
UNIT 6 DATA COLLECTION SOURCES

Structure

- 6.0 Introduction
- 6.1 Learning Outcomes
- 6.2 Primary and Secondary Data
- 6.3 Sources of Secondary data
 - 6.3.1 Internet search
 - 6.3.2 Library search
- 6.4 Sources of Primary Data
- 6.5 How to Store and Save Your Data
- 6.6 Let Us Sum Up
- 6.7 References and Further Readings
- 6.8 Check Your Progress: Possible Answers

6.0 INTRODUCTION

Research in communications is a complex process. That is why this course is organised to assist you to proceed, in a systematic way, to undertake research in the subject. The first block of the course was about introducing you to research and about helping you to decide what kind of research you wish to take, how to define your problem, how to draw a sample, and what is meant by a review of literature. From Block 2 onwards, we shall take you through the entire process of research and explain how to find information, design tools based on quantitative and qualitative approaches, and to analyse data thus collected.

Media research is assuming increasing importance as information is an important part of our daily lives. All organisations, especially those in the media, need data. Data, which are then converted into information and knowledge, goes toward enhancing a given topic and helps in decision making. Data can consist of background information about a celebrity; data can also consist of voting patterns in a previous election. However, not everything published, online or offline can be relied upon to be called 'data'. In understanding data for research, it is important to distinguish data by the source from which it is derived and by the way in which it is gathered.

The most essential component of any research, including mass communications, is that it must be conducted according to sound principles; else, it will not yield accurate results and will be a mere wastage of time, efforts and resources. Such research conducted must use effective data collection sources to ensure that the research study is based on accurate and valid information.

The term 'data' as you may be aware is plural. 'Datum' is the term used for a single piece of data. Our focus is on data. **Data collection** is the process where the information is gathered and measured on variables of interest. It is important at all stages of a research process. Developing research questions and hypotheses and their subsequent tests need good data.

6.1 LEARNING OUTCOMES

After completing this unit, you should be able to:

- describe various primary sources of data collection;
- discuss the secondary sources of data collection;
- explain how to access and use various online databases;
- trace the connection between different forms of research and data collection; and
- store and save your data.

6.2 PRIMARY AND SECONDARY DATA

Research, as you are aware is based on data, not on opinions and hunches. Therefore it is extremely important that the data collection process is accurate and appropriate. Sometimes, even the approach to a research problem can depend on the preliminary information that is collected. What data we collect and how we collect are critical so that we can undertake the following activities:

- frame the research question in a suitable way
- collect relevant data
- avoid having data gaps
- save valuable time and resources
- make accurate, reliable, and sound analyses
- help future researchers navigate the data available
- guide future researchers in further investigating a topic.

As discussed in the previous units the Sources of Data can be of two types - Primary and Secondary. Let us understand them in greater detail.

Data are data which are collected fresh and or the first time. For example, although data may already exist in government databases, one may still want to go to a village and collect fresh information on the number of disabled persons and the types of disability they have. What is then collected is primary data, original in character. During a census every ten years, the government collects primary data; and every few years, government conducts the National Family Health Surveys. Such data collected from the field are primary and they are analysed before being presented to us in reports. When we read the reports, we are looking at secondary data and not at primary data.

There are several methods of collecting primary data. We may collect the data through surveys, through observations, through discussions and interviews with experts and with respondents in our selected research area. We may also collect primary data through photographs, audio, and video recordings that we take when in the field. The key characteristic of primary data is that it is originally collected.

Therefore, primary data are:

- Collected at the time when an event has occurred.
- Data are in original form and are not published anywhere else and have not been interpreted or translated.

- Data thus collected help to create original documents because the origin is the research conducted in the field.

Secondary data on the other hand, are data which are already available and have been collected and analysed by someone else. Such data may have already gone through statistical procedures and analysed for example, Census reports, and national readership surveys. In these cases, an agency has collected data from the field, tabulated and analysed it and has presented it in the form of tables and charts, and reports for others to use. For example, we may refer to the findings of a national readership survey by saying “According to data reported in the latest National Readership Survey, 2018” while reporting findings. Secondary data are also available and collected from a number of sources such as publications of central and state governments, annual reports, technical and trade journals, books, research reports. Some such data may have been published in journals, others may be unpublished reports.

