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

Vijay Kumar Verma

Central Library

IIT Delhi

Week: 05

Lecture 17: Science Communication and Different Metrics

Dear Learners, welcome to our NPTEL course on Science Communication, Research Productivity and Data Analytics using Open Source Software.

Today, I am going to cover Introduction and Application of Bibliometrics and Laws of Scientometrics in the Mapping of Science Communication. The basic objective of this session is to understand science communication. We will be discussing what science communication is. Why do we need different metrics for measuring science communication? After that we will be discussing bibliometrics, its introduction and applications. Then we will introduce the scientometrics, its applications and we will be covering the different laws of scientometrics.

We will be discussing the different applications of scientometrics also. As we know that science is systematic gathering of knowledge. When we say systematic, that means science is something which happens in a very systematic way, it does not happen in a haphazard way, it does not happen in a random way. We are getting the knowledge, scientific knowledge in systematic fashion, in systematic ways.

Science deals with organizing and condensing knowledge into laws and theories and the most important thing is that these laws and theories are universally applicable and can be tested. Another important thing about science is that it is empirical, objective, rational and progressive. When we talk about progressive that means science is not constant, there is progress in each and every domain of science, and scientific knowledge basically deals with the progress which happens in the different domains of science. As we already said, the approach of science is always systematic. When we say science communication, science communication is basically a process of transferring scientific knowledge and discovery to diverse audiences in an engaging and relevant manner in their easy understanding.

That means what you see here, in case of science communication when we are transferring the knowledge of science, of course there would be the different stakeholders

and science communication should be in such a way that it can reach to the diverse audience in such a manner that it should be engaging, it should be relevant to them, it should be such a way that they can understand the thing, and they based on that they can make their own perspective. It's about communicating science and building bridges between people involved in scientific research and different groups of the public. You see there are two groups of people, one who are doing actual science that means they are at the production end of the science, they are producing scientific knowledge and the second group of people is those who are at receiver end, those who receive the communication, who receive the scientific knowledge. That scientific knowledge should be delivered to the receiving people in such a way that it should be transparent, it should be easily understandable, it should be in the proper way, it should be in the language which can be easily understandable by the people. Science communication is not just communicating science, science communication is very important to build the opinion of the people, science communication is very important to make the policy of the government.

For example, we see people make their own opinion based on what they listen from mass media, from social media, from the journal articles, from the newspaper and other sources. We can take the example of genetically modified crops. Now based on the news that they receive from the different sources, people build their opinion and this opinion is very important for the progress of the society. Hence the objective of science communication is to communicate the existing as well as the new knowledge accurately and clearly. Now the question comes, why do we need science communication? What is the basic need of science communication? You see, science is very vast, it is not limited to one subject or one domain. It is not limited to one field of study. There are various factors which are involved in different subject areas and disciplines.

Hence acceptance of scientific information varies. You see science is ever growing, science is multifaceted, science is multidisciplinary, there are various disciplines and areas of science. Hence the proper communication of the result of those areas and discipline is very much required so that people understand the relation between the different domains of science and they can take proper opinion and intelligent opinion. You see, science has grown tremendously over a period of time. There is development in science and technology. There is development in research institutions. There is growth in engineering and technology institutions. All these factors led to the growth of science.

Hence the result coming out from this growth should be delivered to the people in such a way that the common people can understand it. Hence science communication is very much required. Nowadays people are making opinions based on science communication. what they read, what they get information from the mass media, from the social media, from the different sources of science communication. You can easily say the example of artificial intelligence. Now artificial intelligence is a buzzword.

We all are worried about how artificial intelligence is going to play a role in the future. how artificial intelligence is going to affect our day to day life. Important and good piece of study which can be understood by different people would be a good source to make the opinion about artificial intelligence. Hence since artificial intelligence is not related to one facet. It is related to computers also. It is related to other domains of study also. Hence it is required that proper information should come to the people to make their opinion. In science communication there is involvement of a wide range of stakeholders. You see there are different stakeholders and some who do not know at all.

