Towards open collaboration in biomedical and health research: An open collaborative scoping review

Mario Garcia
Copenhagen University

Daniel B. Ibsen
Steno Diabetes Center Aarhus, Aarhus University

Luke W. Johnston
Steno Diabetes Center Aarhus, Aarhus University

2025-01-08

Table of contents

Introduction

- Open science and its role in research
- Collaboration and its role in modern research
  - collaboration spanning from lab, to research center to multiple centers
  - collaboration needed to tackle limitations: small populations, lack of replication, etc
  - re-inventing discovery: a new way of working in science

Scientific research now almost always requires working with other people.

With the growing complexity and specialization in scientific practices and methods, together with globalisation of health and environmental issues, there is a great need for a paradigm shift in research collaboration to be able to tackle these challenges.

- Open collaboration - a combination of two themes
  - definition
  - big potential in science
  - but there very few resources and examples of how to integrate open science into collaborations

We define open collaboration using the definition as found in (1):

> "*an online environment that (a) supports the collective production of an artifact (b) through a technologically mediated collaboration platform (c) that presents a low barrier to entry and exit and (d) supports the emergence of persistent but malleable social structures.*"

With the increasing emphasis on and demand for science to be more open, how we collaborate together is a key component to making science more open from the start of any project. But how do we collaborate in an open and transparent way? What are the best practices and tools we can use? What is an ideal collaborative workflow and how close or far are we from this ideal in reality?

Aim

This scoping review will focus on current practices of open collaboration and open science in relation to collaboration in the field of biomedical and health research.

The specific aims of this scoping review are to:

- Provide an overview of current practices of or opinions about research collaboration that follow basic open principles (e.g., transparency, accessibility)
- Summarize existing online tools and resources available to improve open collaboration in research

We've expanded on our original aims to include a **secondary aim** of building an open source R-based pipeline for conducting scoping reviews. The entire source code, as well as text and collaboration workflows, are found on our GitHub repository `science-collective/scoping-review`.

Methods

The full protocol for this scoping review was uploaded to the Open Science Framework (2). As with any research project, things evolve compared to what was originally intended or planned. In this section we briefly re-state what we described in the protocol and especially we describe what was changed from the protocol.

This is a scoping review, so we followed the framework described in (3) as well as the guidelines outlined in the PRISMA-ScR statement (4).

Deviations and challenges

We originally aimed to review individual archives that included PubMed, bioRxiv, Scopus, and several others (the full list is in the protocol), and to use R packages with web API connections to each of these sources to programmatically extract the sources we wanted. However we encountered substantial issues that completely changed how we actually found and extracted the sources.

The first challenge we encountered was that, while most of the source databases had (hypothetically) exposed APIs along with R packages available to access them, they didn't always work well or had complicated instructions for actually using them. For instance, the preprint repositories bioRxiv and medRxiv didn't have an up to date R package to access the preprints nor was the web API well described, nor well designed. In order to effectively use it, you need to download the entire database of preprints locally before being able to search for the preprints you want. This made it effectively impossible to use.

The second challenge we encountered was a difference in results between using the web API compared to using the web search interface for some source databases. There were some small differences in search results between the API query compared to the web-based query. We couldn't identify or understand how these differences happened or way, though they were quite small differences (less than 10 out of thousands of scholarly works from the search results).

The third challenge we had was that some source databases, like the Web of Science or Scopus, didn't work consistently between the authors. In order to use the web API with the R packages, you need to generate a token for authentication. However, for some authors, the token worked fine, and for others, it didn't work at all, with no explanation or error message. Given our interest in and aim for reproducibility, since the same code ran differently between authors, we decided to exclude these source databases.

We also originally intended on searching websites, blogs, and other online resources, but it was very difficult to programmatically and systemtically achieve this. So we ended up only searching for scholarly work that were indexed in databases.

Given these challenges however, we unintentionally found OpenAlex, which is a public database of scholarly output that aggregates dozens of other source databases into one, easy interface. This resource also has an R package called `{openalexR}` (5) with a very well designed interface, which simplified or made redundant much of our code we wrote. It also made it unnecessary to use the individual R packages for each source database that we originally listed in the protocol. By using this archive, it resolved all our challenges and barriers we had before. For a full list of where they get their source data from, see their "About the data" page.

Document selection

We developed the initial search strategy in consultation with a research librarian. We collected the data via systematic searches of databases. All authors were involved in reviewing the collected sources, from reviewing the titles, to abstracts, and finally to full-text documents.

Information sources

We used OpenAlex, after encountering the challenges described above, which is an open database with scholarly works from Microsoft Academic Graph, Crossref, ORCID, ROR, DOAJ, Unpaywall, Pubmed, Pubmed Central, the ISSN International Centre, Internet Archive, Web crawls, subject-area and institutional repositories from arXiv and Zenodo.