The key factors to remember about secondary data are:

- Secondary data are not original.
- Secondary data have already gone through some analysis, so the use of secondary data means that we are examining already examined data and conclusions.
- Secondary data may have been taken from an original source, but have been repackaged for a set of objectives which may be different from our research.
- Secondary data may have been analysed differently.
- Use of secondary data is a cost effective method of data collection but such data must be reliable, valid and accurate. For example, Wikipedia is not considered as a reliable source of secondary data for academic research.

There are some advantages of using secondary data in research. The use of secondary data is cheaper, time saving, and is a quick way of gaining an understanding of the subject at hand. Sometimes, reanalysis of secondary data may yield different results from the original.

But there are also limitations to secondary data. Because such data come from already published reports, information may be outdated. For instance, much of the data in the Census Report 2011 is dated because the report gives information about the situation in 2011/2012, but cannot be counted upon to be accurate in 2018. This is also the case with many field based reports, because there is always a time lag between the collection of primary data, the analysis and the publication as a report in respected journals. Sometimes this time lag can even be three or four years long.

Requirements of Secondary Data

How does one then make the necessary choices about when, where, and how to use different data sources? We suggest some simple ways of deciding when to use secondary data, when to use primary data, and what are the yardsticks for evaluating secondary data.

Most research in mass communication uses a combination of both primary and secondary data. The extent to which a data source is used depends on the kind of research being undertaken. Then, the first issue that comes to mind is: what is

the kind of research we are undertaking. Let us examine this in terms of the classifications of research that were presented in Unit 2.

There are four requirements in selecting secondary sources of data.

- 1) **Availability:** Secondary data must be available. For instance to take the example of the physically challenged persons cited earlier in this unit, if details about the physically challenged in a village are not available in the census or other documents, then we would have no choice except to collect it afresh. We may also get, at most, a number of physically disabled persons, but may not know specifically who they are in a family.
- 2) **Relevance:** Data must be relevant to the problem under study. Do not use data sources which are not relevant.
- 3) **Accuracy:** Secondary data must be accurate and reliable. There must be clarity on how the data were collected, what methodology were used; what the margin of error was and whether the sources were dependable.
- 4) **Sufficiency:** There must be enough data available. If enough data are not available, primary data collection is a must.

Check Your Progress: 1

Note: 1) Use the space below for your answers.

2) Compare your answers with those given at the end of this Unit.

- 1) Define Primary and Secondary Data.

.....
.....
.....
.....

- 2) What are the important factors in choosing secondary data?

.....
.....
.....
.....

6.3 SOURCES OF SECONDARY DATA

Let us look in more detail about the sources of secondary and primary data especially in terms of the processes that are followed in carrying out research studies. Communication is an interdisciplinary subject therefore we often research topics which are not just about media; but also about many other disciplines. This is why we must cast our nets wide in our search for secondary data and must look at a range of sources for information.

When we are about to embark on a research project or study, many things are unclear. At this stage, we are still looking at a broad area and we do not know

what research on the given topic is available, who has done research before, and where is the information to be found that will guide us to proceed. What are the secondary data sources that you are likely to search for? Some of these are:

- Library resources
- Previous research reports/ Project reports
- Historical accounts
- Books and materials
- Official/non-official records
- State and district gazetteers, Census documents, Statistical reports
- Autobiographies, letters, diaries
- Documentary films, Programme transcripts
- Photographs, Maps
- Newspapers, Magazines
- Discussions with experts /People

This is the first step and it is here that secondary data and secondary data sources are most important.

6.3.1 Internet Search

In today's world, however, an Internet search is the first place to start with. This is because most publications are now indexed online, and hard copies of the indexes may or may not be available in a library. Hard copies are published year wise, while the online index allows you to search across several years in one place. Use a search engine like Google and use key words to search for the topic under study. The search engine will throw up vast amounts of material, some usable, some not usable.

For academic research, some Internet based secondary sources in the field of social sciences include Google Scholar, the Social Science Citation Index, The Arts and Humanities Citation Index, The Communication Initiative Network, ERIC, JSTOR, Science Direct, Infilibnet among others.