As we know that we all, most of us study science at our school level. When we grow up then our knowledge of science is limited. Even those who are practicing science, they are masters in their domain but they don't have knowledge of every domain. Hence communicating that scientific knowledge to different stakeholders like administrators, policymakers and many others is required. Their needs would be different. For example, if a person is engaging in science, he or she may master his or her field. But, it is not necessary that every domain the person knows. Hence if we are going to communicate science to the person, the same piece of information would have different opinions for a common man and the person who are engaged in science. The same piece of communication would have a different perspective for the people who are administrators or who are policymakers. Hence we know that there are different stakeholders. And to meet the needs of those stakeholders we need science communication. And that science communication should be such a way that the needs of different stakeholders can be met. Now what are the different key aspects of science communication? We have already seen what science communication is, how science communication is important, how science is multifaceted, and how there are different stakeholders in science communication.

Now questions come, what are the different key aspects of science communication? One of the most important aspects of science communication is disseminating scientific knowledge. The basics of science communication is to disseminate scientific knowledge. It is an engagement and dialogue between the provider and recipient. It's not only a one way process. it's not like that, that you say what you want to say and other people are not doing anything. Other people are building their opinion based on the communication which the person is receiving from.

Another aspect of science communication is accessibility and clarity. Science communication should be accessible to all. it should be clear. There should be no opaque, it shouldn't be opaque in nature. Crystal clear information should be there so that it helps people to make their opinion, to build their opinion and to use science in their day to day life. Because if there is correct information, if there is clear information, if there is clarity information then it will build trust and confidence.

And ultimately science communication is also required for the promotion of scientific literacy because I have already mentioned that there are different stakeholders from the common man to the person who knows about the science, the layman, the policymaker, the administrator. Hence for the promotion of scientific literacy we need science communication and its impact on science policy. So these are the different aspects of science communication. Now the question comes, what's the importance, why science communication is very much important to us. Science communication is the bridge between science and society.

You know there is a development in every aspect of science. There is a development in research and technology. There are various institutions which are engaged in research and development policy. They are doing research. But, unless and until the research of the institution, the research of the entity is not conveyed to the society, it's not possible for the society to come to know about what's happening. Hence science communication is very much important to bridge, to act as a bridge between science and society. It's also important in the personal and professional arena of our life. You see in our personal life, in our professional life we take various decisions. It's very important that our decisions should not be biased, our decisions should be based on facts, whatever we have.

Hence here science communication plays a very important role to make unbiased decisions, to make logical decisions. Hence it is very important to develop our personal as well as professional arena. Science communication is also important to learn about science. Science communication is important because people come to know about science, people come to know about the development of science and they debate over the development in their life. they make their own opinion based on the information whatever they have. That's the very important aspect of science communication.

Science communication is very important for public engagement and support. because it bridges the gap between scientific research and the public allowing everyone to engage with and benefit from scientific advancement. You see, science basically acts for the betterment of society. And that betterment can be conveyed to the society in the form of science communication. Hence with science communication there is public engagement. There is public support. And individuals take informed decisions. Science communication is also very much required to address misinformation.

You see now there is information overload, information explosion. everywhere there is a plethora of information. We are finding plenty of information everywhere. It is very tough for us to find authentic and accurate information. Here the role of science communication is very much important. Science communication handles the problem of misinformation.

It provides us with the right information at the right time to make the right decision. Hence to avoid misinformation to help the people in taking the right decision in their personal and professional life, science communication is very much important. Hence we need science communication so that people think critically, people avoid misinformation, people avoid fake information. Science communication also has policy impact. You see there are various policies which are impacted by science communication, proper science communication.

For example, if the government wants to make any policy, people have the knowledge about the area in which the government wants to make the policy, any particular scientific policy. Then there would be a rational debate. because people know about the area where the government or any organization wants to make any particular policy. Hence people can influence the policymakers. Another would be influence on the policy decision and it helps in making evidence based policy on the various burning topics. Since the people will have the knowledge of the topic with the information of proper science communication, hence it would be helpful in making an evidence based science policy.

