Different Types of Data Collection Method with Advantage & Disadvantage

Unit: III

Semester: II

Paper Code: GIS 10
Name of Paper: Research Methodology and Application of Remote Sensing and GIS Techniques in Research

PG Diploma in RS & GIS

Dr. SHYAMA PRASAD MUKHERJEE UNIVERSITY, RANCHI

Introduction:

Data is a plural form of a Latin word *datum*, meaning 'facts or statics used for reference or analysis'. In social sciences, it is treated as a plural noun and used with a plural verb, for instance, we write 'the data were classified' and use the term datum in its singular form. According to Hicks (1993: 668), data is "a representation of facts, concepts or instructions in a formalised manner suitable for communication, interpretation, or processing by humans or by automatic means". Data is basically the combination of different kinds of observation either in quantitative or in qualitative form which is needed to express any problem of the society as well as to find out possible solution of that problem through using appropriate methods of calculation in a research study. Data consists of full information regarding a particular aspect; it summarises a set of facts. Reading, scores, points, goals, measurements obtained from observation of salient features of a certain reality and systematically recorded in a standard form, are, therefore, called data (Singh 2011).

The procurement of observation from different sources for a single object is known as collection of data. So, data is the only leader, whose systematic arrangement can help to reach the actual objective of any study. In any research, data are used to draw conclusions, which help us to understand our environment better. We use data to adjust ourselves to the environment or to change it.

Types of Data:

There are many ways of classifying data. A common classification is based upon *who collected the data*. The term 'primary data' is used to refer to data collected by the investigator himself/herself for a specific purpose. For instance, a scholar carrying out field research is bound to generate primary data from his/her point of view. In many instance agency/organizations collecting primary data allow others to read and use them in original form. Census data, for instance, are primary data. Similarly, National Sample Survey Organization (NSSO) or National Crime Record Bureau (NCRB) do publish their original data and make those available to researchers. As against this, the term 'secondary data' is used to mean data collected by someone else for some other purpose, but being utilized by the investigator for his/her own research purpose. For example, we all use research findings of

other scholars to analyse a situation or to prove a point. While using such secondary data, the original data is not cited; rather the analysis of the concerned

author is used. It is therefore called 'secondary'. There are several advantages and disadvantages of both these types of data. As primary and secondary data also refer to two different sources of data, we would discuss these in some detail later.

Another classification of data is based on objective and subjective points of view. Objective data are those that are external to the individual and to which the researcher has assigned meaning a priori. Dimensions of a house, age of a respondent, education of the family members, etc. are examples of objective data. Subjective data are those relating to the subjective makeup of the individuals studied. Attitudes and opinions are typical illustrations of subjective data.

This objective point of view accepts that the data are validated in the sense that they must be measured and recorded accurately. Data can be shown to be true if they correspond to reality.

The objective view makes the following assumptions about data:
- ✓ They are factual, resulting from recording of measurable events, or objects.
- ✓ Data represent information and are the only way we can make information explicit.
- ✓ They record particular instances of reality.
- ✓ They are explicit as they are in a fixed in recorded form.
- ✓ Hence, they can be communicated digitally.

On the other hand, the subjective view makes the following assumptions about data.
- ✓ The data are not necessarily true or accurate as not all errors can be detected automatically and not everyone will necessarily agree that they are a true representation of a particular fact.
- ✓ Some data record subjective opinion, not facts. If data can represent opinions and concepts, they are not truly objective.
- ✓ Data have absolutely no meaning. They acquire meaning only when appropriated and analysed by a human recipient in a particular context.

Another popular way to classify data is to refer to their numerical and non-numerical characters. We prefer to call them quantitative (numerical) and qualitative (non-numerical)

data. Information about the age of respondents is quantitative because age of the respondents is expressed in number, say 38 year. It is possible to quantify many aspects of social living. For instance, apart from age, we can also quantify height, weight, literacy, income or distance. We also further divide such quantitative data into two more groups – discrete and continuous. As compared to numerical data, religion of a group of people is qualitative because religion cannot be stated in numerical terms, a person should either be Hindu or Muslim or Buddhist or Christian. There are many non-numerical concepts like love, hatred, relationship, sentiment, emotion, feeling that are difficult to quantify. Even though it is often possible to numerically present a qualitative aspect of social living, say measuring maturity of a person by his age and work experience, the depth of meaning attached to the concept is often lost in the process. Let us discuss the quantitative and qualitative types of data in some detail to reveal their nuances, advantages and limitations.

