizing his literature library with Papers, a user-friendly application that automatically renames files according to any scheme he chooses. Other tools offer similar functions, except for RefME — a website and mobile app — which stores only lists of references and not the PDFs themselves.

CORE FUNCTIONS

Most of the tools help researchers to import literature from a variety of online sources. Many offer in-app searching of external databases such as PubMed and Google Scholar, as well as web-browser plugins that grab reference data (and sometimes, associated PDFs) from journal websites and other pages.

“We tend to end up with a folder with 3,000 really weirdly named PDF files.”

Zotero — a free, open-source software project — was founded ten years ago specifically to tackle the problem of extracting information from a web browser, says project director Sean Takats of George Mason University in Fairfax, Virginia. “That’s the key feature of Zotero, and remains one of its strongest compared to other reference managers,” he says. RefME offers the unusual option of adding references by scanning a barcode with a smartphone camera.

One of the best-known features of reference-management software is the ability to insert in-text references in a research paper and to create bibliographies in any format. EndNote, a widely used commercial package, has offered this feature for decades, but now faces competition from many modern tools.

Many tools interface with common word-processing software (usually Microsoft Word, but sometimes OpenOffice and related free-ware suites as well) so that a user typing up a research article need only select the papers that they want to mention and click a button to have codes inserted into the document to mark

REFERENCE-MANAGEMENT SOFTWARE

Eight of the most popular tools.

Product	URL	Platform	Free?
colwiz	colwiz.com	Desktop/web/mobile	Yes
EndNote	endnote.com	Desktop/web/mobile	Yes, with some limited features
F1000Workspace	f1000.com/work/	Web	No
Mendeley	mendeley.com	Desktop/web/mobile	Yes, with some limited features
Papers	papersapp.com	Desktop/web/mobile	No
ReadCube	readcube.com	Desktop/web/mobile	Yes, with some limited features
RefME	refme.com	Web/mobile (only stores references)	Yes
Zotero	zotero.org	Desktop/web/mobile	Yes, with some limited features

See the online version of this article at go.nature.com/xbp9ot for a fuller comparison.

the in-text reference. Later, the user can create a bibliography and in-text citations according to several thousand journal styles, picking his or her choice from a pull-down list.

Most tools include built-in PDF readers for reading and annotating articles — typically allowing users to search through comments and notes — as well as cloud-based capabilities for syncing those comments (and the PDFs themselves) between, for example, an iPad and a desktop computer. But ReadCube and colwiz try to offer richer PDF reading experiences. In ReadCube, for instance, in-line citations and author names in PDFs are rendered as active hyperlinks to provide direct access to cited articles and publication lists. The same functionality is available when viewing and annotating PDFs on the websites of partnering publishers (including, for ReadCube, *Nature* and Wiley; and, for colwiz, Taylor & Francis).

Many of these tools can identify articles related to specific items in a library, or recommend articles on the basis of the library’s content overall. F1000Workspace — like ReadCube — uses an algorithm to do this. It also taps into recommendations made by a community of 10,000 or so specialists. However, many other stand-alone software products also recommend papers (see *Nature* 513, 129–130; 2014).

SET TO SHARE

Many tools now allow researchers to set up group libraries or share key papers with distant collaborators, although this process is carefully managed to prevent violation of publishers’ copyright. Those in public groups using Mendeley, for instance, can share only information about a paper — the equivalent of a library-catalogue entry. Only users in private groups can share and modify PDFs (and groups must upgrade to a paid account to add more than three individuals).

Brenton Wiernik, an organizational-psychology PhD candidate at the University of Minnesota in Minneapolis, uses a shared library in Zotero for collaborative projects involving systematic reviews and meta-analyses of the literature in his field. Such efforts might involve 15–20 people, he says: some downloading

articles into a shared library; others reading them; still more adding annotations and tags and logging key data.

According to Wiernik, the process is akin to using a shared Dropbox folder, with the added benefit that Zotero tracks and maintains metadata, notes and annotations. For instance, researchers can use a dedicated tag to indicate that they are processing an article, thereby signalling to collaborators that they should work on a different article to avoid duplicated effort.

F1000Workspace and colwiz both extend sharing to include features for preparing manuscripts and managing projects. With F1000Workspace, researchers can use a plugin to upload Microsoft Word manuscripts to a secure location, thereby enabling team members to comment on the shared copy — although the text cannot be edited in the browser, says João Peres, the company’s product-development manager. Peres plans to implement a ‘one-click’ article-submission feature that sends papers directly from F1000Workspace to journal editors, starting with the journal *F1000Research*. And colwiz also permits users to share documents to an online drive for team members to view and comment on.

Given the highly overlapping feature sets of these tools, a user’s choice often comes down to particular individual priorities. Richard Karnesky, a materials scientist at the Sandia National Laboratories in Livermore, California, supports Zotero for its open-source ethos, for example.

Perhaps the best reason for using a reference manager is the technology’s ability to provide a form of searchable memory. Imagine, says Boyd Steere, a senior research scientist at pharmaceutical firm Eli Lilly in Indianapolis, Indiana, a desk piled high with printed papers: Post-it notes hanging out, writing in the margins, doodles, notations, arrows and more. Today’s PDF-filled, digital folders are in many ways no easier to navigate. With a digital reference manager, however, buried knowledge is just a keyword search away. ■

Jeffrey M. Perkel is a writer based in Pocatello, Idaho.