Google Scholar

When you first search Google for a topic, the search engine will first show you what articles are listed and emerge in Google Scholar. Google Scholar is web search engine which specifically focuses on academic and scholarly articles. With a single search, you can search across many sources: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites. Google Scholar will then provide you with the link telling you where the book or article exists, sometimes provide you a link through Google books (where you can read the book or article) and also show you how to access it. It is a good place to start your hunt for secondary sources. You will also find out if the full article is available free of cost, on payment, who the copyright holder is, and how the article is to be cited in bibliographic citation format needed in academic publications.

Social Science Citation Index

The Social Science Citation Index (<http://mjl.clarivate.com/cgi-bin/jrnlst/jloptions.cgi?PC=SS>). covers about three thousand journals across 50 disciplines in the social sciences. You can search the index by title, author, subject, and journal or ISSN number. The ISSN number is the international classification of journals. You can view a list of the all the journals included in the Index and search for materials in both current and older journals. If you do not have any preliminary information, search by subject, using key words.

The Arts and Humanities Citation Index

The Arts and Humanities Citation Index (http://mjl.clarivate.com/scope/scope_ahci) is similar to the Social Science Index and is available from the same online provider. This index covers more than 1700 arts and humanities journals and has both indexing and abstracting. You can search by topic or title or journal.

The Communication Initiative Network

The Communication Initiative Network (<http://www.comminet.com/global/content/overview-communication-initiative>) is a network of more than 100000 individuals and organisations working in the field of communication and development. You can search their database for research papers, case studies, knowledge summaries on important topics. The initiative provides free access to a vast repository of materials on various aspects of communication and development.

ERIC (Education Resources Information Centre)

ERIC is an online digital library of education research and information sponsored by the U.S. Department of Education. (<https://eric.ed.gov/>) ERIC provides access to 1.5 million bibliographic records (citations, abstracts, and other pertinent data) of journal articles and other education-related materials, with hundreds of new records added every week. The ERIC Collection, begun in 1966, contains records for a variety of publication types, including journal articles, books, research syntheses, conference papers, technical reports, dissertations, policy papers, and other education-related materials ERIC also has a connection of grey literature, i.e. materials produced by individuals and organisations outside the normal means of distribution such as journals and published books.

JSTOR

JSTOR (<https://about.jstor.org/mission-history/>) is a digital library for scholars, researchers, and students which covers more than 2500 journals from 57 countries. JSTOR provides free and low cost access to academics and scholars in the development countries through partnerships with institutions and philanthropic institutions. Currently, JSTOR has in its digital collection more than 10 million academic journal articles, 50000 books, and 2 million primary source documents. In 75 disciplines. After Google, this is a good place to begin the search for secondary data.

Science Direct

Science Direct (<https://www.sciencedirect.com/>) has a huge collection of Social Sciences and humanities journals and books, highlighting historical context,

current developments, theories, applications, trends and more. When you go to Science Direct, you may think that there are no materials in mass communication. However, what you must understand is that because mass communication is an interdisciplinary subject, secondary sources of data in other subjects like Psychology, Education, etc. are also important. Also, if you are looking for research on theme areas in development, such as health communication, there will be a wealth of articles available through this digital repository of articles and journals. Currently, more than 250,000 articles in Science Direct, are open access, made free to read, download and use keeping in line with the author's licensing.

Inflibnet

Inflibnet, an Inter-University Centre of the University Grants Commission is India's digital network of libraries and resources. It was set up to provide support to academic research and learning through cooperative facilities among universities and other academic institutions. Inflibnet has many activities, one of which is E-ShodhSindhu (<https://www.inflibnet.ac.in/ess/about.php>). ShodhSindhu provides current and archival access to more than 15000 core and peer reviewed journals and a number of bibliographic, citation and factual databases in different disciplines from a large number of publishers and aggregators. Students and scholars can access these databases through member institutions including centrally funded institutions, universities and colleges. Through its e resources, you can access journal articles and reports and databases. Through its activity IndCat, Inflibnet also provides access to an online catalogue of holdings (books, journals, theses and dissertations), in Indian University Libraries. This online catalogue is very helpful for students and researchers as it informs you in which library a particular book may be available. An interlibrary loan of materials is facilitated through Inflibnet.