1.Quantitative Data:

The quantitative character of data is technically called numerical variable. Numerical variable is a quantitative character of an object/matter and its values can always be measured. For example, the students of a department may be classified according to their weight as follows:

Table 1: Weight wise distribution of Students

Weight in Kg.	Number of Students
35-45	12
45-55	109
55-65	68
65-75	17
75-85	6
Total	212

Such a distribution is known as frequency distribution. In this type of classification, there are two elements, namely (i) the numerical variable, i.e. the weight of the students in the above table in five groups or classes, and (ii) frequency, i.e. the number of students in each group or class. There were 12 students having weight between 35 to 45 Kilograms, 109 students having weight between 45 to 55 Kilograms and so on. Thus, we can find out the ways in which the frequencies (i.e. the number of students) are distributed.

Continuous and Discrete Variables:

A frequency distribution refers to data classified on the basis of some variable that can be measured, such as price, wage, age, number of unit produced or consumed. The term variable refers to the characteristic that varies in amount or magnitude in a frequency distribution. A variable may be either continuous or discrete.

A Continuous Variable, also called continuous random variable, is capable of manifesting every conceivable fractional value within the range of possibilities, such as the height of the children of an ICDS centre or weight of persons donating blood in a camp. In a continuous variable, data are obtained by numerical measurements rather than counting. For example, the height of an individual may have any value between, say, from 60 inches to 74 inches. It may be 66.4 inches, 66.47 inches or even 66.46589 inches provided we could measure the height so precisely. So a continuous variate can take any infinite number of values within a given interval, however small it may be. Height, weight, temperature, density etc., are example of a continuous variate. Generally, the continuous data are obtained through measurement. Series which can be described by a continuous variable are called continuous series.

A Discrete Variable is that which can vary only by finite "jumps" and cannot manifest every conceivable fraction value. The number of children per family, the number of members per Self Help Group (SHG), the number of classrooms per primary school are example of this type. The number of classrooms per primary school can take only the values 1, 2, 3, 4, ……, i.e. whole numbers; it cannot

take any value, e.g. a fractional value. Similarly, the number of members in SHG is a discrete variable. Such kinds of data are derived through counting. Series represented by discrete variables are called discrete series. The following are two examples of discrete and continuous frequency distribution:

Table 2: Classroom wise distribution of Primary School in Raina – I block of Burdwan District

Number of Classroom	Number of Primary Schools
1	52
2	87
3	106
4	271
5	169
6	43
Total	685

Example of Discrete Frequency Distribution

Table 3: Age wise distribution of members of SHGs in a Village in West Bengal

Age in year	Frequency (Number of SHG Members)
15-20	6
20-25	13
25-30	28
30-35	78
35-45	63
45-60	22
Total	210

Example of Continuous Frequency Distribution

Let us discuss some other examples:

(a) Family Size is a discrete variable. Because, it can take only some isolated values. The family size may be either 1 or 2, or 3, or 4 etc. It cannot take any value like we cannot speak of 3.46 members in a family.

(b) Family Income per Month is also a discrete variable. It can take values like Rs. 6890.50, or Rs. 6890.75, etc., but not any value as income in fraction of Paisa.

