

## PRODUCT REVIEW / ÉVALUATION DE PRODUIT

**Product:** Colandr

**Purpose:** Systematic review software

**URL:** <https://colandrapp.com/>

**URL community:**  
<https://www.colandrcommunity.com/>

**Cost:** Free

**Bottom line:** Colandr is a free product that uses machine learning to facilitate screening and data extraction for comprehensive literature review projects. Although Colandr currently has limited functionality, it has the potential to become more useful as it develops further. As an open source application, whose code can be used and adapted for other communities or purposes, Colandr contributes to efforts to create efficiencies in executing certain tasks for comprehensive literature review projects.

### Purpose and intended audience

Screening large numbers of citations to identify relevant articles, and extracting data from accepted articles, can be time- and labour-intensive processes. Colandr uses text mining and machine learning to automate specific aspects of screening and data extraction, making some tasks related to systematic reviews more efficient. Developed in 2017, Colandr is a collaborative project between the Science for Nature and People Partnership, Conservation International, and DataKind (SNAPP, n.d.). The original purpose of this tool was to support the needs of environmental and wildlife conservation researchers conducting large-scale reviews of research literature. According to the developers, Colandr now houses projects by researchers representing a variety of fields, including conservation, medicine, and education (Augustin, 2018).

### Product Description and Cost

Colandr is a free, browser-based product. Researchers register for an account, create a project file, and enter details including the research question, search terms, inclusion and exclusion criteria, and data extraction fields. Colandr uses this “planning phase” information to highlight relevant terms in titles, abstracts, and full text for quicker screening. At the full text review and data extraction stages, Colandr uses information gleaned from the planning and screening stages to compile and display potentially relevant excerpts of each full text article.

### Features

I took a test drive of Colandr, using a previous systematic review project I had collaborated on. Unfortunately, I could not test beyond Colandr’s Planning Stage, as the program would not import references, despite my making multiple attempts, using a variety of browsers, and seeking assistance from the developer. Accordingly, the planning section of this review is informed by hands-on experience, but the screening and data extraction aspects of the product are based on information gleaned from available documentation and email communications with the Colandr team.

Project workflow is organized into planning, citation screening, full-text screening, and data extraction phases (Figure 1). In the planning phase users enter their research question as a sentence, as well as in PICO format, and list keywords for each search concept (called “key terms”). Colandr generates a basic Boolean search query exclusively from the user-provided keywords. Unfortunately, the search string generated connects all concepts with OR, so is not currently usable (Fig. 2).

Next, users enter inclusion and exclusion criteria for use during screening phase, add metadata fields for use during the data extraction phase, and then import citations for screening. According to Colandr’s FAQ,



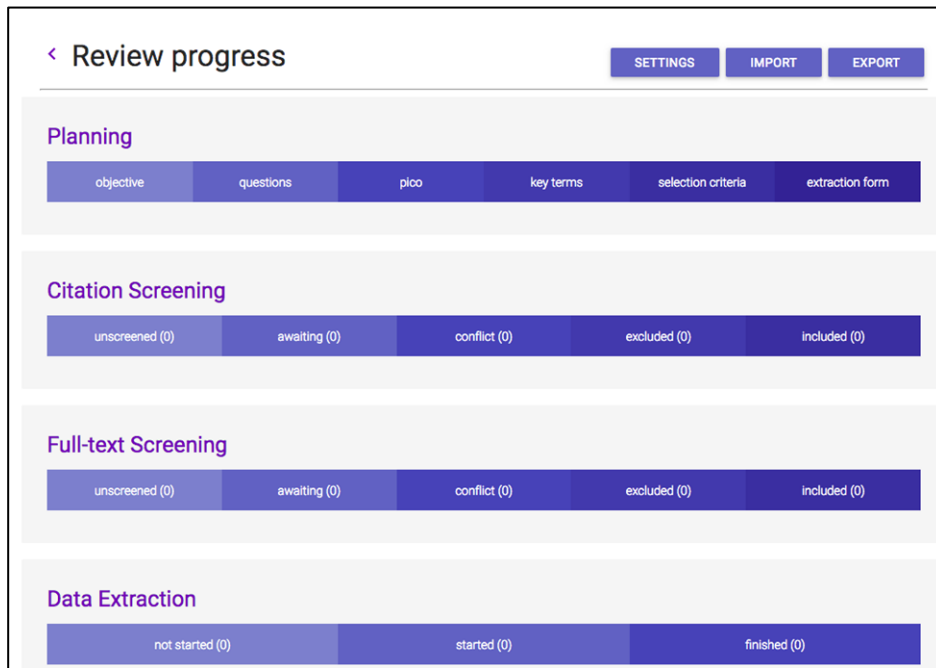
This article is distributed under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>

the program can import .bib, .ris, and .txt files, with a maximum file size of 40 megabytes (typically 15,000-20,000 citations). The total project size is unlimited.

