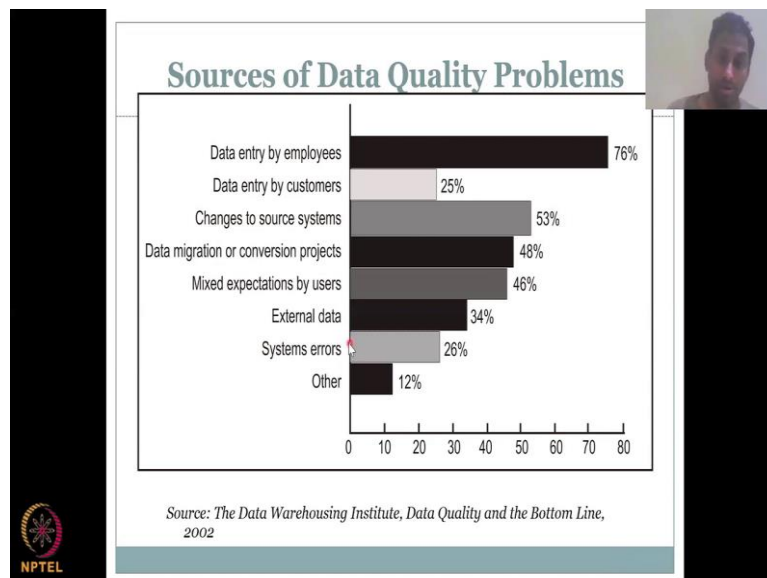


**Rural Water Resources Management**  
**Professor Pennan Chinnasamy**  
**Centre for Technology Alternatives for Rural Areas**  
**Indian Institute of Technology, Bombay**  
**Week 08 - Lecture 05**

**Rural Water Management Issues: Convergence of Funds and Data**

Hello, everyone, welcome to Rural Water Resource Management NPTEL course, week 8, lecture 5. We have been looking at the issues for rural water resource management, ranging from the application of funds, ownership, and also data issues. The last lecture we looked at why data is not enough? And even what are the issues in the current data available? We will continue that before we look at the recap of this week.

(Refer Slide Time: 0:56)



So, sources of data quality problems, data quantity is an issue and data quality is an issue. When you talk about data quality, it is either the data is there or not. Not a lot of data might be present, for example, instead of weekly is monitored, monitored quarterly. And not enough wells, for example, if we have monitoring of groundwater as a problem, you are not doing it at every village, but 1 to 2 stations per state.

So, these are the quantity problems, what is the quality problems? There are a lot of problems that arise because of data downloading, entering, et cetera. Let us look at what are the key errors as per this publication. it is a big error, it is big issue and that is why there is a lot of studies on this. So, look at it data entry by employees is pretty big, almost 76 percent of all the errors is because of data entry, a human interface that enters.

And also, automatic data, which gets converted into one from one form to another form data entry by customers is also there some others would also enter the data changes to source systems, this is your instrument which has been changed maybe the instrument was collecting data for 10 years. And then after that, it slowly gets into errors, because it is a machine and we have to replace the machine.

Data migration or conversion projects, this is where I said when you convert from one data together or from one instrument to your laptop to your dashboard, there could be some data issues because of format because of how you show the data unit conversion all these things make sense especially expectation by users. So, some people may want it at a different unit. So, that is where it is an external data is when data comes into your system, your system errors and others.

So, these are really small compared to the major which is the data entry by employees and then data migration and instrumentation errors. So, the quality is pretty big, the issues with quality and please understand that if quality is not good, then it is as equally to have no data which means data quality issues. So, both are not available. There is no point in having bad quality data, which out of which you cannot get anything.

So, in that report I mentioned in the last class, which was done by World Bank, we found that the data is shared by some countries of all South Asian countries, some countries readily shared the data, but the data was in such a format that not much could be taken out for climate change impacts. And then we have to stop and say the first recommendation is please collect good data. And this is very, very important even now, and most important now because of climate change.

(Refer Slide Time: 4:24)



So, let us look at the ways forward enough of the problem statements and looking at sources for data issues and errors. Moving forward, we want to understand how can we rectify these issues to collect better data... So, the data quality is there with data governance wherein data governance is kept as your data sharing and how do you really collect and store data into a particular matrix and then do the analytics.