Although the theoretical distinction between continuous and discrete variation is clear and precise, in practical numerical work it is only an approximation. The reason is that even the most precise instruments of measurement can be used only to a finite number of places. Thus, every theoretically continuous series can never be expected to flow continuously with one measurement touching another without any break in actual observation.
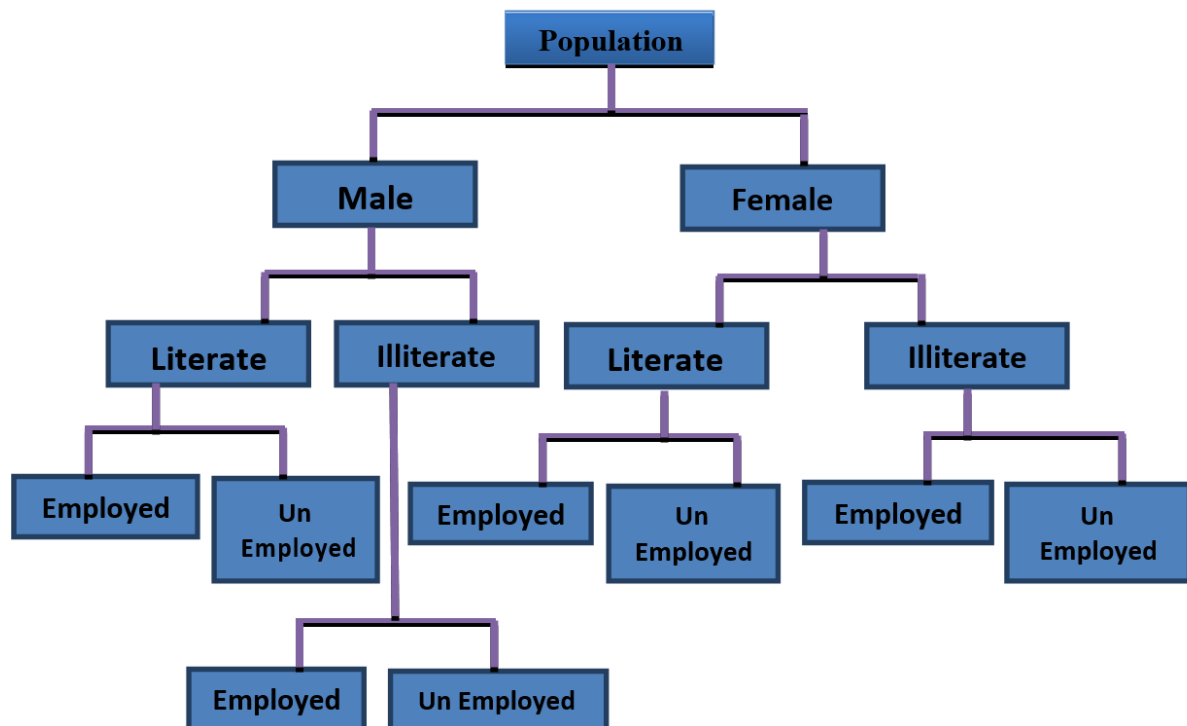
2. Qualitative Data:

In qualitative form, data are classified on the basis of some attribute or quality like sex, literacy, colour of hair, religion, etc. In this type of classification, the attribute under any study cannot be measured; it can be classified into different groups and one can find out whether it is present or absent in that unit under study. For example, if the attribute under a study is human population, one can find out how many persons are male and how many persons are female. Thus, when only one attribute is studied, two₁ classes are found. This type of classification is known as simple classification.
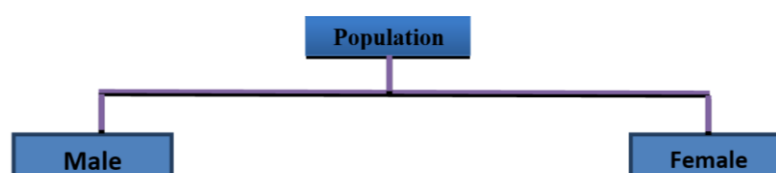
The type of classification where only two classes are found is also called twofold or dichotomous classification. If instead of forming only two classes, we farther divide the data

on the basis of some attribute or attributes so as to form several classes, the classification is known as manifold classification. For example, we may first divide the population into males and females, on the basis of the attribute *sex*; each of these classes may be further subdivided in to *literate* and *illiterate* on the basis of the attribute *literacy*. Farther classification can be made on the basis of some other attribute, say employment. The flow chart stated below tries to make the classification clear:

```
                          ┌──────────────┐
                          │  Population   │
                          └──────────────┘
              ┌───────────────────┴───────────────────┐
        ┌──────────┐                          ┌──────────┐
        │   Male    │                          │  Female   │
        └──────────┘                          └──────────┘
       ┌──────┴──────┐                        ┌──────┴──────┐
  ┌─────────┐  ┌──────────┐              ┌─────────┐  ┌──────────┐
  │ Literate │  │ Illiterate │              │ Literate │  │ Illiterate │
  └─────────┘  └──────────┘              └─────────┘  └──────────┘
   ┌────┴────┐      │                     ┌────┴────┐    ┌────┴────┐
┌────────┐ ┌──────┐ │              ┌────────┐ ┌──────┐ ┌────────┐ ┌──────┐
│Employed │ │  Un   │ │              │Employed │ │  Un   │ │Employed │ │  Un   │
└────────┘ │Employed│ │              └────────┘ │Employed│ └────────┘ │Employed│
           └──────┘  │                          └──────┘            └──────┘
                ┌─────┴─────┐
          ┌────────┐ ┌──────────┐
          │Employed │ │Un Employed│
          └────────┘ └──────────┘
```

3 Quantitative vs. Qualitative Data:

Glesne and Peshkin (1992) have made a useful comparison between the characteristics of quantitative and qualitative modes of enquiry as under:

