Lecture 3: Open Science and Open Access

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

Dr. Mohit Garg

Central Library

IIT Delhi

Week: 01

Lecture: 03

Dear learners, welcome again. So, before starting the second part of this week, let us have a quick recap of what we have discussed in the last two lectures. So, in the first lecture we have discussed the basics of science communication, why it is important for the scientific community to communicate and the different channels of communicating science. In the second lecture we have discussed the different characteristics of some of the channels of science communication like journals, conference proceedings, edited books, articles, data papers, erratum, retractions, etc. I hope you are able to identify some of these channels from your domain. Please share the link of the example with us in the discussion forum.

Those who are not able to find or face any issue or need to have some clarity, please let us know through discussion forums. We all the instructor and teaching assistant will do our best to explain your queries. So, in this part I will be discussing open science and its different components like open access, open data and open peer review. For the concept of open science, we have already discussed what science is.

So, the main focus is what is open. So, open means that anyone can freely access, use, modify and share the knowledge. So, there is some problem about the channels that we have discussed in the last lecture. So, the first key issue is most of the channels are generally behind the payables. So, this particular image shows what this means. This means you need to pay some amount of money for accessing the content published in those channels.

So, the access in these channels will be either by paying some amount or through the libraries that subscribe to these channels. But those who do not have the money or access to the library would not be able to read the content published in those channels. So, this way only fewer people can enjoy the access to knowledge. So, science must be accessible to everyone. So, if you remember when I was discussing the retraction, there were some

articles which were published and later on they were retracted due to peer review, compromising or due to data fabrication.

And there are many various reasons for retracting the paper. So, all these reasons demanded the need of the opening of this review process and the data so that retraction can be minimized and more transparency can be brought to the science. So, generally the scientific community used to share the final results, but the data were mostly unshared. So, this is where the concept of open science emerged. So, let us first discuss what open science is.

So, UNESCO defined open science as a set of principles and practices that aim to make scientific research from all fields accessible to everyone for the benefit of scientists and society as whole. Let us take another definition by these authors they have given in 2018 that open science is a transparent and accessible knowledge that is shared and developed through collaborative networks. So, if we see this particular definition, it has four key parts. So, the first part is transparent. So, this is our first part.

So, transparency, what transparency means? Transparency at all stages of research, not at the final stage, but like when the data was generated at this stage, how the, what were the variables? So, transparency at all stages of research, not at the final publications, but opening that peer review, what were the reviewer comments? How was the data collected? All those stages are transparent so that research can be reproducible. Then the second is accessible. The second is accessible. So, what does "accessible" mean? So, it means that not only the final publications are accessible, but the research data is also accessible to everyone around the world. Now the third important part in this definition is shared.

So, it means sharing and using all available knowledge for new findings. The last is the collaborative networks. So, what collaborative network means? It means global collaboration for the creation of knowledge and it includes collaboration not only among the researchers and the countries, but also it includes collaboration among researchers, institutions, countries and other disciplines. So, open science helps in various ways. So, the first is it makes multilingual scientific knowledge openly available, accessible and reusable for everyone.

It increases the scientific collaboration and sharing of information for the benefit of science and society. And the third is it opens the process of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. So, open science helps in multiple ways. So, first is to make multilingual scientific knowledge openly available, accessible and reusable for everyone. And the second is to increase scientific collaboration and sharing of information for the benefit of science and society.

And the third is to open the process of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. So, UNESCO has come up with these four values and six principles that aim to ensure not only that scientific knowledge is accessible, but also the production of that knowledge like peer review and research data. So, which makes science inclusive, equitable and sustainable. So, let us first discuss what these four values are: quality and integrity, then collective benefit, equity and fairness and diversity and inclusiveness. So, the first value is quality and integrity, which ensures that high quality standards for science are followed that enhances the transparency and accuracy of evaluating scientific methods and outcomes.

The second value is collective benefit. So, what collective benefit means is that scientific knowledge is not the exclusive property of any individual organization or nation, instead it is a public resource that belongs to the public. So, it is a public resource that belongs to all the humanities. Then the third value is equity and fairness, which means that equitable and fair, excess of science irrespective of any kind of discrimination like nationality, race, age, gender, socio-economic status, discipline, language, religion or any other grounds. So, the last and the fourth value is diversity and inclusiveness.

So, what it means is to bring diversity and inclusiveness in all aspects of science. These are the four key values as per UNESCO. Now UNESCO has given six principles for open science. So, the first is transparency, scrutiny, critique and reproducibility. Second is equality of opportunity. Third is responsibility, respect and accountability. Fourth is collaboration, participation and inclusion. Fifth is flexibility and the last is sustainability. Let us now discuss these principles. So, the first principle is transparency, scrutiny, critique and reproducibility.

