Lecture 4: Open Research Data and Open Peer Review

Science Communication: Research Productivity and Data Analytics using Open Source Software

Dr. Mohit Garg

Central Library

IIT Delhi

Week: 01

Lecture: 04

Dear Learners, Welcome again to the second part of this first week of this course. I hope you all are able to identify some of the journals related to your domain which are using different open access models. So, in this lecture I will be discussing the two of the open science components. In the last lecture I have discussed one of the components of open science is that open access to publications. Then under which we have discussed different open access models and the licenses under which we have discussed what exactly open access is, what the different myths have been circulated about open access, what the different open access models are, and what different licenses for publishing the open access work are. So, in this lecture I will be discussing two other components of open science: one is open research data, and the second one is open peer review. Let us start first with the second component of open research data, and then we will discuss open peer review.

So, what happened is that there is a common practice that we use to share the final results. But, throughout this research journey from here to there, we use to generate a huge amount of data which is not shared with the community. So, the community felt the need of opening this research data so that more explorations can be done using the same dataset. So, open research data simply means the opening of the research data like opening of the publications.

So, research data refers to the information collected, observed or generated during the research. So, it is of various types from discipline to discipline. Some of the types of data are listed here, for example, experimental results, software codes, transcript of interviews, questionnaire surveys, historical documents, social media data, logs of web server traffic, or another activity, and many more. There are different types of research data depending on the type of research they are doing. A huge amount of data is generated during that process. This data comes in variety of formats, like textual data which can be in PDF or Word, then the tabular data in CSV, TS, TSP or Excel, or maybe the web data in HTML

format, or XML format, or the databases like MySQL, Oracle, or MariaDB, images, video data, geospatial data in shape files, or GeoJSON files. So, it can be in different formats of data, and there can be different other formats of research data. The questions come why exactly we need to open the research data. So, what are the reasons why the community felt the need of opening the research data? So, the first and the foremost advantage of opening up the research data is collaboration.

So, let us assume that there is this Group A. So this is a Group A and this is a Group B. So, this Group A is of one particular subject discipline. This one Group A of one subject discipline is researching some phenomena, let say for example this is a phenomenon they are exploring, and say for example this phenomenon is Y. Now during their investigation, they will generate the data and based on that they will present the results. So, Group A will study this phenomenon and they will generate the data and based on that data they will present the findings of their research. So, this is another Group B which is another discipline but working on the same phenomena with some other parameters. If Group A does not share the data, then Group B will have to invest a huge amount of time and money in collection of this data again.

So, if data is shared, Group B can collaborate with Group A, and can bring diverse perspectives to the scientific inquiry. It will also save a lot of infrastructure for another group. Then the second reason why research data should be open is reproducibility. So, reproducibility is the key aspect of science which helps in reproducing the findings of the research again from the data. So, open research data helps in replication of earlier experiments and to produce the same research findings. It will not only help in enhancing the reliability of scientific knowledge but also help in decreasing the cases of scientific misconduct due to data fabrication and all.

So, with the availability of data, the reproducibility can be achieved and results can be validated. So, it will authenticate the quality of research and enhance the chances of publications. It will also increase the visibility, credibility and impact of the work. So more citations can be attracted to the research paper. And it will receive more citations. Many of the journals, funding agencies, and organizations have a mandate to share the research data. So opening of the research data can help in complying with the rules of journals and funding agencies. Also, like many of the journals are asking for data availability statements, and they are giving like this data is particularly available here. So, that kind of thing can be complied if the research data is shared.