Quantitative Mode	Qualitative Mode
Assumption	**Assumption**
Social facts have an objective reality	Reality is socially constructed
Primacy of method	Primacy of subject matter
Variables can be identified	Variables are complex and interwoven
Measured Etic (outside's point of view)	Emic (insider's point of view)
Purpose	**Purpose**
Generalizability	Contextualisation
Prediction	Interpretation
Casual explanations	Understanding under perspective
Approach	**Approach**
Experimentation	Naturalistic
Deductive	Inductive
Component analysis	Searches for patterns
Reduces data to numerical indices	Makes minor use of numerical indices
Detachment and impartiality	Personal involvement and partiality

Sources of Data: Primary and Secondary

The investigator, while collecting data is faced with one of the most difficult problem of obtaining or gathering the desired information or data. Utmost care must be taken while collecting data because data constitute the foundation on which the superstructure of analysis is built and policy decisions are taken. So, if the data are incorrect or inadequate, the whole endeavour becomes useless.

Data may be obtained either from the Primary Source or from the Secondary Source. Primary data are original in character and are generated through surveys/field work conducted by the Government, individuals, institutions and research bodies. Primary source usually has more detail information particularly on the procedures followed in collecting and compilation

of the data as compared to the secondary data. For example, data obtained in a population census by the Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, are primary data. Such data may be both quantitative and qualitative in nature. Any ethnographic data collected through participant observation, for instance, is also primary data.

By comparison, secondary data are those that were previously been collected by some person/agency for one purpose and these were merely complied from that source for use in different research. For example, a person/agency conducting a research might use the findings and analysis of any other researcher to argue a point. In the chapter on Review of Literature, in particular, scholars writing a thesis or article use arguments/findings of other scholars to arrive at certain assessment of situation. These data/findings/analyses are secondary for any one carrying out research. It appears that primary data lose their 'primary' character when subjective assessment and analysis are carried out on them after their collection. All researchers collecting primary data through field research do so as they have to generalise on those and link those with the existing body of literature. It may be argued that even raw field data supplied by the Census, NSSO or NCRB officials may become 'secondary' if these are compiled again or reorganised for the purpose of any research. Such transformation of data signifies increasing use of primary data for the purpose of analysis and understanding of social situation

Advantages for use of Primary data

- ✓ It is preferable to make use of primary source wherever possible for the following reasons:
- ✓ The secondary source may contain mistake due to errors in transcription made when the figures were copied from the primary source.
- ✓ The Primary source frequently includes definitions of terms and units used.
- ✓ Primary source often includes a copy of the schedule and a description of the procedure used in selecting and in collecting the data.
- ✓ Primary source usually shows data in greater detail.
- ✓ Through Primary survey one can get any data to express the problem better and to find better correlation among certain factors or to notice change in field situation; but in secondary sources all relevant data may not be available. There is, therefore, every need to continuously collect primary data to assess the current field situation.

Advantage for use of Secondary data

- ✓ It is highly convenient to use information which someone else has compiled. There is no need for printing data collection forms, hiring enumerators, editing and tabulating the result, etc. Researcher alone or with some clerical assistance may obtain information from published records complied by somebody else.
- ✓ If secondary data are available, they are much quicker to obtain than primary data.
- ✓ Secondary data may be available on all most all subjects where it would be impossible to collect primary data. For example, census data cannot be collected by an individual or research organisation, but can only be obtained from Government publication.
- ✓ In almost all research, scholars have to take note of research findings that are already been done on the chosen theme. Such a review of literature is required to make assessment of "what is already being known".

We all know that research is not 'reproduction' of opinions of other scholars, yet, use of such literature to develop an argument or to develop a distinctive position on the subject is essential. This is because, the basic objective of any research endeavour is to

a) reject an established explanation/theory,

b) modify them, and/or

c) strengthen them.

A social research must also have relevance, a depth of concern for social issues. How would one prove the relevance of any research or the uniqueness of its objectives? It is mainly done by relating the research questions with the broad body of existing literature in the field. Hence, secondary data are useful for conducting any research.

Choice between Primary and Secondary Data

Selection of use of primary data and secondary data depends on following factors:

- ✓ Nature and Scope of the enquiry: The nature and scope of a study dictates to a large extent whether the study would be based on primary or secondary data. As for example, if it is an evaluation study like the impact of MGNREGA for improving the standard of living of the population of a village, then the sources of data must be primary. On the

other hand, if we are involved in assessing the increase in per capita income in India in the last one decade, then we have to rely mostly on secondary data.