This is one of the important ways by which science communication builds the opinion among the people and helps in making information. There are various reasons for the publication. But one of the most important is to spread scientific findings. Scientific findings can only be spread by publication. If there is no publication, it is not possible for people, it is not possible for the citizens to know about what is happening. Hence to spread scientific finding publication is required. And that publication can be in various forms. This can be a journal article, this can be published in proceedings, this can be published in mass media, like newspaper, magazine and other mass media channels. All these basically spread scientific knowledge. Hence to spread scientific knowledge publication is required. Another important thing for the publication is to protect intellectuals. When we talk about intellectuals that means when any scientific development happens, someone contributes for that development. Scientists and researchers contribute to that development. Now that development, that particular invention, that particular discovery, that particular news is the intellectual output of the person. Hence the person wants to protect his or her intellectual output. That protection can be granted in different forms like patent and other. Hence to protect intellectual output publication is required and to gain fame.

When we talk about gaining fame that means we know we all are human beings. We all want to get known by other people. If we are doing something, if a scientist is doing something, if a researcher is doing something then it is his or her desire for people to know about that invention, that particular discovery, that particular piece of news. Hence for getting, for gaining fame publication most of the scientists publish it. Hence we can say that scientists make their work available to the public and they also access the work of their peers.

It's not only one way communication. it's not only that a scientist is publishing, a scientist is publishing but at the same time the scientist is accessing the work of other scientists. It helps scientists to broaden their knowledge. because as I have already said that science is progressive, it's not constant and that progress can only be possible when there is two way communication. Now there was Logan Wilson. Logan Wilson in 1932 used the term Publish or Perish. As the term, as the phrase indicates it says that any scientific outcome should be published, a scientist should be published otherwise the knowledge would be in silo, it would be Perished.

The knowledge won't be available to different people, the knowledge won't be available for discussion, knowledge won't be available for growth. Hence the Publish or Perish concept given by Logan Wilson in 1932 is very much required for the growth of science. The growth of science can only be possible if there is publication. Why is the publication necessary? We have already covered some of the factors which are required for the publication.

Hence publication is very much required. In 1967 Derek John D. Sola Price suggested that research findings should be published. He said the nation publishes or perishes. That means the prosperity of a nation depends upon publication. How much does the nation publish? It's very much required for the growth of science. It's very much required for the growth of a nation that publication should happen.

Price was a physicist and historian of science. And he was credited as a father of Scientometrics. Hence the contribution of Price, Derek John D. Sola Price is very important in science communication in Scientometrics. Now Price's work primarily focuses on quantitative analysis of scientific publication. And when you talk about the quantitative analysis of scientific publication, the scientific publication which can be measured.

There are two ways of measurement. One is quantitative, the other is qualitative. Price's work is basically for quantitative analysis and in the later part of my lecture I will be discussing different quantitative techniques which can be used for the measurement of science. Now you see since the De Sola Price has mentioned that science communication should be measurable. Now the question comes: what is the need for the metrics in science communication? Why do you need different measurements in science communication? The first important thing for metrics is to get the quantitative and qualitative evaluation. We need metrics in science communication to find out to evaluate both quantitatively and qualitatively.

We need to evaluate science communication in respect of quantity and quality. Hence we need metrics in science communication. Another important thing for the need of metrics in science communication is understanding audience engagement. You know that there

are different types of audiences engaged. We have already seen that common man, policymakers, scientists, administrators, there are different people who are engaging in science.

Some of them are producing science, some of them are actually using science in their day to day work to understand the audience engagement and how they are using it. For example, a scientist is using an article in writing another article in the form of a citation. A common man is using that piece of information to make any logical decision. Another administrator took that piece of article for another purpose. Hence to understand the audience engagement we need science communication.