So, it means that every process of science should be open and transparent further to encourage the researchers to regressively scan the science and submit constructive critiques that strengthen the overall quality of the scientific output. Then the second principle is equality of opportunity. It means all scientists have equal opportunity to access and contribute in science irrespective of any kind of constraint. Then the third principle is reproducibility, respect and accountability. It means that public scientists should be aware of public accountability and potential conflict of interest.

Then the fourth principle is collaboration, participation and inclusion. It means collaboration, participation at a larger level breaking down any geographical language barriers, amplifying diverse voices and harnessing collective wisdom to solve the various societal problems. Then the fifth is flexibility. Sustainability promotes the idea that not one kind of format and structure is applicable to all. So, it encourages us to explore different pathways following the core values.

And the last principle is sustainability. It encourages building long term infrastructure services and funding models so that the global scientific community from all the geographical regions are included. So, these are the four values and six principles of open science as given by UNESCO. So, let us now understand the component of open science with the help of this monitor. It represents the three key characteristics of open science.

So, the first is open access to publication. Then the second is open research data. And the third is open scholarly communication. So, the first component of open access to publication includes open access publication, preprints, alternative publishing platforms, funder policy on open access, journal policy on open access. Then the second component open research data includes research data repository, funder policies on data sharing, researcher attitude towards data sharing.

And the third component of open scholarly communication consists of open peer review, journal policies on open peer review, use of Altmetric platform and correction and retraction. So, if you so, if you have noticed that two of the things are common in two of the components. So, the first is the funder policy. So, funder policies on open access and then second is funder policies on data sharing. So, what it means is how the funding agency promotes the open access publication and sharing of research data.

So, second is the journal policies on open access and the journal policies on open peer review. How a journal takes the open access publication and opens up the peer review process. So, we will discuss each of these components in detail. Let us first discuss what open access is. So, the idea behind open access is that whole knowledge should be open for everyone.

So, let us understand this open access with the definition given by Budapest Open Access Initiative. So, what exactly is it? So, the first thing is free availability on the public internet. Then the second is permitting any users to read, download, copy, distribute, print, search or link to the full-text of these articles. This is the second part of what open access is. Then the third part is to scroll them for indexing, pass them as data to software or use them for any other lawful purpose.

And the fourth part is without any financial, legal or technical barriers other than those inseparable from gaining access to the internet itself. So, open access not only means that the access should be open but the whole idea of open access means that it should be freely available. And the second is it allows users to read, download, copy, distribute, print, search or link to the full-text of the article. Then the third aspect of open access is it should be to scroll them for indexing, pass them as data to software or use them for any other lawful purpose. And the fourth aspect of open access is it without financial, legal or technical barriers other than those inseparable from gaining access to the internet itself.

So, there should be no such kind of financial and legal or technical barrier which creates hindrance in the access of knowledge. So, let us now see the whole journey of the open access movement. So, the journey of open access started with a meeting of supporters when they met in Budapest on December 1st to 2nd in 2001. So, the goal of this meeting was to enhance research articles from all academic fields that are freely available on the internet. So, it recommended to achieve this meeting suggested two key strategies.

So, it promotes two things: the first is self-archiving and the second is open access journals. So, this Declaration has been signed by various organizations and individuals around the world that include researchers, universities, laboratories, libraries, foundations, journals, publishers, etc. Then following this in the year 2002 a license known as Creative Commons emerged that supported the Open Access Movement. So, CC licenses are the public licenses that promote copyright-free distribution of a copyrighted work.

It allows the creator of the work to give the right to share or mix or distribute to other people. There are six types of CC licenses. So, we will be discussing some times during this lecture itself. So, in the same year another initiative was started as Sherpa Romeo Services. So, Sherpa Romeo Services basically it is a web-based platform that has the feature of searching the copyright and other archiving policy of the journals.

So, in the year 2003 two meetings were held at different places that supported the open access movement. So, the first meeting is known as the Bethesda Statement on Open Access. This meeting was held on 11th of April 2003 in Maryland, USA where the biomedical research community met and discussed open access. Then the second meeting was held on 22nd of October 2003 in the Berlin known as Berlin Declaration on Open Access that discusses the opening of the knowledge of science and humanities. So, both these declarations supported open access to scientific knowledge. So, in the same year another initiative was started which is known as the Directory of Open Access Journals. So, it is an online catalogue and provides access to open access journals from various disciplines worldwide. So, in the same year PLOS biology was published PLOS also known as Public Library of Science is a non-profit open access publisher that provides the access of high-quality peer reviewed scientific journals. So, in 2005 another initiative was started similar to the Directory of Open Access Journals. So, this was the Directory of Open Access Repositories.