So, the quantity part the governance part is on one side whereas you have data quality issues on the other side and there is an overlap, how do you manage get to get good data out. So, in data quality parsing, matching and profiling are done to get better quality of data parsing is conversion from one form to the other and matching data from primary secondary and profiling, you do have some overlaps.

Some methods where data quality can be enhanced in the data governance side data availability is big, you can have open source data and then some compliances of data, how do you use data who can store the data, those rules and regulations can be kept, but what is more important is the central part to move forward. What do you have is an overlap between the data quality team and the data governance team which is both require data cleansing, to have good data quality, data cleansing is important.

And same for data governance data cleansing is important, important of the methods enrichment of data, which means if you do not have good data or data gaps are there, you can enrich the data, you can add data to fill the gaps through multiple mathematical models, monitoring can be increased, increased in monitoring, generalize cleansing, you can clean the

data for using statistical and other methods and standard the solicitation wherein you create a standard or normal to the data.

In the data consistency issues, lot of steps can be taken consistent Analytics, you do not run to some statistical packages, just to make the data look good. It has to be consistent, what has been done across the world, what is accepted in literature should be done. So, this please make sure that you understand when you do data management, the rural data issues, there are a lot of statistical packages that can make your data look good.

But that is not the point. So, when you do analytics, data analytics and data work, it is important to have consistent analytics so, that you can be readily compared last 10 years last 5 years. And you can also manage to compare with other studies. Metrics and reporting are also very important. How do you create new metrics for data indicators and how do you report it and data selection?

So, this is also important you cannot take one station for explaining the rainfall in the village. And then suddenly, after 5 years or 2 years take another station, either you have to combine these stations as one unit, or only select one data and then make converted reformatted into another data. So, data issues do exist, both in quality and quantity. However, the methods in addressing these data quality and quantity issues are almost the same in multiple settings.

And that is what we are coming here is we can have some overlaps between data quality and data quantity management, but how to tackle data quality and data quantity and you can do these exercises in a good fashion. Remember, the aim is to better manage the water. And if you see if you want to make the data look beautiful, and then clean the data without any objective, then the whole point of data for rural water management is taking a right, which may just not correct.

So, if there is data issues, it is better to try to salvage or get as much information as possible. Otherwise, you just have to say the data is not good. Let me collect the data to understand the problem. Do not jump into the solutions with bad data. That is bad science. So, all of you are going to be young scientist, or young researchers learning these data methods. There are a lot of methods now to make the data look better or look good. Do not fall into those traps.

The basics have to be strong and how you select a model is very important for data cleansing. For example, I will give you an example of data enrichment for data gap filling, some people can use an average some people use 10 years average some people use nearest neighbors

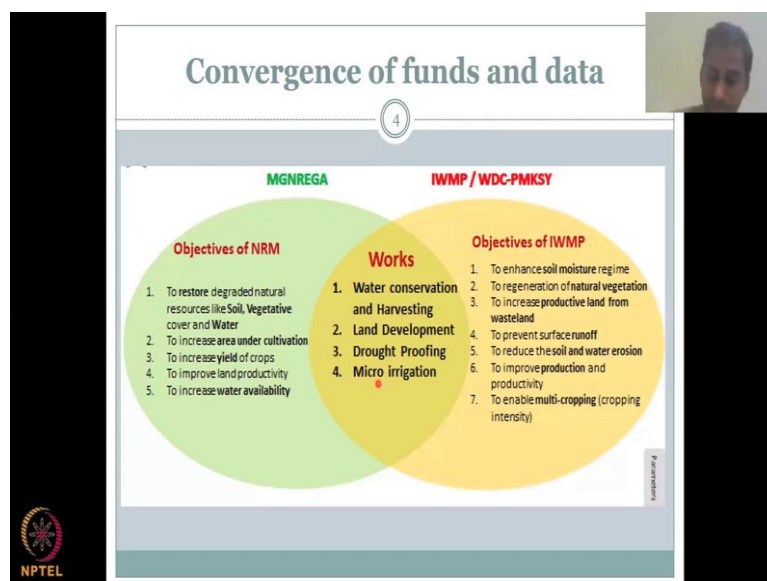
imputation methods. So, there are multiple methods to fill the data gaps. But what is more important is, is it making sense logically with other data is making sense.