Science communication is very much important to demonstrate impact and values. If you are measuring science communication then we demonstrate how impactful that particular piece of research is. Hence to demonstrate the impact and value of that piece of research we need the measurement of science. Science measurement metrics are also needed for accountability and transparency. Based on the quantitative result there would be transparency among the different pieces of science communication hence is required.

Metrics are also needed for resource allocation because based on the metrics we can find out which area of science needs funds where there are more chances for research, which area is lagging behind a particular area. Hence for all these things for allocation of funds science communication metrics are very much required. Now, since science is multifaceted, having different stakeholders. Hence it is not possible for a single metric to measure the various aspects of science communication. Hence there is a need for different metrics to measure science communication. because no single metrics can capture all aspects of impact and values of science communication.

Hence we need different metrics. Different metrics provide different insights and provide various dimensions. For example citation impact, societal relevance, collaboration network all these are measured with the help of different metrics. There is a need for more than one metric to measure science communication. because more than one metric can judicially measure the different aspects of science communication. Using a combination of metrics enables a more comprehensive and value added evaluation of research. If you are using different metrics, if you are not fixing yourself with one metric then it is possible that a combination of metrics can be used to make the measurement of science communication more fruitful. using various metrics enable a more comprehensive study of science communication. You see it allows a tailor communication and measurement approach to meet the specific requirement of the diverse audience.

Different metrics can satisfy the needs of different audiences. Hence we need different metrics. There are different stakeholders whose needs are different hence we need

different metrics or various metrics to measure science communication. These also needed different metrics are also needed to background understanding of different domains of science and also the methodological ranges. Because you see as we see science is spreading its multidisciplinary is multifaceted. there would be the different methodological processes and there would be the different methodological ranges to cover all these aspects. We need different metrics so that various aspects of science communication can be measured and can be covered. Different metrics are also needed for balance assessment and transparent evaluation. Hence we can say different metrics in science communication offer a more comprehensive contextual and balanced understanding of science communication. It helps to judge the research impact and quality. They cater the needs of diverse stakeholders and being a multidisciplinary nature multifaceted nature it caters to the need of science also it can cover different aspects of science also. Now we have covered the need of the different metrics. Now a milestone achieved in 1963 when Eugene Garfield founded Science Citation Index SCI. This is a very important aspect in measuring science. Science Citation Index helps to measure science using quantitative and objective methods.

Basically when Eugene Garfield developed the Science Citation Index in 1963, his aim was to use citation analysis as a legitimate and practical tool for evaluation of scientific production. He wanted to evaluate scientific production by using citation analysis. We will be discussing what citation analysis is. I think you might have learned about different aspects in your previous lectures. We are not going to detail about that. But, citation analysis is a tool by which we can measure the impact of a researcher. When Eugene Garfield founded this in 1963 his aim was to use citation analysis as a legitimate and practical tool for the evaluation of scientific productivity. And this was the most important step in science communication.

Then in 1969 Pritchard coined the term bibliometrics. Pritchard was a British librarian and information scientist and his study was focused basically on publications, citation analysis and the measurement of citation research output. And in 1969 Pritchard published an article. The title of the article was statistical bibliography or bibliometrics in the Journal of Documentation and in this particular article he introduced and defined the term bibliometrics. As we know biblio means book, metrics means measurement. But, Pritchard defined this term as something broader in aspect. He defined bibliometrics as the application of mathematical and statistical methods to books and other means of communication. He was not confined to books only. But, his definition of bibliometrics covers the application of mathematical and statistical tools to measure the quantity of information in books and other means of communication.

And basically it was the Pritchard work which laid the foundation of the development of bibliometrics as a distinct field within the library and information science. In library information science bibliometrics is a very important field. And it was the work of

Pritchard which led to the foundation of this particular metric of bibliometrics. Now when we come to know about bibliometrics, bibliometrics is nothing but it is a quantitative method to study and analyze publications and their impact. We use statistics. bibliometrics is a statistical method of studying and analyzing publication and its impact. This involves measuring and analyzing various aspects of scholarly literature such as publications, citations, authors, journals, institutions to gain insight into the patent trend.

