An Activity Module: Creating a Citizen Science Reader

Images and media can be provided.

By Team CEVOpen, GitHub: [CEVOpen](#)

The module is designed to enable a hands-on activity for the easy use of modern Open Access infrastructures involving – finding, using, and sharing scientific literature. CEVOpen has developed open-source software to search across multiple research literature repositories such as Europe PMC or bioRxiv (pronounced "bio-archive") enabling the creation of an automated [Literature Survey](#) within minutes – presenting the user with a summary of findings and allowing the download of the full articles.

CEVOpen is an example of an *open search* framework based on text data mining (TDM). Having 'open search' systems is important as search engines are the gateway to scientific knowledge. Search engines can be *gamed* to bias certain outcomes or be based on faulty algorithms (Kraker 2018). CEVOpen is built using Open Science methods so all parts of the system are open and verifiable, even down to a specific 'search results' query.

Who is the activity module for?

The module is for researchers using Citizen Science in their research project to add a module to engage the participants in conversations in formulating research questions and consulting on what is known about a topic in the existing scientific literature corpus. The activity starts with the participants creating a simple ten word dictionary of terms related to their topic and a bot retrieving a sample of 100 papers from Open Access repositories.

Infobox: About CEVOpen open-source software

CEVOpen is a project that aims to develop knowledge resources and tools to help tackle the research questions on a global scale – viral epidemics, climate change – down to niche questions.

Despite over $100 Billion being spent on medical research worldwide, much knowledge is behind publisher paywalls. Moreover it is usually badly published, dispersed without coherent knowledge tools. This particularly disadvantages the Global South. The project aims to use modern tools, especially Wikidata (and Wikipedia), text mining, with semantic tools to create a modern integrated resource of all current published information on viruses and their epidemics. It relies on collaboration and gifts of labour and knowledge – to find out more see the [Getting Started Guide](#) or the [How Can I Help](#) section.

CEVOpen practices [Open Notebook Science](#) which means there is no insider knowledge and all work is open and licenced for the freedom of reuse.

Video link: Example short video showing an application of CEVOpen.

The activity

Image: A simple one slide image summarising the process? A, B, C, D. Review > Dictionary > Summary > Review (repeat).

The activity will involve using the '5 minute literature search' from CEVOpen – this involves using a web browser to search literature repositories and sort the results to get up-to-the-minute relevant research related to the questions asked.

The activity can be used to share results on a public webpage and to update the search as often as is required – say once a week.

Note: link to an example existing research results and dictionary.

A Step-by-Step Guide

In this example the Citizen Science project is looking at the topic of 'zero-carbon plans'.

Note: Can we create and host this example search? or have a list of other searches already done?

CEVOpen carries out two types of search:

- Firstly, it searches repositories on the net and retrieves the papers;

- Secondly, it then analyzes the local *full text* copies of the papers that have been downloaded. The result is a swift and verifiable literature survey that might have once taken days, weeks, or months to complete if done manually.

1. The researcher (includes citizens) formulates a question they are interested in,
    i. e.g. What 'zero-carbon' plans for tackling the problem of climate change are reliable enough for further adoption in cities or regions around the world, for example in: green energy, transport, and housing? Many policies are being implemented and rely on such unverified plans, for example the EU's 'A European Green Deal' but if we want to ensure the public, industry, and governments buy into them – such plans must be 'Open Science proof' AKA open and available for scrutiny – verifiable, reproducible, and reusable.
    ii. e.g. A simpler version of the same question could be 'What zero-carbon plans can be used for the future of my local schools, city public transport, or municipal buildings, etc.'?
2. From the 'research questions' a dictionary of terms important to the topic need to be made. Ten dictionary terms is a good start. These terms are then input into CEVOpen in the browser and it goes off and collects the top one hundred research papers from your repository of choice – we use Europe Pubmed Central as the default as it aggregates many other sources, but many other literature repositories could be used.
    i. Our example dictionary terms for 'zero-carbon' plans would be: rapid decarbonisation; zero-carbon; low-carbon; energy planning; decarbonisation; low energy transport; policy and planning; policy; low energy housing; low energy city planning; low energy schools.
3. CEVOpen local search on the downloaded papers after giving the papers a scan read. The local searches can be focused on paper sections – introductions,

findings, etc., or on content types illustrations or tables – informed by what is thought to be the most yielding in the papers.

4. Refine and repeat depending on what looks useful. The dictionary of initial search terms should be updated as well as reviewing the local full-text search.
5. CEVOpen downloads the full-text of the papers, as well as PDF copies. It also makes a summary of the results of the frequency of the terms. The whole *search results* package can then be published and shared online.
6. In addition Wikidata can be used in relation to the dictionary terms being used and this allows for more advanced semantic queries to be carried out as well as being able to retrieve multilingual Wikipedia pages of say English terms used in papers.

Acknowledgements

TBC

References

Kraker, Peter. 2018. "Illuminating Dark Knowledge." Text. *Generation Research* (blog). December 3, 2018. https://doi.org/10.25815/FN1Q-KC98.