Course Name: Technical Communication for Engineers

Professor Name: Prof. Arun K. Saraf

Department Name: Department of Earth Sciences

Institute Name: Indian Institute of Technology Roorkee

Week- 06

Lecture 06 Data inputs and analytical methods

Hello everyone, Namaskar and welcome to a altogether new discussion under this technical communication for engineers course. And in this lecture, we are going to discuss about data inputs and analytical methods, how this would be presented and what are the important points or importance of all this. because depending on the type of data which one is using in his research for trying to publish that how it should be presented and some difficulties, challenges, intricacies we are going to discuss. So, we know that so far what we have done, we have completed up to background 5 and this is data inputs analysis. data sets, field data sets, etc. And of course, analysis methods are also here.

So, before that I would like to emphasize you know because the data word is being used sometimes very wrongly. Some people say datas and therefore, I thought that first I will have a discussion what is basically data is. Singular form is datum and plural is the data. So, therefore datas is a completely wrong word of English and data itself is a plural.

This is the important point here one should know. Further what is basically data means? Data means one raw fact. or one numerical value. Because if I say data, data are raw facts and suppose somebody has went in the field, measured certain thing that become our data, if there are multiple values which one has written then it becomes data. If a single value has been recorded then it becomes datum.

So, datum means one raw fact or data means raw facts. or one numerical value, it can be also numerical value. It can be alphanumeric values or just alphabetical value also. So, the word datum is rarely used unfortunately in the form of communication. And most of the time people are just using word data.

And data is when they say it is a wrong because data itself is a plural but this is how it is being used. So, we know that a data is a collection of information or data is raw facts that means collecting the information gathered through observations, measurements, research or analysis. So, almost everything which you have gathered through all these things observation, measurements that become our data. And, data can consist of facts, raw facts

I have already mentioned, numbers, names, figures or even description of the things that can also become data. And, it can be organized in the form of graphs, charts and tables.

Some of the examples here in this particular lecture we will see. in form of tables others but otherwise we will see in subsequent discussions. So, data can be described as a collection of discrete or continuous values that convey information. Now, there is another differentiation which we will see just within few seconds that data and information, these are two different things and that we will see what is the difference is here. Because it is a collection of discrete values that convey information.

Data and information are not same. Systematic record of a specific quantity. And, that becomes our data. A set of facts and figures are useful in a particular purpose like survey or in the analysis. So, whether you say a field investigations or it is in the lab investigations or through some other sources the data has come that can be considered under this data or data sets.

So, a set value of qualitative or quantitative variables about one or more persons or objects that becomes our data. And data can be presented in numbers, letters or form. Let me give you an example. Suppose the census data is collected. So, lot of variables which are collected in form of quantities or numbers and many other variables which are collected in terms of alphanumeric numbers or just proper noun things also.

So, everything about a household that becomes a data. So, and then for a village, city and state and country that everything is then data. And as we know that data can be presented in numbers, letters or other form and it can be stored and processed by a computer. So, this is the point now comes about the computer. So, now we are looking data, information and knowledge.

And there are three different things. If you think in hierarchy then data is at the bottom. It is a just simple raw fact, simplest form consisting of raw alphanumeric values. These can be alphanumeric, these can be purely numeric values or these can be just proper noun kind of thing or names. So, information is created when data are processed.

Now, one must be getting now sense what I am trying to say that once the data is processed then it becomes information. and organized or structured to provide context and meaning. So, when suppose if I give again example of consensus data like census data, then it is collected. The people are not analysing. Some people are just collecting the data and depositing and giving to some higher ups who would analyse, who would process the data.

And then data that becomes information that how much population is there, what are the male population, what are the female population and what is the economic background and all that thing that becomes the information. So, information is essential process data. Basically, it is essentially a process data. So, data has to be there. Data is at the bottom of this hierarchy and important.

Without data, you cannot generate information and without information, you cannot generate knowledge. So, knowledge is what we know, what we can understand. So, let me start that data after once it goes through the processing, nowadays we talk only computers. So, data processed through computers becomes information. And when this information is processed like geographic information system GIS where we are using geographic coordinates, then this information can become knowledge.

Let me give you an example. This is an example from JEE examination. When students fill their applications online, now portal I am talking. So, data is collected or students have given their data. Now, then it is processed and then as per the availability of seats in different centres or in different cities, then separate list is prepared.

So, that information that such and such student will go to such and such centre. That is only after the processing because the person who has provided the data does not have that control or does not have that kind of you know capabilities to do. So, now this information that how many people have appeared, whatever success rate and other. But, if we can attach the geographic coordinates and start plotting the successful students, then we can have this plot say having a map of India in the background and we would know that from where the maximum number of students are coming. And once we know this about the clustering or distribution because there is not going to be a systematic distribution as we know.