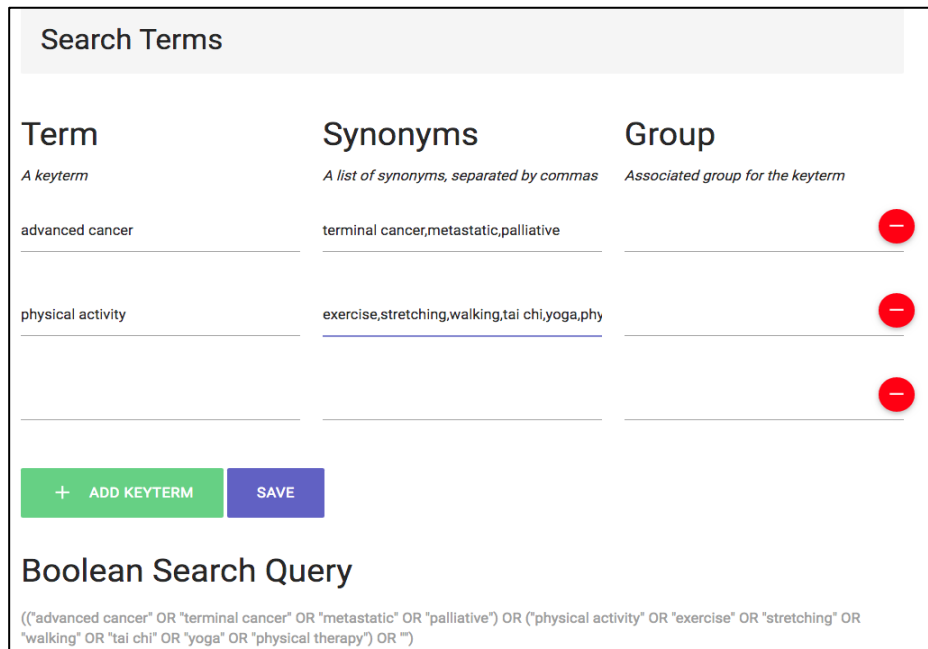
Colandr automatically deduplicates references on import, but it is unclear what the deduplication criteria

are, or whether a user can review or override duplicate removals. Additional user-identified duplicates cannot be removed from a Colandr project.

**Fig. 1.** Colandr project workflow



**Fig. 2.** Planning phase: Search terms and auto-generated Boolean search query



Following title and abstract screening, users import PDFs of full text articles, one at a time, into Colandr for full text screening, review and data extraction. During the citation and full-text screening phases, Colandr searches the title, abstract, and author-supplied keyword fields of the imported citations for the user-provided keywords. It then highlights these terms in the citation, abstract, and full text. Once at least 50 full-text items have been reviewed, Colandr looks for potentially relevant blocks of text within an article and collates them. The idea is that reviewers can focus on those excerpts, reducing the time spent poring over the whole document looking for relevant information. This aspect of Colandr is meant to be highly sensitive, to retrieve potentially relevant content to reduce the likelihood of missing relevant items (Augustin, 2018).

Researchers can export their data extraction files in .csv format; a summary of metrics is provided at time of export (e.g., number of articles imported, accepted, excluded organized by reason for exclusion).

#### *Additional feature*

- Colandr can list citations by expected relevance. Ranking is initially based on the user-supplied keywords from the planning phase (Colandr checks title, abstract, keyword fields of imported citations), but is dynamic—it changes as more items are screened and as data extraction occurs, because Colandr also uses the included/excluded articles to determine relevance.

### **Platform, usability and compatibility**

Colandr is a web-based application, with no local client install. It is platform- and browser-agnostic, and works on mobile devices running iOS.

Training and support resources are very basic. Clicking on Help sends users to the Colandr Community website (<https://www.colandrcommunity.com/>). The Training section includes a slide presentation with an annotated screenshot tour of Colandr. An informal video walkthrough, FAQ, and sparsely populated Google+ community forum/bulletin board for posting questions are also available. A Report a Bug feature is available on the Colandr Community site, but is not accessible from within a Colandr project.

### **Strengths**

1. Open access and open source product—free for anyone to use. Other developers are also able to examine or reuse it, and enhance its code.
2. Browser- and platform-agnostic

### **Weaknesses**

1. Limited to 2 screeners or reviewers per project
2. As of this review, could not import references in .ris and .txt format, even with developer intervention
3. Projects with less than 50 accepted articles cannot take advantage of the machine learning feature for data extraction
4. Missing contextual help buttons and field validation criteria to guide researchers as they enter information; small missteps could lead to data not being saved and having to be re-entered
5. No bulk PDF upload

### **Conclusion**

Colandr is a young product that would be worth reviewing again, once the above-noted limitations have been addressed and the user interface has had time to mature. By sharing their code, the developers of this program have made an important contribution to efforts to automate labour-intensive aspects of comprehensive literature review projects.

### **References**

1. Augustin, C. (2018, April). Finding needles in the evidence haystack. Smart sorting for conservation decision making - Users [PowerPoint slides]. Retrieved from <https://www.colandrcommunity.com/updates/datakind-talk-at-cee-2018>
2. SNAPP Science for Nature And People Partnership [Internet]. SNAPP team: Evidence-based conservation. Available from: <https://snapppartnership.net/teams/evidence-based-conservation>

### **Statement of Competing Interests**

No competing interests declared.

#### **Maria C. Tan**

*Public Services Librarian*

*Scott Health Sciences Library, University of Alberta*

*Email: [maria.tan@ualberta.ca](mailto:maria.tan@ualberta.ca)*