- ✓ Availability of financial resources: As the financial involvement is high in the collection of primary data, one might prefer to conduct a secondary data-based study with limited financial support.

- ✓ Availability of time: Availability of time is an important determinant. Collection of primary data requires much time. Contrarily, a review based on secondary literature may be done within a stipulated period. If we have to conduct a study to prepare the personal profile of the members of the SHGs formed under SGSY scheme in West Bengal within one month, then we have to depend on secondary sources of data.

- ✓ Degree of accuracy desired: The main barrier to the use of secondary data is degree of accuracy. As the society is changing very fast, a study based only on secondary data might not reflect the field reality.

- ✓ The collecting agency: Sources of data for conducting any study also depends on the collecting agency. It is possible for Government body or reputed institutions to carry our large and extended field-based project with proper budget, involvement of a team of researchers and time allocation. By comparison, an individual can collect field data only from a limited number of respondents.

Publications containing Primary and Secondary data

The following lists of primary and secondary data would allow one to look for appropriate sources of data most easily. It should however be noted that these lists are inconclusive.

a) Primary data
- ✓ "Census of India" published by the Office of the Registrar General and Census Commissioner, Ministry of Home Affairs (*censusindia.gov.in/*)
- ✓ Report Published by National Sample Survey Organisation, Government of India, Kolkata *(mospi.nic.in)*
- ✓ "Reserve Bank of India Bulletin" issued monthly by Reserve Bank of India, Mumbai (*https://m.rbi.org.in*)

- ✓ "Indian Textile Bulletin", issued monthly by the Textile Commissioner, Mumbai (*www.textileconnect.com*)
- ✓ "Crime in India" published by National Crime record Bureau of the Ministry of Home Affairs (*ncrb.gov.in/*).
- ✓ Central Statistical Organisation, New Delhi *(mospi.nic.in)*
- ✓ Directorate of Economics and Statistics, Ministry of Agriculture, Government of India (*eands.dacnet.nic.in*)

b) Secondary Data

- ✓ Official documents like "Statistical Abstract of the Indian Union" issued by the Central Statistical Organisation (C.S.O), New Delhi or "Monthly Abstract of Statistics" issued by C.S.O.
- ✓ Doctoral Thesis made available by University Grants Commission (UGC) at Shodhganga website (*shodhgangotri.inflibnet.ac.in)*. Interestingly, keeping pace with advances in the electronic world, many reputed universities across the globe have allowed awarded doctoral thesis to be read and used by other researcher as "unpublished manuscripts".
- ✓ Research articles available at several journals many of which are available on line.
- ✓ Books (even though most important source of secondary data is library, some online libraries like Library Genesis: Scientific Articles (*libgen.org/scimag*), Libgen (*libgen.info*), JSTOR (*www.jstor.org*), *Bookzz.org; booksfi.org* are helpful to locate books).
- ✓ Mass Media Output (all newspapers do publish their online versions).
- ✓ Encyclopaedia and Dictionaries (also available on line).

c) Issues Concerning Use of Secondary Data

Any scholar using secondary literature should also be aware of certain important issues. Thus, to begin with, scholars should be aware of the theoretical framework used to arrive at any conclusion at a particular document. Second, the methodological tool used to collect data might also generate a particular variety of responses. Hence the data source and type should be clearly mentioned while stating the findings of a research. Third, the time-frame of any data used and analysed in any document should be noticed. It is expected that a document based on data collected 15 or 20 years ago may be different from

the one published recently. Fourth, a literature may express the personal opinion or interpretation of the author(s) instead of any broad-based findings. This should be clearly

stated. Fifth, the usefulness of any document for any particular research should also be determined before using it. Finally, all secondary (and primary) sources used in any work should be cited properly to avoid the crime of plagiarism.

Methods of Collection of Primary Data

Like secondary data, students should also take serious note of various issues concerning the collection and use of primary data. We all know that different methods are adopted mainly to collect the *primary data* to fulfil the objective of the study.

(a) Direct personal observation or interview,

(b) Indirect oral interview/investigation

(c) Questionnaire sent by post or mail,

(d) Schedules filled up by investigators

(e) Case Study method

(f) Participatory Rural Appraisal method

(a) Direct Personal Interview

In this method the investigator (or interviewer) collects the required information in a face-to-face contact with the respondent. The investigator asks them questions pertaining to the survey and collects the desired information. Thus, in order to study the infrastructural facilities available for the delivery of quality education in the department of Sociology, University of Burdwan, the investigator has to meet the students of the department of Sociology, University of Burdwan personally and has to collect necessary information. The information is first hand or original in character.