Or you can just say if it does not make sense and you do not trust the methods, just say data is not there, but we will use the other years data to understand what is happening. And data has proved to have errors because it said instrument. Instruments have errors and they will be making some issues due to management and maintenance. Also remember that these instruments are not there to actually monitor these in many cases.

For example, the groundwater again I am saying an example because groundwater is measuring the pressure and the pressure is converted to a water level when you talk about some instruments preserved meters for example, preserved meters is working on a principle that spring is there and as the water level increases, the force on the spring increases and the compression force on spring is converted to water level.

So, see how many changes are there from water level to the force, force on the spring and springs how the spring compresses is converted to another parameter and the parameter goes back to your thickness of water. So, keeping it simple all these are not directly there to measure your water parameters. However, you could convert it into a parameter through algorithms and equations. So, there is a point where you need to be careful about which ones you use for water management, which methods et cetera we clear drawing.

(Refer Slide Time: 12:19)



Another thing is convergence of funds and data. So, as these works are being done, you could see that there is could be convergence of funds to do the work. But there is no data for

monitoring and fund for monitoring. So, why is that happening? Again, as I am saying the mandate for these agencies MGNREGA on the left and IWMP on the right is to manage the water not monitor the water is to create land resources, better land resources for agriculture, not to manage the pond or not monitor the agricultural activity on the on the field.

So, it is important to understand how you could put a monitoring in this funding network you need to be creative talk to the agencies and say that for better for example, I would put it in a way that to enhance soil moisture, I need to have better monitoring of soil to enhance multi cropping I need more data on what crops are necessary how they do it. So, the other one is micro irrigation.

For micro irrigation, you can automate it, which means water can be sent through drip irrigation or a specific point agriculturally precision irrigation. All this requires data a soil moisture can be converted to turning on tap a signal to turn on the tap for your plant. Now, how does how do you how does the instrument know that the soil moisture is low water should be applied it knows because the soil has a moisture sensor that is important for understanding the data for soil.

So, you can indirectly put monitoring costs in these frameworks. That is what I am saying drought proofing. How do you know if it is a drought unless you have soil moistures in the thing, you can say okay rainfall did not come for 2 years. But my point is rather than saying it is a drought after 2 years, you could be able to capture the soils water content or if you have a meter and then that data can be converted to management.

(Refer Slide Time: 14:55)

**Paradigm shift**

**Big Data**

Big Data analysis can identify previously unknown insights and changes which could help create new opportunities for water management.

- Define problem/issue
- Big Data methods
- 4Cs- Collect/Clean/Collaborate/Communicate
- Research/dissemination
- Policy and Banks
- Dynamic and cyclic process

**Extension**

(IWMI, Ahmed et al 2017)

Moving on, I would give you a very important paradigm shift which is needed for converting data into action and then preserving rural water resources. So, let us take an example of a child in the rural village waiting for water when her mother is pumping, the water is not coming. What happens normally in a normal situation when such an issue happens, there is a talk only between this network wherein the policymakers understand Oh, there is no water.

Like for example, mission is there to provide water for the public and they find that the water is not coming. They go back to the budgets, get more budget, and then put it back into the water issue. So, more pumps more augmentation of water. However, there is a gap. Let us take a live example, when Chennai ran out of water last year. So, water was brought in trains and tankers, Delhi, still tankers are coming in from outside. But is there a solution to it?

How long can this happened? So, I am saying suddenly, there is an issue, budget is released to get tanker water, good, happy, everybody's happy, you can move on, then the monsoon comes people forget it. And then suddenly another drought happens. So, it is not a solution. But a short-term fix. Suppose your tap is leaking, you put a tape around it, and then go get the plumber and then fix that.

So, until then the tape acts as a short-term solution. It cannot be kept on going like this. It is a waste of money and waste of natural resources. So, what I am proposing here in this class is, once that issue happens, yes, some short-term fixes can happen. But meantime, the agency plus the monitoring system has to evolve to collect data, there are multiple sources that it can collect data.