So, immediately questions will arise that why there are clusters in the country from where the successful candidates are coming. Is it because of better coaching in those areas or it is because of socio-economic backgrounds or any other factors. So, these questions then can be answered. And then once that we can answer then that becomes the knowledge. Knowledge that these centres are good.

If you go there and have a coaching, there is chances. So, that is what knowledge is. Now, there is one more thing which people have added in this concept of data, information and knowledge and that is the wisdom. So, again as I have said earlier that data comes in this pyramid if I consider that it comes at the bottom and this is the base of the pyramid. If that is not there, rest of the things cannot be built up.

And then you are having discrete elements, then you are having linked element, then in the knowledge you are having organized information and then you are having applied knowledge and that becomes your wisdom. So, the purpose of all this discussion is that whatever the data which is going to be used in your research, one should bring that analysis, up to this level, knowledge and wisdom level, at least knowledge level. That what my paper should not only contain the data and information. My paper, my work, my thesis should bring the knowledge about a particular specialization, particular work and so on. And if we can add to this wisdom that would be the real ultimate aim.

But at least information should be there so people can understand. Another, this cartoon is quite good so I thought that I will discuss this. That data is just you know points you are having then you start getting different information that some points are having different colours that means there are some differences in these points that becomes information. and that is only after the processing of the data. Then this information when it goes with the location and connectivity, it becomes the knowledge.

And then when knowledge goes further then it becomes the insight as you can see that this might be the beginning, this might be the end. Then it is the wisdom and one more step in this that it is the conspiracy theory that if we linked these points, probably we are getting a shape of an animal. But at this stage, at data stage, information stage, knowledge stage, inside and wisdom, you do not have that information. Anyway, so the best thing is if we can provide insight or you know knowledge in our work that would be wonderful. Another way of this important is that the knowledge has to end with decisions.

If I try to link with certain modelling or certain works related with groundwater. So, some data has been collected, it has been processed through computers then the GIS based work has been done that where the recharge structure should be constructed and how much time it would take to recharge our groundwater regime, how much area will get inundated and so on so forth. So, that will become knowledge. And with this, you can provide various scenarios to whom, to decision makers so that better decisions can be made. And that should be the purpose of writing or presenting our data and further processing methodologies in our work whether it is a publication or a PhD thesis or article or anything.

That this is the data how the information has been generated, how the knowledge has been built up and then how better decisions can be taken. So, this section of the paper or thesis when we come after this background or review work, this section is important to provide complete information about data in the study. complete information means here that from where the data has been collected, how the data has collected and what are the

instruments which have been used, what was the frequency of the data. If I link say weather data, weather data about temperature, humidity, and rainfall every day it has been collected, hourly it has been collected and so on.

That information should come. Which type of instrumentation has been used whether automatic weather stations or manual, that information should also come. And where these stations were, if there are more than one station, where these stations were there, that location should also come. and further what was the elevation, what was the surroundings in those meteorological station. That is why the word has been used here that complete information is required about the data. The source of data says must be declared with proper referencing and acknowledgement.

Now, another important point is sometimes another term is used information of data is called metadata. That means the data about the data, information about the data. And that always should be included in your work that how this data has come. Whether you have borrowed the data, you have downloaded the data or you yourself has generated data. So, that information about the data must be there.

So, no further question will be asked about that if you have already provided complete information. Source of data sets, as I have just mentioned that because nowadays lot of data like remote sensing satellite data, lot of data is available. So, you can download the data, there is no problem. But you must provide from where you have downloaded the data and what kind of data you have downloaded. And details about that, so satellite details, sensor details, resolution, time, everything should be provided and proper referencing and acknowledgement that will go in the acknowledgement section that is almost just before the references has to be there.

If data has been collected by the researcher or student then it should be mentioned clearly that this is how the data has been collected. And it is preferable to provide data used information in tabular form. If it is possible to provide in a graphical form, that is very good. But if it is not possible to provide, then at least provide in tabular form. I am giving an example here of a paper, this is my own publication, this is publication from one of my students, Of course, my name was there et al.

So, that is why I said my own publication. So, here see this data used and in this study what kind of data has been used it is mentioned. And then 1 by 1 subsections are there like here topographic maps, then cartosat data and then everything and all details about the data is there. That means the metadata information is very much is required. that what kind of data you have used, when it was published, who has generated the data. All that information should come in your paper because this is very important.

Data is the basis of your work. That is the foundation that comes at the bottom of the pyramid. that is why it is important. Another way of providing as I have said in the tabular form, again it is from our own work and that is like here characteristics of Cartosat-DM, AsterD-GDM and SRDM. So, three types of digital elevation models have been used here and we have provided all relevant details about that. So, people can compare what are the differences between these three inputs and against these characteristics and they can satisfy whatever the queries or questions they might be having.

Complete information, remember complete information has to be provided. If anything which cannot be presented in a graph or thing like here the dates of the scene which we have used in the satellite work which is based on remote sensing. So, different dates and other details are there. Of course, this is again from our work. I am just giving the examples how sometimes because this data, this table cannot be presented in graphical form.

