Questions and answers from Colandr webinar on 6/6/2018
Hosts: Caitlin Augustin and Sam Cheng
Contact: colandrteam@gmail.com

This presentation is being recorded and we will be posting slides. Earlier versions of these slides from CEE are available on colandrcommunity.com
Training materials for creating a test review are available at:
https://drive.google.com/drive/folders/0B62GXhCvxnEaVlJ3cHJfbnNyZUk
FAQs are available at:
https://docs.google.com/document/d/1wysfnmIBRkZOKdxKY5Hx_pg7X6nWuA8bjw5o-KFfo6Q/edit
More in-depth discussion of the algorithms behind this tool are available here from Caitlin's talk at CEE: https://www.colandrcommunity.com/updates/datakind-talk-at-cee-2018

Caitlin Augustin : we are also monitoring twitter if you want to tweet at sam @pilesofsquid or me @augustincaitlin with a #colandr hashtag

Q: I'm sure the answer is "it depends", but do you have a general recommendation for the size (# articles) for the training set for both system 1 (identifying articles) and system 2 (data extraction)?
A: Colandr starts learning in the screening stage (system 1) after 10 articles have been screened for inclusion/exclusion. System 2 requires that the user manually label a minimum of 35 articles before it will start learning.

Q: Does Colandr only work for hard sciences, or can it also be used for social sciences?
A: Colandr works for all topics

Q: Thanks! Does it learn from "rejected" articles too?
A: Yes, it learns from included and excluded citations.

Q: Is there a way to delete reviews?
A: To delete a review, go into the review, go to settings, scroll to the bottom, and find delete review

Q: How many collaborators maximum can you have for a review? In other words, if you have 2 screeners do they have to be the same individuals? (I'm working on a team project with 4 people...)
A: Reviews can have unlimited users. Reviewers for citations can be any two individual screeners.

Q: What is meant by "group"? (in the planning stage for search terms)
A: Devising a search string involves choosing key terms and their synonyms (e.g. well-being, wellbeing, well being). However, often you will wish to search for many alternative key terms in

a group. For example, in a group for outcomes - you may wish to search for (abundance OR richness OR recovery OR diversity). You can connect groups of key terms with an AND. Within the planning window, you can assign key terms to these different groups.

Q: I don't need sistematic review, I need scoping review, for that, i don't have PICO. It is a problems here?
A: no - PICO and objective are just for your organizational purposes and do not inform the machine learning models

Q: Is the Boolean Search Query pre- or post-scoping?
A: You typically iteratively develop your search query as you determine which terms are useful and which are not during the scoping stage.

Q: when i add terms not in consecutive order of their grouping term, the boolean puts them together with AND? is this right?
A: Yes, unfortunately, if you add terms to groups not in consecutive order, it will group them with AND instead of adding them to the terms within a group's parentheses. To ensure you do not have terms appearing in the wrong order, you should add your terms in order of group.

Q: does the search builder offer any support for using controlled vocaulary, such as MeSH headings for use in Medline?
A: No, this does not provide explicit support for MeSH vocab

Q: Can you manually import files so that they can be deleted
A: You cannot delete an import! Apologies

Q: Can you manually upload citations rather than use RIS download
A: Only citation file formats are supported

Q: Since you can't delete uploads, can uploaded things be tagged as "ignore"?
A: You could exclude them with a specific tag (e.g. "ignore") as a workaround

Q: Is there an opportunity to review the identified duplicates? And how does the system identify duplicates across different imports?
A: You cannot review/approve/reject duplicates, but the algorithm used is very conservative and must be an exact match which can mean different formatting from different sources might not be flagged. The deduplication process time can vary depending on volume. Removed duplicate records will appear in the coding record you export

Q: Are you able to highlight irrelevant keywords separately to relevant keywords to quickly scan if a citation is irrelevant for the study?
A: only keywords that appear in your search string will be highlighted in the citations. However, you can always add a tag for irrelevant keywords

Q: I may have missed this, but how do you create specific tags/what do the tags in this review mean (I.e. tag1 v. tag2)?
A: tags can be added to a citation at the bottom of each citation - but I believe you are asking about the inclusion/exclusion criteria. you will need to define those in the planning stage as why you would exclude the criteria

Q: Sorry if I missed this—is something flagged as a disagreement if both reviewers exclude the record but for different reasons?
A: citations in conflict disagree in inclusion/exclusion but not for reason for exclusion but you will be able see differences in exclusion reason in the exported coding record

Q: can you separate that training set from the newer studies in the update? (i may have missed this)
A: you cannot separate the training set from the newer studies (but if the training set is already included, you will only be screening the unscreened new ones)

Q: Can users pre-assign studies unique IDs (for consistency with older review projects)?
A: no you cannot assign IDs to papers, but you could add them as a tag (albeit tedious)

Q: Can you exclude a study at title and abstract without having to indicate which reason it was excluded on?
A: No, you must choose a reason

Q: Just to confirm that I'm following, it sounds like we can't go straight to full-text screening, correct? That can only be applied to papers that have made it through the citation screening? I'm interested in using this for searching federal land management planning docs that won't have abstracts and such, so might not work as well for the citation screening part.
A: Yes, you cannot skip to full text screening without doing the citation screening, however, you could upload the full data set as "included" which would speed up your process

Q: how do you filter by tag? after adding tags they don't seem to appear in the filter…
A: You will have to refresh the page to see the tags in the filter by: dropdown menu

Q: Can you batch upload prior extracted data for training system 2 (full text)?
A: Unfortunately not, but you can manually enter the data if you already know it as the training set

Q: could you please post how we access this visualisation tool?
A: Evidence for Nature and People Data Portal - www.natureandpeopleevidence.org (but we are currently relaunching a new development, so the website will be available starting next week). Github link for Evidence for Nature and People Data Portal -
https://github.com/scheng87/ebc_portal where you can access the code

Q: we are using colandr in a study - what would you need from us
A: if you are interested in testing the performance of colandr - please contact us directly at colandrteam@gmail.com, otherwise, please be sure to cite our paper in Conservation Biology (link is on the community website)

Q: can you customize data extraction? e.g. can I export paragraphs that contain certain keywords?
A: Yes, in the data extraction stage, you can copy and paste text from the PDF into one of your data extraction fields (be sure to predefine it as "text") however, it will not automatically export paragraphs around certain keywords

Q: Loving that this is a free tool! Are there any plans to shift to a paid model?
A: currently no plan to shift to a paid model. DataKind is firmly an open-source, open-access company :)

Q: for allowed values, do you have to list all synonyms? ie tree and trees and plant?
A: (in data extraction) - no you do not - these are merely how you choose to assign your papers to categories

Q: For full text, can it handle XML? (I have docs that I've OCR'd with XML output.) or Word docs?
A: No, only PDF