I will not get into the data that is beyond the course, but some of them are remote sensing data, satellite data, surveys, app, and then big data for capturing these activities. And then it goes to collecting all these into bins and storing them into data archives, and then converting them into maps or indicators and metrics, which we discussed in the previous slide.

So, what happens next is you would take the data, convert to map, transform and clean the data and then go into a data server and then it goes to explore clean, analyze and share. There could be some levels of cleaning. And then once the data is analyzed, it can be shared. So, once the data is clean analyzed, one and can be shared, then more research can happen and teaching and happen using the data.

And suddenly an extension network can come and say for example, your NGNEREGA can come and say okay, I will use your data for better water management. So, it goes like this, I

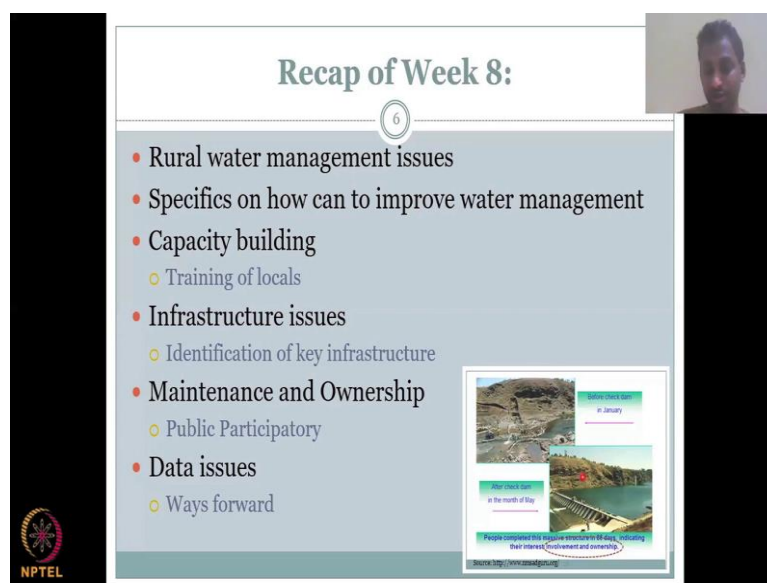
am going to give you the steps. So, define the problem issue, identify the big data methods, where you can collect the data. And then there are 4Cs which are collect clean, collaborate and communicate.

You collect the data using your methods, observations, observation data, primary data, secondary data, then you start cleaning the data to remove uncertainties, bias, errors, gaps, you use statistical methods or proven methods to they make the data better. And then you collaborate with other data or other agencies to make the data look better. And then you run your analysis and communicate the results.

What is the point of running this analysis, if you are not giving it for the general public, so, there is a lot of importance given to communicating the data. Once the data is communicated and results are communicated, a lot of research and dissemination can happen for better rural water resource management. Remember, that all this data comes to the platform where you can anyone can use it.

And then make sense out of it for management for research and dissemination. And then it goes back to the policies and bank where date where money can be now released on a better term, better long-term plans, residence dynamic and cyclic process. So, anywhere there is an issue the system can come back to one step backward, and then work its way through the system. Suppose I am going to the AI model, and I find that some of the data is not correct. So, I go back to the data I have collected.

(Refer Slide Time: 20:12)



**Recap of Week 8:**

6

- Rural water management issues
- Specifics on how can to improve water management
- Capacity building
  - Training of locals
- Infrastructure issues
  - Identification of key infrastructure
- Maintenance and Ownership
  - Public Participatory
- Data issues
  - Ways forward

Source: <http://www.nandaguru.org>

The slide includes a video inset of a man speaking in the top right corner. At the bottom right, there is a small image showing a dam and a check dam with text: "Check Dam in the mouth of Bay" and "Check Dam in the mouth of Bay". Below that, it says "People organized the regular maintenance of the dam, collecting their interest, involvement and ownership." The NPTEL logo is visible in the bottom left corner.



**Paradigm shift**

**Big Data**

Big data analysis can revolutionize urban water utilities and help cities and states make new opportunities for water management.

- Define problem/issue
- Big Data methods
- **4Cs-** Collect/Clean/Collaborate/Communicate
- Research/dissemination
- Policy and Banks
- Dynamic and cyclic process

(IWMI, Ahmed et al 2017)

So, let us take a recap of week 8. This week was very important to look at rural water resource management, after we looked at the hydrological cycle of the key parameters, et cetera. We looked at the rural water management issues, where are the issues, how are they coming, what are the key players for the issues et cetera. And then we looked at specifics on how to improve water management.