That means bibliometrics is not confined only to the information which are in the form of books. But,the area of bibliometrics is quite large, and is quite wide. It covers scholarly literature in the form of publication, citation, author, journal, institution. And all these help to get insight about the happenings in that particular domain in that particular field. Bibliometrics primarily focus on a statistical analysis of bibliographic data. That means when we talk about bibliometrics, it focuses on a statistical analysis of bibliographic data. There are various types of bibliographic data. And with the help of bibliometrics we do the statistical analysis of that data, its significance, its influence, its quality. All these aspects of the data are covered in bibliometrics in a scope of bibliometrics.

Some of the key aspects used in bibliometrics are publications count. I am not going to discuss in detail but I am just going to give a glimpse of how they work, what they are and what their scope in bibliometrics. When you say about the publication count as the name indicates it is counting the publications and publications can be counted by the name of the author, by name of the institution, by the name of the country. Hence when we talk about publications count that means it provides an insight. It provides an overview about the publication of the country or publication by a particular author or publication of that particular institute. For the example you see I have taken this data from Scopus and here you see the affiliation name is Indian Institute of Technology, Delhi and the range is from 1952 to 2024. That means if we want to count the total number of publications using the Scopus database between 1952 to 2024 for Indian Institute of Technology, Delhi we found that there are 30,658 documents. This is not the complete number, this number is given by one database that is Scopus. This number can vary because all the publications of IIT Delhi might not be indexed in Scopus. But you see, using this database we find out what is the output of Indian Institute of Technology within a period of time, within a range of time. This means we are measuring the output of an institution quantitatively.

If we are taking the affiliation country as India that means if we are taking country as India and find out what is the total publications from India indexed in Scopus in the range of 1866 to 2025 then we come to know more than 3.2 million documents published from India during this particular period range and which are indexed in Scopus. That means what we are doing as a country we are analyzing the output of India, output of India within a particular period. This document won't be the same number, it would be higher because all the documents published from India during this period are not covered under

Scopus. You see the same publication count if you want to analyze how they are increasing year by year. You see this graph which means it reaches a peak that is going upward from 1866 to 2026 and then in a particular year it reaches its peak and then again coming down and coming up. That means we are analyzing the productivity, the publication count, the scientific productivity of a country.

This is bibliometric analysis in the form of publication count. If we want to see that out of 3.2 million documents whatever we more than 3.2 million documents which we found using Scopus database who is the most productive author that means who has written the most number of articles then we come to know there is an author by the name of Singh B and the article written by this author is approximately 2400 here. So that means that we are quantitatively analyzing the publication of that country with the help of a database called Scopus. If you want to further analyze this and we want to find out under which subject domain or as a country affiliation of India then we come to know if you want to further analyze and find out publication count or publication document by subject area then we see the share of engineering is most followed by medicine.

So that means during the year the largest number of articles published in the field of engineering followed by medicine and so on. So that means what we are doing is analyzing the publication count of an institute we are finding the most prolific author. We are finding the authors who have written the maximum number of articles. we are finding the subject which is producing the maximum number of documents. That means we are quantitatively analyzing the different aspects of publication count using bibliometrics. This is a technique of bibliometrics. Another aspect of bibliometrics is citation analysis. When we talk about citation analysis, citation analysis is examining the citation within a scholarly publication to understand the relationship between works, identifying influential articles and authors, and major impact of the research.

If we use the same database that is Scopus and without the earlier result of 3.2 million articles from India. if we try to find which article has the maximum number of citations. Then we came to know that an article published in IEEE transaction on evolutionary computation published in the year 2022 received more than 35000 citations. That means we are analyzing the citation received by an article over the period of time using bibliometrics. Another is publication count. If you want to see out of 3.2 million publications, whatever we found earlier, what are the different types of publications that appear. Then we come to know the share of articles is 70%, followed by conference proceedings. Hence using bibliometrics we are analyzing the types of documents which appear in different subject areas in different forms of communication in a Scopus database during the particular period of time. If you want to see the growth of documents, year by year, documents per year by source then we come to know there is AIP Conference Proceedings which contributes the maximum number of documents which is approximately 30,000 over the range of time.