Then another point is that the opening of research data helps in motivating the other scientists to share the data. Also, opening of the research data can help in publishing the data papers and can increase the count of the number of the publications. So, to publish a data paper, or to comply with the journal or funder policy these are some of the things that need to be remembered. So, data should be uploaded on the repository. It should be

on the appropriate identifier. It should have the persistent identifier so that data can be perfectly located. Then the data should have the license and it must follow the FAIR principles. But FAIR principles are something I will discuss in a few minutes. Then it should have the description of data, what data is, how data is collected, what the different variables are there, and then the ethical approval. If any ethical approval is required for that data. So, in order to open the research data so that it is beneficial for all. Research data must follow the international standards, known as FAIR principles. So, what is FAIR? In FAIR, the F stands for the findable. Then A stands for accessible. I stands for interoperable, and R stands for reusable. So, what F means is that research data can easily be located so that research data should be properly describable and have the proper metadata. Then the second point is accessible. So, what accessible means. It means that there is something that needs to be recorded about how the data is available. Is there any restricted access to the data, or is there any embargo period to particular dataset. Then what is the usage license of that dataset?

Then I stands for interoperability. So, it means that data follows the standard. So that it can be used in different environments. Then R stands for reusability. It means that reusability of data is the ultimate goal of opening the research data, the type of license assigned, and whether it allows the reuse of that data. These are the FAIR principles which need to be remembered when sharing the research data. So, opening of research data is not just putting the data on some of the servers and providing the link. It is about how that data is beneficial for society. Say for example if the data is uploaded in some random format, like say for example, XZQ but there are no tools available for accessing the data, or converting the data. So, this is not about opening the research data. So, to open the research data we must follow these FAIR principles, so that it will be beneficial for society.

Now different kinds of repositories are available. So, Zenodo, Dryad, Dataverse, OSF are some of the popular data repositories that enable researchers to deposit and share their various kinds of research outputs like datasets, software, and publications. These repositories provide DOI and permanent links of the work, so data can be uploaded here. And it will provide the DOI and then it can be linked. If you are publishing a data paper, it can be provided and those DOI can be shared for receiving the citations also. This is a brief about the open research data.

Let us now move to our last component of open science that is open peer review. Let us first understand what peer review is. So, peer review is considered as the important quality measure in publishing scientific research. In this process one or more subject experts known as peer reviewers assess the research work submitted by the authors. So, generally in peer review either two or three are the experienced experts who evaluate the works on different parameters to assess the quality of the work. After the assessment, they submit their opinions to the editor. And based on those recommendations the editor

either accepts or rejects the publication for the publishing in a journal, or any of the other sources. So, this peer review based on their expertise contributes to assessing the quality of the research work. Generally the task of the reviewers is to assess the novelty of the work, like originality of the work, assumptions, principles, methods, findings of the work, content of the work, data, figures, tables and the references. Based on these, they rate the research work on different parameters like how technical the content is, how much it is important to the field, conciseness of the presentation style and clarity, and how much the research work is complete or incomplete. Based on these, they give their final recommendations to the editors for the publishing either to accept the research work with minor or major revision, or reject the research paper.

So, let us now understand what this peer review process is. So, this peer review process generally has this like say for example. So, this is a Group A which has already conducted the research and drafted their research work into a paper. Now they send this to this source. So, let us say for example this is a source. So, let us say for example this is a journal and they send this here. So, this journal is maintained by the editors. So, these are the editors. As soon as this submitted research work reaches the editor they do check some of the parameters and if it does not follow those parameters they send it back to Group A either with the suggestion or to reject. Then if that particular research work follows those parameters then the editor further sends this research work to the reviewers. So, each journal has its own database of reviewers. If the research work fulfills the initial parameters of the editorial board, they send the research work to the reviewers. So, each journal has their own database of reviewers to which they send the work. Now these reviewers review the work. Then they send the recommendation to the editors. And based on this recommendation the editor sends the message to the authors either to accept, or reject, or any minor or major revisions are required. This is how the whole process of peer review works. Generally it is of two types, one is blind peer review and another is of open peer review. In blind peer review it is that the editor knows the reviewers. But the reviewer either knows the author or doesn't know the author. So, under blind peer review it is of two types one is a single blind peer review and another is a double blind peer review. So, in single blind peer review this Group A author doesn't know who the reviewer is, but the reviewer knows who this author is. So, under single blind peer review, reviewers know the authors. But the authors do not know who the reviewers are. So, in double-blind peer review the authors don't know who reviewers are, and reviewers don't know who exactly the authors are. So it is through the editor only they have these all. So in double blind peer review the author does not know who exactly reviewed their contents, and the reviewer does not know who exactly the authors are. So, this is about the blind peer review. So, in both those review systems, what reviews are given by the reviewers and what the responses are given by the authors are generally with the editor only. So, in both the cases of either single- or double-blind peer review, the reviewers' comments and the author's responses are with the editors only. So, either the reviewers

know what review they have given or the author knows what review they have received or the editor knows what reviews they have received. But the outside world does not know all these reviews. The outside world only knows the final product of those final publications.