The sources for secondary data given above are largely free databases. However, on the Internet search, you may find many articles freely available in the public domain and those that are copyrighted and not accessible unless you make a payment. In developing countries like India, the cost of paying US\$35 just to view an article can be very prohibitive. This is where access to a library becomes important.

6.3.2 Library Search

Before the age of the Internet, the library was the place where a researcher would go to collect data. The library is a storehouse of books, journals, indexes, reports, and documents, newspapers, magazines and even unpublished documents. Not only do libraries hold published materials, many libraries also hold collections of old manuscripts, scrolls, both in hard documents and within their microfilm and microfiche and e resources sections. Libraries are also able to access e journals and books through interlibrary loan facilities.

Unless you have a borrowing facility as a library member, most libraries, you will find, will allow you to read the books and journals there in the library itself. The first place to start in the library is the card catalogue. Many libraries have both online searches as well as drawers full of library cards organised in an alphabetical manner. Search the card catalogue first by the topic of your study. This is because, at this stage, you may still be unaware of the names of scholars

or titles of books on the topic under study. You can also search by author and/or title of the book. Books are arranged on shelves by the order in which they are numbered in the library catalogue. Usually it is the Dewey Decimal Classification (DDC) system or the Library of Congress (LCC) system which a library follows.

After your book search, proceed to the Reference section of the library. This is where reference materials, such as Encyclopaedia, Yearbooks, and even the hard copies of the Citation Indexes mentioned in the earlier section are housed. Once you find the journal article reference, go to the journals section of the library. This is where journals are kept and you may find the issue of the journal there. Newspaper articles are usually kept in the microfilm or microfiche collection of the library, because it is difficult for a library to stock print copies. If you cannot find what you want, ask for help, and the librarian will tell you whether it is available in the library or at the NASSDOC library or can be obtained on an interlibrary loan from some other library.

You may also like to explore the NASSDOC library of The Indian Council of Social Science Research. NASSDOC provide documentation and information support to researchers in social sciences and contains a huge collection of documents for use. Scholars working in autonomous research organisations, universities, government agencies, can use this facility to obtain data very easily.

Newspaper and Magazines

In media research studies, newspapers and magazines can play an important role in collection of secondary data. With the development of internet, e-papers are also getting popular. Thus, getting information on any topic is very easy. There are opinions, columns, news stories, editorials, features, reports and essays that can be referred to for research. However, it is advisable to use newspapers and magazines carefully and only for some background information, and to get a time reference for an event. Newspapers and magazines are not recognised as valid sources for academic research. Equally, articles in some open access journals (which are not peer reviewed), Wikipedia sources, and blogs of individuals are not recognised as valid sources for academic research.

Much of secondary sources of data are consulted at the formative stage of a research project, in fact, much before the researcher goes to the field. Secondary sources of research form the heart of the Review of Literature discussed in Unit 5. In turn, the review of literature forms the basis for developing research questions and hypotheses, and which serves later to link the current research to a body of theory on the subject. Being able to identify, sift the useful from the irrelevant and find knowledge gaps which need to be filled, needs that we understand how to find secondary data and how to use it. And finally, because our sources of secondary data have to be reliable, we need to know how to identify reliable secondary data.

Check Your Progress: 2

Note: 1) Use the space below for your answers.

2) Compare your answers with those given at the end of this Unit.

1) Name a few freely available Internet sources for secondary data.

.....

.....

.....

.....

2) What is JSTOR?

.....

.....

.....

.....

3) Name two salient features of Inflibnet which are useful for Indian academics and researchers?

.....

.....

.....

.....

6.4 SOURCES OF PRIMARY DATA

Primary Data are original research data that are collected by an organisation or an individual i.e. first-hand data that are collected. It is often assumed that, since the data are collected by the researcher him/herself, such data are more reliable. This is not necessarily so, unless rigorous and systematic steps are taken while collecting data. How to collect accurate, unbiased, and reliable and relevant primary data is discussed at length in various units of this course. At this stage, we will simply discuss what are the various ways in which primary data can be collected. Collecting primary data is often time consuming, and also more costly to carry out.