Remember in the rural water management issues, we looked at other structures are their importance given to rural water and then we went into water management issues, we looked at how we can improve the water management by having collaborative work between the private and the government sector. We looked at the role of NGOs, how they could help and most importantly of the issues was the capacity the local people have to be trained.

So, that they can address these issues in the longer term. In the infrastructure issues, identification of key infrastructures is important. There is a lot of losses of these structures, check dams washed away groundwater wells default all those terms and that happens when there is an infrastructure issue and non-representativeness. So, take your time to understand these issues.

Especially the infrastructure there is something called a one size fits all approach, which means suppose check dams worked in Tamil Nadu. It does not mean it has to work in Karnataka, Kerala, which is nearby Tamil Nadu or it is no reason it may not work in longer distances from Tamil Nadu. So, it is very important to understand that there is no one size fits all I cannot use one check dam across India.

There are multiple methods to construct check dam, what civil work you do for check dam, et cetera, et cetera. There are multiple designs for the second time. So, you need to understand all these four infrastructure issues. The most important issue we saw is the maintenance and ownership. It is not right for the locals. Be happy when there is a water infrastructure, but not taking care of it.

It happens to every single system I have seen, I have seen check dams being not maintained I have seen wells not maintained polluted tanks, people misusing the water et cetera. So, at that time, what happens is there is no ownership for maintenance and conservation of water bodies. And that is where we saw that in those cases, public participatory exercises have shown to be very beneficial for the local setting.

I showed you the case study of the NM Sadguru foundation in Gujarat, where a location and see they have taken a location with a lot of water losses from the watershed. And we will check dam the check dam was built using community participation and involvement because they have been struck by low water for a long time. What has happened is in beautifully the water has been stored and when the water is storing all the banks have been having increased soil moisture.

Because of what was their soil can pull the water to a particular level and they could see that the irrigation which is happening on the top from this water through lift irrigation. What is happening is the entire range is getting a lot of greenery soil moisture increase because of these check dams. Because of these infrastructures, the infrastructures may be small, but impact can be big depending on the geological setting.

So, to move on, the maintenance and ownership was the key with a lot of stress was given on creating these public participatory networks. We looked at NGOs as discussed what is it NGO, how they play a very vital role in bringing your stakeholders of government agencies and the public to communicate each other with each other and manage the water resource. Again, the NGOs duty is not to stay in one particular location, and take care of that village or district or block.

But it is mostly to learn these new methods apply it, and then take the learnings to the another center, and then rework the same thing to see how they produce it. Kind of upscaling. But more importantly, it is not just that there are some NGOs where they be on a particular location, because they suddenly start to invest in more branches for example, just not water. But let us talk about agriculture harvesting.

Let us talk about taking the harvest and preserving it and exporting to foreign countries, for example, grapes in Maharashtra are exported to foreign countries, and it can be tied up to the water because grapes have a good amount of water. So, all these are necessary. And even though you want to set up these rural water management structures, and you want to work with the government, at the end of the day, you need data and multiple issues for data that we looked in the class today.

And we also trained some of the students in using different statistical methods, or mathematical methods that are apt for taking the data issues down. When I say trained, I meant off you have been given a recommendation of specific methods. I am not talking about average mean more, or the AI methods of CNN KNN those kinds of things. But more I am talking about that you do have parsing you do have a data cleansing method, go and look at the literature, which method is more appropriate for your work, and then use it.

There are many, many methods that is why I did not bring all the methods here. But depending on the data you download. For example, groundwater and I need to interpolate into a surface I will use inverse distance weighting method IDW or Kriging method based on the thickness of the aquifer and if it is a shallow unconfined aquifer or find that... So, issues are there. But hopefully there are multiple ways to move forward.

And I hope this continues to go forward. For better rural water resource management. There is a need of ownership. There is a need of public participatory networks and NGOs and good government agencies along with them. And there is more importantly, to understand the data issues and create more data for better management. The end of the day. I would like to conclude by saying you cannot manage what you cannot measure or monitor. Thank you.