When the researcher and the respondent are present in the same location, they face each other, the interview is called *face-to-face interview*. But in some cases, the researcher and the respondents are separated by the distance and the researcher uses telephone for communication. It is called *telephone interview*.

Focus group interview or focus group discussion (FDG) is a type of interview that facilitates collection of qualitative data. Even though FGD is a form of 'group interview', the difference between the focus group method and the group interview is by no means clear and the two are frequently employed interchangeably. In focus group interview, the researcher interviews a group of respondents at the same time. Focus groups typically emphasize a specific theme that is explored in depth in an unstructured setting as compared to any formal individual

interview. Alan Bryman (2009: 346) argues that the focus group is a form of group interview in which a) there are several participants in addition to the moderator/facilitator, b) there is emphasis in the questioning on a particular fairly/tightly defined topic, c) the accent is upon interaction within the group, and d) joint construction of meaning as individuals discuss a certain issue as members of a group, rather than simply as individuals. In other words, FGD allows the participants to respond to each other's views to build up a view out of interaction within the group.

(b) Indirect oral interview/investigation

In this method, the investigator contacts third parties to get information. The method is generally adopted in those cases where the information to be obtained is of complex nature and the informants are not inclined to respond if approached directly. For example, drugs addicts may be reluctant to provide correct information about their own habit. As a corollary, most of the Commissions of Enquiry or committees appointed by the Government collect primary data by this method. The accuracy of the method depends largely upon the type of persons interviewed and hence these persons have to be selected very carefully.

(c) Questionnaire sent by post or mail

Under this method questionnaire is sent to different respondents by post or mail. A request is made to the respondents through a covering letter and possible guideline for how to fill up the questionnaire and send it back within a given time period. This method can be adopted where the field of investigation is very large and the respondents are distributed over a wide geographical area. It is also relatively cheap.

But this method can be adopted where only all respondents are well educated. It involves some uncertainty. Cooperation on the part of respondents may be difficult to presume. It is also very difficult to verify the accuracy.

(d) Schedules filled up by investigators

It is most widely used method of collection of primary data. Here investigators are employed for data collection. The investigators carry with them printed schedules specially developed for the purpose. They fill up these schedules themselves on spot based on answers received from the respondents. The method is very popular and many a time it yields satisfactory result. Much of the accuracy of the collected data however depends on the ability and tactfulness of the investigators, who are given special training as to how they should elicit

the correct information by developing rapport and through friendly discussions. This method is adopted during the decennial census of population in this country.

(e) Case Study method

Case study method enables a researcher to closely examine the data within a specific context. In most cases, a case study method selects a small geographical area or a very limited number of individuals as the subjects of study. Case studies, in their true essence, explore and investigate contemporary real-life phenomenon through detailed contextual analysis of a limited number of events or conditions, and their relationships. Yin (1984) defines the case study research method as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used.

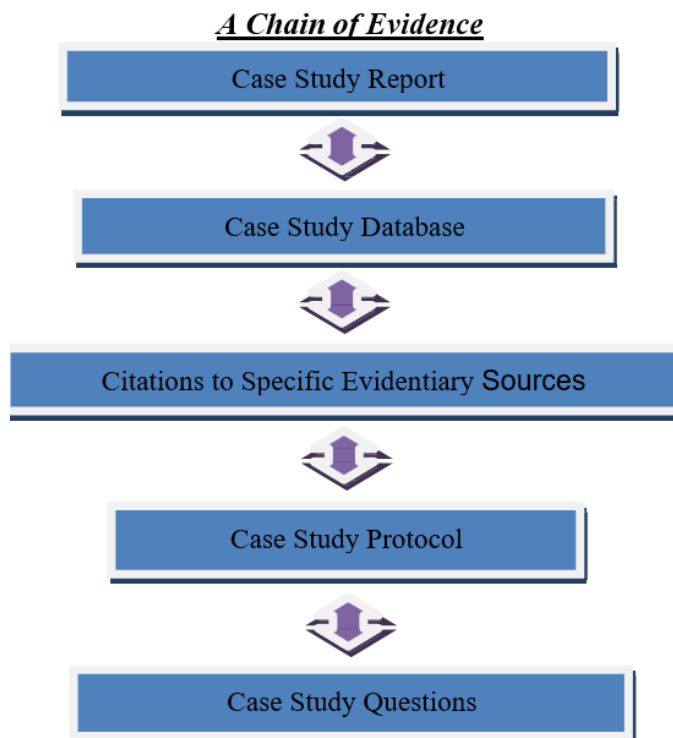Yin (1994) has mentioned that the case study design must have five components:
- ✓ The research question(s),
- ✓ The research propositions,
- ✓ Unit of analysis,
- ✓ How the data are linked to the propositions, and
- ✓ Criteria to interpret the findings