Primary data are also referred to as **raw** data. However, once such raw data are cleaned and processed by one team of researchers, the data become secondary data for another group of researchers.

What are the characteristics of primary data?

- It is data collected in its original form.
- Such data have not been analysed or published anywhere else. It may be fresh data collected in the field, relevant only in terms of a specific date, e.g. exit survey data during elections.
- Such data are always specific to the researcher's needs.

What are the sources of primary data? There are many kinds of primary data. These include original letters, historical documents, photographs, audio and video recordings, archives, surveys and observations can also be original data. We will discuss different kinds of primary data in mass communication terms of the different objectives of a given research effort.

If the objective of your research is to understand the role of the press in the Independence struggle in India, you will definitely look for books and articles

written about the topic. This is **historical research**. You may also go to original pages of newspapers, and original audio recordings. For example, in a pioneering research on the role of Gandhi and Mao as communicators, Dr. Kusum J. Singh (1980) used the original recorded messages of Gandhi and Mao during the period 1942-1944. She listened to the recordings, transcribed them and analysed them.

For her research, these original recordings were primary data.

You may study the original discussions of the Constituent Assembly to see what was discussed when India's fundamental right to speech was included in the Constitution. You may also study parliamentary proceedings, to understand what was discussed in the Parliament. You may study the diaries, personal letters, photographs, and other communications made by leaders and historical figures. In all cases you would use original materials. They include letters, life histories, memoirs, diaries, biographies and autobiographies etc. These help in getting information from the person himself and thus are reliable enough to be considered as primary data. If you are doing legal research, you would study case files, legal decisions and previous case histories and decisions. These documents give information regarding happenings in the past and are written by an individual to narrate incidents of his/her life.

Suppose, the objective of your research is to analyse the gender-based communication patterns in a village, especially the access and use of the Internet but you have little readily available data on which to base your hypothesis on. In this case, you would carry out **exploratory research** in the village, using various primary data tools to collect primary data. Without going into detail because these are discussed in subsequent units, we will now quickly look at various ways of collecting primary data.

Survey

As a first step, you may carry out a household door to door Survey (discussed in detail in unit 7) to determine your sampling frame accurately. This survey would give you information about all the members of each household, their gender, age, and educational level, etc. Such information would help you to select your sample. As such, this would be an exploratory survey. However, if the household survey does not yield enough information to enable you to make your research design; you may consider other methods such as observation or case study.

Observation

For a study on gender-based Internet access and usage cannot be complete until you understand a) if Internet Access is available in the village; and b) where WiFi hotspots are geographically located. For this you would need another tool - **observation**. You would have to walk around the village, see and locate the WiFi hotspots, the cellphone towers, and routers. You may take some photographs and recordings.

You may participate in the village activities in order to get a sense of dynamics. You would then be using a tool called participant observation, also part of ethnographic research. You would then record your observations of the participation in a diary or journal. The diary or journal notes become part of primary data. These tools are part of **Observation** which, in turn, is part of ethnographic research discussed at length in Unit 12 of this course.

Interviews

You may also go to locations where you see men and women gather and talk to them about Internet Access to gain some insights. By talking to them, you are carrying out **Interviews**, whether with individuals or women in groups. The process of collecting data through Interviews is discussed in detail in Unit 10.

Experiments

A significant amount of primary data can be collected through experiments. Discussed in detail in Unit 9, experimentation is the practice of gathering data by selecting matched groups of people, giving them different treatments or scenarios, controlling related factors in their environments, and checking for differences in their responses. By collecting all this preliminary information in an exploration, you are collecting **primary data**, which will then help you to frame your research questions, hypotheses, sample sizes and samples.

Primary data are not only collected at a preliminary stage. Primary data are also what you will collect during the field-work part of your research. In fact, all the data that are originally collected as part of your field-work are primary data, which will be tabulated, analysed, and reported as part of the research findings.

Both kinds of data, secondary and primary are extremely important and are necessary for research. Without secondary data, one cannot begin to sharpen the focus of research, to define a problem in addressable terms, frame questions and hypotheses, to know what information is necessary to be collected and what is not needed and to make judgments on reliability. When primary data are collected and analysed, findings have to be presented in terms of prior research and evidence collected from secondary sources. Without such a connection, the continuity of the research tradition gets affected.