It basically provides a directory of all the repositories worldwide which are in the open access. And similar to the Directory of Open Access Repositories, another initiative was started known as Directory of Open Access Books. This was started in 2012. So, it has catalogued all the open access books. So, in 2018 another initiative called PlanS was initiated. So, where the research funded by grants should be available via open access, and it is supported by the cOAlitionS which is a group of funders, institutions, and

organizations that support this. So, in the year 2021, the UNESCO Recommendation on Open Science came. Let us now discuss what are the key advantages of open access. So, the first and foremost advantage is that everyone has access to your findings. So, the reader need not pay any kind of charge for accessing the content.

So, the second advantage of open access is that more exposure for your work and it is understood that when your research work is freely available anyone can see the finding of your research that means the more exposure will come to your work. Just a very interesting example I would like to give you is that in 2003 PLOS Biology published the first issue within a few hours 5 lakh visitors had visited that site. This kind of exposure you can expect by publishing in the open access. So, the third key advantage of open access publishing is that practitioners can apply the findings of your research directly into their area.

Then the fourth and the most important point is citations. Citation is considered as the key parameter for evaluating the impact of scientific productivity. When your work is more exposed and more people are reading it is obvious that you will get the higher citations for your research work. Then the fifth point is that when it is published in open access it will travel faster and may reach the policy maker and policy can be made based on your findings. Another key advantage of open access publishing is that taxpayers get the benefit and mainly like developing countries where there is already a scarcity of the resources. So, the researcher from the developing country can access your research easily.

So, the last key advantage is compliance with the grant rules. So, many funding agencies have the mandate of publishing the outcomes of the funding research in an open access source. So, with these various advantages of open access there are like some of the myths which have been circulated about open access. Let us now discuss this open access myth. If you see the first myth is that open access is available for stem discipline.

So, this is not true as open access publishing is available for all the publications and the field. So, you can understand this by a small amount of data. So, this is data taken from Scopus of all open access publications. If you see there are like in social sciences and then if you see others also in others like arts and humanities and then multidisciplinary psychology. So, this shows that it is just a myth that open access is available for STEM disciplines.

Now take the other myths of open access publishing. So, these are like open access publishers are predatory and of poor quality. Open access publishers do not follow peer review and open access publishing is a fast process of publishing. So, these are the common myths circulated among open access publishing and there is no direct relationship between open access publication and the quality of the journal. So, it is not

that publishing an open access journal is easy and fast. Open access journals have strict quality measures and follow a peer review process which takes months in publishing.

So, there may be some journals which follow some unethical practices but that is also seen for paywall journals. So, all these three myths are not true. Now take another two myths. One is that open access does not have an embargo and no license for open access content. So, these licenses and embargoes are the policy matter and it varies from journal to journal.

So, embargo is actually a time period for which the access of articles is restricted before making it too online. So, open access publication generally prefers public licenses like creative commons for the publications. So, open access publications have embargoes and again it varies from journal to journal and generally it is based on the open access model, but the model of open access is that journal is following. So, it is not like that all open access journals have embargoes or all of open access journals do not have embargoes. So, the next myth is that open access publishing requires fees like APC article processing charges.

So, not all journals ask the APC to publish the research. So, it again depends from source to source what source exactly publishes the open access, and exactly what kind of model they are using for publishing the open access research. There are some journals which do not take a single penny for publishing the research in open access. We will discuss what these different kinds of open access models are. So, now move to the next myth which is commonly circulated about open access is that bibliographic databases, like, Scopus, Web of Science do not index open access, and open access does not have reputation and good metrics. So, if you see this chart there is research that has been published in open access and they are indexed in Scopus.

And similarly open access research is indexed in the Web of Science and other databases. So, it is not true that open access publications are not indexed in bibliographic databases or they don't have good journal metrics. So, there are various open access journals which are indexed in Scopus and Web of Science, and also have good journal metrics like Citescore and Impact Factor. Some of the myths are basically based on the open access models, which open access model exactly that journal is following.

So, let us now discuss these open access models. So, these open access models are basically the business model which publishers follow to publish the open access research to recover their costing and to generate the revenues. So, in this the first model is green open access model. In the green open access model, the researcher publishes in a journal and shares their work by archiving it in some repositories. So, these repositories can be like institutional or own hosted repositories. So, in green open access journals, the final accepted version is archived and there is an embargo on the final published version.