According to Stake (1994), types of case studies depend upon the purpose of the inquiry: an *instrumental case study* is used to provide insight into the issue; an *intrinsic case study* is conducted to get in-depth knowledge regarding the case; and the *collective case study* is the study of a number of cases in order to inquire into a particular phenomenon. Stake has also put emphasis that many other types of case studies based on their specific purpose, such as the teaching case study or the biography. Yin (1994) points out that case studies are the preferred strategy when "how" and "why" questions are posed.

According to Yin, a good case study report must have following components:

- ✓ The report itself should make sufficient citation to the relevant portions of the case study database,

- ✓ The database, upon inspection, should reveal the actual evidence and also indicate the circumstances under which the evidence was collected,

- ✓ It should be consistent that the specific procedures and questions contained in the case study protocol, to show that the data collection followed the procedures stipulated by the protocol,

- ✓ A reading of the protocol should indicate the link between the content of the protocol and the initial study questions.

Again, Yin has mentioned that one can move from one part of the case study process to another, with clear cross-referencing to methodological procedures and to the resulting evidence. This is the ultimate "chain of evidence" that is desired (see chart below).

A Chain of Evidence

Case Study Report

⬍

Case Study Database

⬍

Citations to Specific Evidentiary Sources

⬍

Case Study Protocol

⬍

Case Study Questions

(*f*) *Participatory Rural Appraisal Method*

Participatory Rural Appraisal (PRA) approaches have become increasingly popular among qualitative researchers in recent times. Such approaches are based on post-positivism and combine the strengths of "constructivist" paradigm and that of "critical realism". From the "constructivist" paradigm, PRA approaches seek to understand human experience as it is lived and felt by the participants, the context of such experience in an interdisciplinary framework. PRA as a methodology puts emphasis on the people's capabilities, field-based learning and innovations.

According to Kumar (2002), there are four pillars of PRA: methods and tools, process, sharing and attitude and behaviour. Again, he considers that attitude and behaviour in PRA is the most important pillars and it is central to PRA and more important than methods. The attitude and behaviour in PRA include:

- ✓ Self-awareness of one's behaviour
- ✓ Accountability to the poor
- ✓ Self-respect and respecting others
- ✓ Good Listeners
- ✓ Ability to 'handing over the stick'

PRA as a tool or a method, or a tool and method is very useful to collect primary data from the field in a very short period of time with the active participation of local people of the area. Initially to apply this method and tools, the researcher has to build up a sound rapport with the participants. Then she/he has to explain the purpose of the exercise to the participants. PRA is conducted mostly during the leisure time of the respondents.

A wide range of applications of PRA have evolved and still is it continuing. According to Chambers (1997), most application of PRA are in three areas, namely

a) topical investigation,

b) research, training, and

c) an empowering process of appraisal, analysis, planning, action, monitoring and evaluation.

Five sectors like, natural resources management; agriculture; people poverty and livelihood; health and nutrition; and urban attracted more PRA.

As per the need of the application of tools it is categorised under sectors, like, Space-related PRA methods, Time-related PRA methods and PRA relation methods. Here we would cite one example of a Venn Diagram under PRA relation methods.

Venn Diagram

Venn diagram is one of the commonly used methods in PRA to study institutional relationships and is sometimes also referred to as institutional diagram. It is, however, popularly known as *chapati diagram* as this method uses circles of various sizes to represent institutions or individuals. The bigger the circle, the more important is the institution or individual. The distance between circles represents, for example, the degree of influence or contact between institutions or individuals. Overlapping circles indicate interactions and the extent of overlap can indicate the level of interaction.

Applications

The Venn diagram method in PRA has been found very useful to study and understand local people's perceptions about local institutions, individuals, programmes, etc. The method provides valuable insights into and analyses of the power structure, the decision-making process, etc. The need to strengthen the community's institutions can also be ascertained. The relative importance of services and programmes has also been studied using the Venn diagram.