This had to be presented in this tabular form and there is no harm that you present data in tabular form. But if the data can be presented in graphical form, always prefer that mode rather than tabular form. Because here it is just simple data but a graph becomes sort of information. That is the important point. Now, key contents of technical communication here and in which now we are coming or indirectly we have also touched, Analytical methods, analysis methods, methodologies, approach, algorithms et cetera.

So, now in this section of this current lecture, we are going to discuss. Now, depending on your work, your analytical approach, methods, methodology is going to be different. So, what we are going to discuss in this section is a very generalized approach. Depending on your work, you have to go through but must include all those details in order to get a successful publication or submission or whatever. So, methodology section of a research paper is again key part of any academic or scientific work.

How the analysis has been done? We have data inputs. We have already discussed that how data has been collected. Now how data has been analysed, processed that has to be described in this particular section. So, research methodology section in which we describe how a researcher systematically designed a study to ensure valid and reliable results. The main purpose of this section is to address what you performed and how you did. So, your results and how you have achieved those results that has to be explained here.

Letting readers assess the reliability and validity of your study. So, after describing data

then you should tell that how the processing has been done. You may have used certain algorithms, you might have modified the existing algorithm models, some approaches and already existing methodologies or new methodologies. All those details should come in this section. So, this section should begin with a description of theoretical or conceptual framework.

outlining the parameters that the author plans to operate within. That kind of beginning has to be there in this section. Further, you should outline the method selected and used in the research process. as well as factors for data analysis. And in this section, methodology section of a research paper manuscript, you should let your readers analyze your approach to the study.

In that manner, it has to be written. So, you should start first topic of your study. as like in introduction, like in background. So, here just try to link this part or this section of your work with the topic. What are the tools, techniques which one has used and how this analysis has been conducted. So, you know in case of background, we have discussed certain tips.

So, here also we are discussing certain tips for writing a methodology for research paper. Restart your thesis or research problem. Here again you are saying why this analysis is being performed as per the topic or your specialization or your research problem which you have taken. So, here you would restart again. But very brief manner, very focused manner just one or two sentences.

You are trying to link the methodology. Explain the approach you choose. Whether you have chosen certain algorithms then why you have chosen, certain methodology you have chosen, why you have developed, why other already published work or methodology were not useful. Every detail should come at this level. Explain any uncommon methodology you use.

If you have used anything but purpose here is to provide all details. about your approach, algorithm, methodology or analytical methods. Describe how you collected the data you used. That section will be there in data inputs but you have to link with that. that each section should be linked with another in a sequential manner.

Otherwise, the data section will lie in complete isolation. So, here you have to connect that one. Explain the methods you need to analyze the data you collected. So, why a particular approach has been selected that you explain. Evaluate and justify the methodology chosen humanly. That certain algorithms may have not worked in this area as people have already published.

So, I have chosen this particular algorithm, this approach, this method and therefore, I am choosing. Now, further tips are the focusing on information on how you intend to analyze your results. So, this information one should keep in mind. that you are going to provide details how the things have been or data has been analysed to achieve the results. And another point here, another tip is here describing how you plan and intended to achieve an accurate assessment of the hypothesis, relationship, objectives, patterns, trends, distribution associated with the data and research purpose.

So, all those details should also come here. As good information, accurate and focused information you would provide in this, there are always high chances that one will get success in this. Be very specific about your analysis. Here as I have said focused.

So, that is what meaning is. Specific, not here and there. Why this analytical approach has been chosen? And why others not? If they are available, very briefly you can mention and justify your choice. We are almost approaching the end of this discussion about analysis methods, methodology approach, algorithm, etc. So, basically in this section of the thesis or paper, manuscript, report, it is important to provide complete information. So, in summary, we are saying complete information has to be there. If there any improvement over published work, then it should be mentioned clearly.

And of course, it should be there. Otherwise, it is not you know recooked the same work which has already been done or published by others. So, there has to be always improvement about the knowledge, may be new datasets, may be new instrumentation, new type of analysis. So, the improvements over already published works should be highlighted in this section very clearly. And if one is using a particular methodology instead of writing in the text, the same may be presented through as easily understandable flowchart. This is the method which I always prefer most because we know that a picture tells thousand words.

So, instead of writing step by step methodology which you have chosen, very briefly in the text you can mention. But, use a flow chart and through those boxes and arrows you can show that how the analysis has been done, at what step what has been done. So, this is the flow chart approach. in case of methodology should be preferable, always good and impactful for this section of your work.

And finally, it is preferable to provide complete details about the methodology. This information, this part we have already discussed few times. Complete information, complete details are always preferable, always required to write a good section about

data methods, methodology, approach, algorithms, models, etc. So, with this I thank you all. Namaskar.