To bring transparency in this peer review system and to reduce the biases, another type of review system has emerged which is known as open peer review where the whole this process is open. Let us now see some of the platforms which are using an open peer review process. So, OpenReview.Net is a platform that aims to promote openness in the peer review process. So, various conferences, workshops, and funding agencies are using this platform. Let us have a quick demo of the OpenReview platform.

So, let us take an example that I am searching for. If you see this article. An abstract on all those things is visible to you. But the replies of the reviewers are also provided here. So, what did this reviewer think about this article, what was the rating they have given, what were the strengths and weaknesses. So, this whole review of this particular work has been open to everyone. So, you see that there is no login, no such thing is required, everything is in public. So, this is how this platform promotes the open peer review process. Everything, what are the reviewers' comments, what are the replies of the author are available to the public and then based on this they can evaluate the quality of the content.

So, Qeios is another platform which promotes an open peer review process. So, it is a web-based platform, it provides a space for researchers to share their work openly and engage in transparent and open peer review. So, it is available in both as a free service or a subscription-based service based on those extra features. So, let us take one example of how the peer review process works in Qeios. So, let us take an example of an article. So, I just put it down by only articles and I put another filter like by time. So, let us take an example of this particular item. If you see here this is a version of September 8, 2023. You can see the latest version also. So, this is the latest version from September 12, 2023 and you can see here the number of times this content has been viewed, number of times this particular content has been downloaded, and the number of reviews have been given. So, if you see in the earlier system in blind peer review there were either 2 or maximum 3 reviews were there. But here the content is publicly available and anyone can submit their review. So, if you see here like these, some of the reviews have been given about this work, about the language, or maybe the comments, the rating also, how they rate the content whether it is 4-star, 5-star, or 3-star. Then you can also see here that the DOI of this article and the license is CC-BY. This is also another platform which promotes an open peer review process.

Let us move to a one more example of the platform which promotes open peer review which is known as F1000 Research is an open access publishing platform that supports

open peer review. It has APC charges based on the 3 kinds of the papers, what kind of papers you are submitted. So, it has some APC for that and let us take one example here also. So, I am going with the same word like sentiment analysis. If I search it down. And let us take this example modeling sentiment based on objectivity and subjectivity. So, if you see here, you have all the details about the reviewers. So, these are the 3 reviewers who have submitted their review. So, this was the first version, you can see this version also and you can see the version 2 also and this was the version of 4 October 2021 and this is the version of 17th May 2022. Now you can also see what review report has been given by these reviewers. So, you can see here like if you click here, so this is the review report of reviewer 1, and similarly you can see the review report of another reviewer also. You can also find like this many times this article has been viewed or the number of downloads are there. Then also like if you want to comment on this article, you comment here also. So, this is about the 3 of the platforms which support open peer review. So, this is about some of the platforms which follow Open peer review process and so with this our second part as well as the week 1 is completed.

I hope you have enjoyed the lecture. So, in this week we have covered the basics of science communication, and the different channels to communicate the science, what are their different key attributes. Then we have discussed under Open Science what are different components like Open Access, Open Research Data, and Open Peer Review. Then we discussed the models of Open access. We have discussed different models of Open access, then Open access licensing, and then repositories to upload the research data. In the end, we have discussed some of the peer review platforms.

So, please let us know in the discussion forum if you have any of the questions or queries related to this week. We will help you to answer all your queries in the discussion forum. Thank you. See you next week. Thank you.