Process

The suggested steps in the process of doing a Venn diagram are as follows:

- ✓ Explain the purpose of the exercise to the participants.

- ✓ Ask them to list the various institutions, individuals, etc., as per the objectives of the exercise.

- ✓ Ask them to write and/or depict them on small cards. Visual depiction becomes necessary if there are non-literate participants.

- ✓ Ask the participants to place the cards on one of the variables of study, e.g., perceived importance of the institutions, in a descending order. Once the cards are arranged in an order, confirm the order. Encourage them to make changes, if they are interested.

- ✓ Ask them to assign paper circles of different sizes (cut and kept ready) to the institutions or individuals in such a way that the bigger the circle, the higher that institution or individual ranks on that variable. Paste on the circles the cards with names of

institutions or individuals. If you want, you can simply note down or depict the institutions or individuals on the circles.

- ✓ After placing all the cards, confirm the placement. Encourage them to make changes, if needed.

- ✓ In case, there are certain institutions/individuals who interact or work closely, they could be placed with an overlap. The degree of overlap indicates the degree of interaction.

- ✓ Ask them to discuss and explain why they placed the cards in such a manner. Note down the points of discussion and explanation.

- ✓ Copy the output onto a sheet of paper. Record the name of the village, participants, date, legends, what the size of the circle represents and what the distance represents.

- ✓ Triangulate the diagram and the major findings with others knowledgeable about the situation to ensure that your information is correct.

In order to facilitate easy making of this diagram, you should follow a step by step approach and need not explain the whole process to the participants at the outset. For instance, ask the participants first to list the institutions. Once the list has been made, ask them to put them in descending order based on each variable. Next, ask them to assign paper circles of different sizes and so on. Also ask them what they mean by the two variables. Make sure that the participants are clear on which dimension represents what variable. One simple way is to write it down legibly in bold letters and keep it in front when the exercise is on.

Materials Required

Paper circles are the most frequently used materials in Venn diagramming. It can also be drawn directly on the ground or on paper, but that does not allow the size or location of circles to be changed. Sometimes, after the circles are drawn, participants discuss the diagram and want to change the size or location. They hesitate to do so if the Venn diagram has been drawn, but if the circles are cut from paper, they find making modifications easy at any point in the process.

Time Required

Time required for a Venn diagram may vary considerably depending upon the details that are being represented. However, you should plan to spend 2-3 hours on the Venn diagram and the subsequent discussion.

Limitations of Venn Diagram

There are certain limitations of this method. Venn diagram generally becomes difficult and complex when the number of items increases. Relatively inexperienced facilitators find it difficult to explain the Venn diagram process to the participants. Another practical problem with Venn diagramming is that sometimes it can become sensitive. In the presence of some of the individuals or representatives of institutions that are being rated in the Venn diagram, the participants may play safe. Hence the output in such cases may not reflect the realities.

Steps for Collection of Primary Data

The different steps of collection of primary data are as follows:

(1) *Planning the Study:* The planning is an essential component to conduct any research study without which the data collected may not be found suitable. The following points should be considered at the planning stage:

(a) Objective of the study should be clearly mentioned.

(b) Sources of data, whether primary or secondary, should be identified.

(c) Type of study, whether census or sample survey.

(d) Definition of the unit of the study, whether individual or household.

(e) Degree of accuracy.

(2) *Collection of Data:* The collection of accurate data is the most important part in the whole investigation. The method of collection depends to a large extant on the nature, objective and scope of the study and the availability of time and money. We have already discussed several methods that can be applied singularly or in mixed form to collect field data.

(3) *Editing the data and its tabulation:* Soon after the collection of data, arrangements should be made to scrutinise them. If they are available in written form, those should be checked to limit inconsistency, errors and omission. After such check, the numerical data should be classified and tabulated, if required. Qualitative researchers also carefully read their diary or notebook to organise and classify their material under certain theme

(4) *Analysis of data and Interpretation of Result*: Data analysis is a method of abstracting significant facts from the large mass of data collected during the field work. If the given sets of data are numerical, then it involved determination of various statistical measures, the estimation of statistical constants and subsequent test of significance. The results of analysis are then interpreted and conclusions are drawn. Ethnographers also follow certain standard practice to organise their field data. Availability of computer aided qualitative programmes has made such analysis very easy these days (read module RMS 30 for details).

(5) *Preparation of Report:* After completion of the total process, it is necessary that a detail report is to be published. It should contain a detail description of all the stages of the survey. Charts and tables are also included in the report to represent/classify data rationally and consistently.