Search terms

We used the following search terms when searching for scholarly work:

> (open[title]) AND (science OR research) AND (collaborating OR collaboration OR collaborate OR team OR cooperate OR cooperation OR cooperating) AND (technology OR technologies OR tool OR framework OR guideline OR principles OR practices OR systems OR resources)

Inclusion and exclusion criteria

Inclusion criteria included any document where open collaboration practices are not the primary focus. We relied on the definition of open collaboration from (1) in determining whether the records were relevant. Any published document with reporting on current open collaboration practices. Any published document with advice, guidance, tools, and/or recommendations for improving open collaboration. Article language in English. Exclusion criteria were documents that do not report on specific open collaboration practices.

While we didn't change our search terms for the initial search and extraction, we found that our search terms were not precise enough and got many irrelevant scholarly works. The full R code to filter down the search results is kept in the `R/exlusions.R` file on the GitHub repository. For example, we extracted a large number of scholarly works describing surgery such as "open wound", electronics such as "open circuit", or environmental such as "open water" that we had to post-process and exclude.

Charting procedure

At least two of us extracted data using a standardized and tested template. Data regarding the data source (e.g., author, title, publication year), open collaboration practices, and any other relevant information, will be extracted. Extracted data will be summarized with the descriptive analytical method described in (3), which is aimed at identifying and summarizing different open collaboration practices.

Search period

We were not able to dedicate as much time to this project as we had initially planned, so our "stopping rule" for end date didn't match with our actual search period. The actual end date for when we did the last search results extraction was 2024-02-29.

The start date remained 5 years from our end date, which would have been approximately 2019-02-28.

Results

Scoping review identifies only 11 papers discussing open collaboration

Practices of open collaboration

Across the papers found in our literature search, X out of 11 discuss actual experiences on implementing open collaboration, albeit in different organizational levels. These range from lab level (Turoman et al) center level (Bush et al and Grange et al), multi-center level (ManyPrimates et al and Grange et al) and field level: psychology (Alessandorni et al), Spinal Cord injury (Torres-Espin et al) and neuroimaging (Niso et al). All of these share in common the need for an infrastructure that promotes open science and open collaboration via starting building

a community of collaborators or making tools (usually online resources) that researchers can use to reach FAIR (Finndable, Accessible, Interoperable and Reusable) goals while doing research. Before delving in those, here we provide the specific needs for implementing open collaboration in each organizational level, as well as describing the differences in approaches among experiences between and across levels (Figure 2).

Implementation of open collaboration practices at laboratory level

Implementation of open collaboration practices at center level

For center-level only one paper was identified (Bush et al), providing the example of a center of neuroimaging describing first how they identified the needs for open collaboration, then the individual-level practices that researchers should adapt and then the center-level changes to promote reaching FAIR goals.

Implementation of open collaboration practices at multi-center level

Interestingly, the level of organization with most papers is the multi-center level with two papers from ManyPrimates et al (ManyPrimates 2019 and ManyPrimates 2021) and Grange et al. In this organization level the most urgent needs are: 1) identifying the needs for open collaboration and 2) creating an infrastructure that allows those needs to be satisfied. Indeed, Grange et al describes a systematic review as an exploration of the needs for 11 UKRN centers to implement open collaborative practives. In this paper they specially focus on tools and resources among the 11 institutions and make clear the need for starting to utilize open resources as a first step to adapt open collaborative practices. The first paper from ManyPrimates et al [REF] describes the start of another multi-center collaboration, albeit from different institutions. It is interesting to observe that in this situation, the complex issue is to start the collaboration itself, which was ignited thanks to a symposium and an email chain among future collaborators. In contrast, ManyPrimates could quickly develop an open collaborative structure with relative ease in comparison to changing an already solid infrastructure, like in the case of Grange et al.

Implementation of open collaboration practices at field level

Useful tools for open collaboration

Discussions

Conclusions

Contributions

We originally had another author involved, HC, who is on the protocol author list. Since the time we uploaded the protocol, she has moved out of academia and could no longer participate in any work.

Otherwise, the remaining authors (MG, DBI, and LWJ) had the following contribution roles, following the CRediT taxonomy of conceptualization, data curation, formal analysis, investigation, methodology, writing - original draft, writing - reviewing and editing, project administration, software, validation, and visualization.

Bibliography

1. Forte A, Lampe C. Defining, Understanding, and Supporting Open Collaboration: Lessons From the Literature. American Behavioral Scientist. 2013Jan;57(5):535–47.
2. Johnston L W, Urena M G, Chatwin H, Ibsen D B. Towards open collaboration in biomedical and health research: A scoping review. OSF Registries; 2022.
3. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology. 2005Feb;8(1):19–32.
4. Tricco A C, Lillie E, Zarin W, O'Brien K K, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Annals of Internal Medicine. 2018Oct;169(7):467–73.
5. Aria M, Le T, Cuccurullo C, Belfiore A, Choe J. openalexR: An R-Tool for Collecting Bibliometric Data from OpenAlex. The R Journal. 2024Apr;15(4):167–80.