So, this is done by the publisher so that they can generate some revenue to cover the costing of publishing and it has a license based on authors or publisher choice. So, the next model of open access is the gold open access model. So, in gold open access as soon as an article is published it is made under the open access. So, the difference is the author has to pay here some amount which journal called as APC for publishing the open access article and journals which follow the gold open access model, they use either open or creative common licenses. Then the third model is the diamond model which is partly different from green and gold.

So, in green there is some embargo on publishing the final version. So, in diamond there is no such embargo. As soon as the article is published it will be made under open access. In the gold model for immediate publication of articles under open access it requires APC under diamond model there is no APC and in the last diamond open access model journal follow creative commons licenses like CC by or CC by essay. So, the next model is a hybrid model. So, the hybrid model is another model where the author can publish the article as open access in a paywall journal. The whole journal is not open access, so the journal is paywall but the author can publish that article as open access but the author has to pay the APC amount for publication of that article in that journal.

It also uses CC license but here the CC license is only for that article for which author has paid the amount for publishing it in open access. Then another open access model which is not that quite popular is bronze model. It partly shares the features of the green and gold open access model. This is about the different open access model.

Let us now see what these different licenses are. What exactly this CC license is. So, a license is a legal framework that grants permission to use the work under certain conditions. One of the myths about open access publishing has been circulated that open access publishing is done without any licenses and anyone can reuse the content or distribute without any license. So, their public licenses come into the picture where they grant the permission to the author what part of their work they want to share and how they want to share. So, there are many advantages of public licenses and it helps in many ways in supporting open access like clarifying what are the permissions associated with the work and encouraging sharing and distribution of work, and promoting fair use of work.

So, there are various kinds of licenses available based on the task. So, creative common is one such example. So, there are the four key attributes of the CC licenses. So, the first is credit must be given to the creator. So, if you see any of the licenses, if this symbol is there that means you must give the credit. Second is if any changes modification remixing is done it should be shared under the same terms.

So, this is the second also known as share alike. Then third is representing that it is only for non-commercial use and the fourth is no derivative of this work is allowed. So, these are the four attributes of CC licenses based on these four attributes CC has six licenses and another license known as CC0 which is a public domain license. Let us discuss each of these licenses. So, the first license is allowed to distribute, remix, adopt and build in any way even for commercial use but the only thing you have to remember is that if this license is used you have to give the attribution to the creator.

So, this first license shows that attribution should be given. Then the second license is the same as the first one. One thing is you can remix the content but only thing is that you have to give the same license under the same terms. So, the third license is quite similar to the first one but the only difference is that this license is given to any of the work that means that work should be used for only for the non-commercial purpose. So, you cannot use that work for commercial purposes. Then the fourth license is a combination of second and third licenses that is you have to give the attribution and it should not be used for non-commercial purposes and if anything, any modification is done it should have the same license under the same terms.

Then the fifth license is a strict one compared to the other four licenses. So, the only thing is that from the fifth license that you have to give the attribute you have to give the credit to the creator. The only thing is that no derivative can be made from the work which are licensed under this fifth license and the last license is a combination of the third license and the fifth license is under which the work should be used for non-commercial purposes only and no derivative of this work can be made. And in the end if you see this license this is also known as CC0. So, this CC0 means that it is in a public domain and there are no such conditions applicable to this work.

You can do anything without worrying about any of the conditions. So, let us now see some of these open access tools which can be useful for finding the open access repository and the journals and their corresponding policies. So, the first one in this category is a DOAJ which is a Directory of Open Access Journals. So, DOAJ is a Directory of Open Access Journals. It indexes the open access journals published worldwide and provides comprehensive information about them. So, if we search on this for example let us say we want to see for management. It will list all the journals of management that include journal scope, website, editorial information licensing details, APC and the type of article it published.

So, provide the different features so that results can be refined based on the search results. So, we can further refine these results based on languages then the licenses. So, what all licenses that journal uses. So, let us now take the example of this journal of engineering project and production management. If you see, it is showing in this journal that this is the ISSN number in the print form and this is in the online form.

P related to this journal so there is no publication fee and then the journal uses this license. So, then the publisher of this particular journal is this. Then this is the publication timeline of publishing the articles in this journal. So, it is expected that it takes on average 12 weeks from submission to publication and it gives the DOI.

This is how open access journals can be searched in DOAJ. So, next is this Sherpa Romeo service. It is a web based online platform that provides comprehensive information about the open access policy of journals and publishers. What you can archive, what version of the article you can share in the repository and what are the rights you will get for publishing the article. You can search the journal here by either the title or ISSN number.