Check Your Progress: 3

Note: 1) Use the space below for your answers.

2) Compare your answers with those given at the end of this Unit.

1) List three sources of primary data.

.....

.....

.....

.....

.....

2) Are original interviews primary data?

.....

.....

.....

.....

.....

- 3) I went to a village and saw that men gather in a tea shop while women gather at the temple. What kind of data have I collected?

.....
.....
.....
.....

6.5 HOW TO STORE AND SAVE YOUR DATA

At this stage, it is important to know how to store and save your secondary and primary data so that you do not have problems when a) citing sources, and b) doing your research, and c) writing your report.

All secondary and primary data must be stored for at least one year after completion of the research. These data will include your surveys and questionnaires, diaries, interview notes, field notes, observation notes, photographs, audio and videotapes, and anything else you may have collected as part of the research. There are simple reasons for this. First, you might want to consult a source again. Second, you might need to include a bibliography at the end of the research. Third, you need to get the citation correct. Fourth, you may want to write another article or research paper on a different aspect of the project for which the same data can be repackaged. Finally, if you are writing your thesis, the examiner may want information about your source. You should have it readily available for scrutiny when questioned. You may also have to add some more data to strengthen your study, when asked to do so by the examiner, or the agency paying for the research. If you do not take these steps at the beginning itself, you may land yourself in a nightmare later when writing the report, presenting it, or defending it at a viva voce.

In secondary sources, all Internet based sources provide details on how the book or article should be cited in a publication. Follow the rules and methods used in the American Psychological Association (APA) style sheet (<https://www.apastyle.org/>) on how to reference, a book, a journal article, and/or a web page. Save those details either in a folder or file on your computer, or on library cards, with each card for each citation.

All books and reports also provide you with full details on title, authors, publisher, and dates of publication. This information is usually provided in the inside cover page. Note down that information again, in a folder or file so that you can access them readily.

6.6 LET US SUM UP

In this unit, we discussed two types of data collection sources used in research-Primary and Secondary. It was explained that data collected by others, analysed and published in various books, journal articles, reports, etc. constitutes secondary data. Secondary data are very important for research since they help the researcher to set the tone of research, define objectives and questions more sharply, and most important, set the contexts and boundaries within which a research study can be undertaken.

Primary data are what is collected in field, either at a preliminary stage, to provide accurate information on which to build a sampling frame, or to set the field realities. Primary data also form the heart of research, because they are data on which, the findings and conclusions are based. The researcher's primary data becomes the secondary data for the subsequent or later researcher guiding the latter in the way that secondary data guided the study undertaken.

We further examined how we can collect data from Internet sources available for everyone to use. We also explored what are the data sources available in India and how you can use the library effectively and efficiently without much loss of time and effort. You should explore all these sources in doing your research.

Through the discussion we emphasised the importance of data and data sources in communication research because the failure to use proper data sources will affect the quality of your research, as well as the reliability and accuracy. Undertaking research is a time and effort consuming process and there are ways of making it simpler. These ways of sourcing data in a simple way were explained in this unit.

6.7 REFERENCES AND FURTHER READINGS

Chawla, Deepak and Sondhi, Neena (2011), *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing

Hamilton, John (2005), *Primary and Secondary Sources*. ABDO Publishing.

Singh, Kusum J (1978), *Gandhi and Mao as Communicators: A Comparative Study of Practice and Theory*. Unpublished doctoral dissertation. University of Pennsylvania 1978

6.8 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress 1

- 1) Primary Data are collected fresh and for the first time. Secondary Data have already been collected, computed, analysed and reported by some other person or agency.
- 2) Availability, relevance, accuracy and sufficiency are some factors in choosing secondary data.

Check your Progress 2

- 1) Google Scholar, Social Sciences Citation Index, Arts and Humanities Index.
- 2) JSTOR is a digital library for researchers and students with more than 2500 journals from 57 countries included.
- 3) Two important features of Infolibnet useful for researchers are E-ShodhSindhu and IndCat.

Check Your Progress 3

- 1) Surveys, observations, and interviews.
- 2) Yes, original interviews are primary data.
- 3) Primary data.