Hence we analyze using bibliometrics what are the different types of documents produced by a country year by year. Finally we can do author analysis, we can investigate the productivity of authors. We can investigate the collaboration of authors. whether the author is working in isolation or is a single author, or which other country the author is collaborating with. What's the impact of the citation received by the author? How many citations are received by the author during the period of time. All this can be done using bibliometrics over a period of time. For example if you want to find out which author from India has received the maximum number of citations. Then we come to know the authors Singh B, AlHaddad K, Chandra A received approximately 2000 citations for the article published in IEEE Transactions on Industrial Electronics. If you want to analyze the publication of the author of India we already saw that Singh B is the most prominent author from India in our earlier slide. Now if you want to limit ourselves to Singh B only in a Scopus database and using an affiliation country as India we can find that he received approximately 2000 citations for his article over the period of time.

Similarly journal analysis can be done based on the impact factor, citation pattern, in which quartile the journal is, and what the publication policy. All these are part of bibliometrics. For example using a Scopus database if you want to find out which journal has the highest percentile. Then we came to know there is a journal called this particular journal CA of cancer journal of clinicians having the highest percentile of 99% having the citation of approximately 70,000 and percentage citation 94. That means what. We are analyzing the journal based on the different factors and all these are done using bibliometrics. Some other aspects and techniques used for bibliometrics are co-citation analysis, bibliographic coupling and the various other metrics. I am not going to explain all these things you might have learnt about all these in detail in your previous lectures or incoming lecture.

Now the question comes if bibliometrics is so important then what are the different applications of bibliometrics? Bibliometrics is very much required for research evaluation. We have already seen how bibliometrics help us in finding the most prolific authors, finding the articles which received the maximum number of citations, and finding the journals which are at the highest percentile.

Hence bibliometrics is very much required for research evaluation; it is widely used for evaluating the research, and the quality of research. Bibliometric is also used for journal ranking based on their impact factor, based on their influence, based on their specific field in which quartile they are falling with. Hence bibliometrics are used to give the ranking of the journal. They are used for the journal ranking. Bibliometrics is also used for research trend analysis. I have already explained in my previous slides how bibliometrics is used in using Scopus data and how different subject fields are associated for a particular area of growth.

Hence bibliometrics is used for the analysis in the research trend of which particular area is growing, which particular area is receiving the maximum number of citations, which particular area is not sought after or maximum sought after among the researcher. All these things are needed for analysis. And it is very important for the development of science. Hence bibliometrics is used for the research trend analysis. Bibliometrics is also help in research collaboration, it is help in network with bibliometrics we come to know which particular scientist or which particular researcher is working in that domain of research, what the area they are working with, which are the different networks, which are the different agencies which they are collaborating with. By that way bibliometrics is used for the research collaboration. Bibliometrics is also used for the funding decision because if a particular funder or a government wants to fund then with the help of bibliometrics the government or the funder come to know which are the trending area. which area there is a need of fund, which area there is need of development. Hence it's important to use bibliometrics in funding decisions. And it helps the funder to check informed decisions based on what the quantitative data we have, what the data we have available with. A funder can take informed decisions based on the data available. Bibliometrics is also used for science policy and planning. Science policy and planning is very much important for the progress of any country. Now with the help of bibliometrics the planners, the policymakers come to know that with the help of bibliometrics the planners and policymakers come to know which area they need to give more strength, which area their development is needed.

Hence for science planning and policy bibliometrics is very much needed. Bibliometrics is also needed for the patent analysis for getting the intellectual output of a particular country or a particular research organization or a particular researcher. Thanks for your patience.