You can also browse the list as a country or browse by publisher by clicking on these two tabs. Let us know if you don't know the exact name of the journal. So, you need not to worry, you have to just type in some keywords and it will auto suggest some of the journals which you can select from this drop down. Let us take the example of how I write indigenous. It will automatically show that these are journals related to indigenous and like say for example I want to search.

So, as soon as you select it will take you to the journal information page. So, this page has the information under three categories. So, one is this publication information where it has title, ISSN, URL, publishers, DOAJ, details and whether it requires APC or not and what is the status. Then the second category it has is this publisher policy. So, publisher policy for published version, accepted version and submitted version and the third is the record information.

The record about this journal is created and when was the last modified. So, the main information that we require is in the publisher policy. If you see it has the policy for three versions of the article. So, the first version is the published version, then the second is the accepted version, and the third is the published version. So, the published version is the final published article. Accepted version is one which is accepted after the revision if any, and the submitted version is one which was submitted initially.

For this particular journal the policy is that it includes open access publishing. There is no embargo. Then it follows the CC-BY 4.0 license, and the copyright owner is the author. Location is where you can deposit the research work. So, you can deposit this work on any website or journal website. And these are some of the conditions that the published source must be acknowledged with citation and must link to the published article with DOI. And then for this accepted version. And for submitted versions, so this policy does not allow for open access for this version. So, this is how you can use Sherpa Romeo Services for knowing the policy of that journal which you want to select for the

publication. So, OpenDOAR is another Sherpa service which is an online directory of open access repositories.

The interface of OpenDOAR is also quite similar to the Sherpa Romeo site. So, if you search the repository here either by country wise or using the advanced search. So, if I click on this, browse by country. So, if you can see that here all the open access repositories are listed country wise. So, like for example for India 106 repositories are listed in this directory.

Say for example I want to search for a repository like ShodhGanga. Let's take the example of ShodhGanga. And this OpenDOAR page has four pieces of information. So, one is repository information, then the second information is organization, then the third is about open access policy and the fourth is record information. So, in the repository information it has all the details like name, type, URL, then the software used for building this repository, what are the different subjects, then organization, name and URL and then open access policy. Open access policies are listed here for this repository and then in the final the record information when this record was created and last updated date.

So, this is about OpenDOAR service. So, there is another Sherpa service which is Sherpa Juliet. So, Sherpa Juliet is a service which is a web based online platform that provides the details of funder policy and their requirement on open access publication and data archiving. You can search similar to Sherpa Romeo services and OpenDOAR service. You can just type in down here the funder's name and it will list the whole policy and also you can browse the funders. So, if you don't know the funder's name you can browse the funder's by clicking on this tab.

So, funder's like say for example these are the different funders. So, you can go country wise also. So, it will display all the funder's which are index in this Sherpa Juliet service and you can see their details. So, this is another Sherpa service known as FACT, Funders & Authors Compliance Tool. So, here you can see whether a Journal follows the funder open access policy and its structure and its interface and other features are similar to other Sherpa services. So, recently all these four Sherpa services are combined and it is known as one single service which is still in beta.

So, for that you can click on here and you can go to this URL. It's still in beta. It's under development, but you can search all details about funder's, journals, publishers or repositories at a single place and you can search like say for example a journal and you can write management. So, these are some of the journals. So, this is how you can use this platform for all the details. But this is still in beta and I hope in the future it will be a full-fledged single platform where you can see the policies about funders, journals, publishers and the repositories. So, the platform until now we have discussed is about

knowing the journals what their policies are and what they allow you for archiving or sharing the content. Now you have already published the content and you want to share your content and you want to know about the policy about your research paper.

So, for that you can use this service called share your paper. So, it's another web-based platform that allows you to search the open access policy, about your published work and then deposit it into another repository. Let us take one example of how we can search for articles on this platform. So, here you have to just type in down the DOI of your paper, and once DOI has been submitted it will search for that journal. So, it is saying that they have checked and this particular journal LHTN encourages you to freely share your paper, so colleagues and the public can freely read and cite it.

So, this is how you can use this platform to know the policy about your article. So, these are some of the open access tools which can be used to know about the open access policy of the journals or the funding agency or the repository or to know about whether your article can be shared. I will stop here in the next lecture. I will discuss two of the other open science components that are open peer review and open research data. So, I request you all to share at least two of the journals from your domain that follow either the gold open access model or green open access model and diamond open access model. Please share this detail in the discussion forum. Thank you, see you in the next lecture.