

MNM-031

Internet Research Methods

BLOCK 1	
Digital Research Methods	7
BLOCK 2	
Internet Research Methods-I	89
BLOCK 3	
Internet Research Methods-II	167
BLOCK 4	
Analysing and Interpreting Data	221

Disclaimer: Any materials and images adapted from web-based resources in this self-learning course material are being used for educational purposes only and not for commercial purposes.

PROGRAMME COORDINATOR: Prof. K.S. Arul Selvan

BLOCK 1			
Unit 1	Chhavi Garg, Assistant Professor, Department of Mass Communication and Media Studies, Central University of Punjab, Bathinda, Punjab	Unit 5	Dr. Anubhav Mathur, Associate Professor, Department of Media Studies, Bennett University, Greater Noida, UP
Unit 2 & 12	Dr. Sonali Srivastav, Assistant Professor National Institute of Fashion Technology Panchkula, Haryana	Unit 8	Dr. Poarkodi Natarajan, Assistant Professor, Department of Visual Communication, Loyola College, Chennai
Unit 9	Dr. Anupa Lahkar Post Doctoral fellow ICSSR New Delhi	Unit 10	Dr. Jayaseelan R Assistant Professor Department of Communication, PSG College of Arts and Science Coimbatore, Tamil Nadu
Unit 3, 4, 6, 7, 14, 16 & 17	Adopted from MAJMC of SOJNMS, IGNOU	Unit 11, 13 & 15	Adopted from E-PGPathshala, Ministry of Education, Govt of India
Unit 18	Prof. Nandini Lakshmikanth Former Director, Manipal Institute of Communication, Manipal University, Manipal		

PRODUCTION TEAM

Mr. Sanjay Agarwal
Asst. Registrar
MPDD, IGNOU, New Delhi

May, 2024

© Indira Gandhi National Open University, 2024

ISBN:

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on the Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by the Registrar, MPDD, IGNOU, New Delhi.

Laser Typeset by Tessa Media & Computers, C-206, A.F.E.-II, Okhla, New Delhi.

Contents

BLOCK 1	DIGITAL RESEARCH METHODS	7
UNIT 1	Internet Media Research	9
UNIT 2	Internet Research Traditions	25
UNIT 3	Identification of Research Problems	38
UNIT 4	Sources of Data	52
UNIT 5	Online Sampling Methods	73
BLOCK 2	INTERNET RESEARCH METHODS-I	89
UNIT 6	Interview Method	91
UNIT 7	Experimental Research Methods	105
UNIT 8	Visual Analysis	120
UNIT 9	Social Network Analysis	138
BLOCK 3	INTERNET RESEARCH METHODS-II	167
UNIT 10	Web Archiving	169
UNIT 11	Textual Analysis	189
UNIT 12	Online Ethnography	198
UNIT 13	Narrative Analysis	211
BLOCK 4	ANALYSING AND INTERPRETING DATA	221
UNIT 14	Fundamentals of Data Analysis	223
UNIT 15	Digital Data Analysis	238
UNIT 16	Data Interpretation	263
UNIT 17	Reporting Writing	285
UNIT 18	Ethics of Internet Research	301

COURSE INTRODUCTION

Information and Communication Technology advancements have revolutionised our daily routine communication activities and created a new academic discipline called Internet studies. This discipline is all about human communication and its allied activities in cyberspace. It is a stronger interaction between human beings and technological platforms. This unique combination requires a separate understanding to study on this platform, the same as the case of data collection tools. This is precisely what this course describes: the research methods and data collection tools and their techniques. This course has four blocks and 18 units, and it is the MNM031 Internet Research Methods of MA in Journalism and Digital Media programme. Let's briefly understand the whole structure of the course.

Block 1 Digital Research Methods would give a broader orientation on the basics of Internet research processes. In Unit 1, "Internet Mediated Research," we explore Internet research's evolution, significance, and challenges. Discover methodologies crucial in journalism while navigating ethical considerations. Moving to Unit 2, "Internet Research Traditions," uncover the multifaceted realm of Internet studies, exploring its scope, characteristics, and methodologies like digital ethnography and data analytics. In Unit 3, "Identification of Research Problems," learn the critical process of selecting and formulating research inquiries, emphasising literature reviews and crafting research questions. Unit 4, "Sources of Data," explores primary and secondary data gathering methods, including Internet and library searches. Finally, in Unit 5, "Online Sampling Methods," we explore probability and non-probability sampling techniques, exploring various methods and addressing ethical considerations in Internet-based survey sampling.

Block 2 of Internet Research Methods – I would clarify various data collection tools. In Unit 6 explains about the techniques of interview and online sampling methods, and the multifaceted realm of interview techniques. From informal to structured interviews, including telephonic and online surveys, we will navigate various formats while emphasising interviewing skills and ethical considerations. Unit 7 takes you into "Experiment Methods," exploring classic research designs and the intricate experimental process. We will discuss controlling extraneous variables and randomisation, exploring experimental design variations, and field experiments. In Unit 8, Visual Analysis, we will understand the complexities of perception and interpretation in images, exploring methodologies like Semiotics and Discourse Analysis and examining historical, contextual, and personal dimensions. In Unit 9, "Social Media Analysis," we explore the intricate web of digital interactions, exploring Social Network Analysis, Data Management, and Network Theories in Cultural Anthropology.

Block 3 Internet Research Methods II, continues the discussion from the previous Block. Yet another set of data collection tools are explained in this Block. In Unit 10, "Web Archiving: Understanding Internet Histories and Beyond," we uncover the significance of preserving digital footprints.

Discuss the origins, ethical considerations, and pivotal role of web archiving in Internet history research. Unit 11 introduces "Textual Analysis," where we navigate diverse perspectives, types, and presentation methods. Learn effective strategies for collecting and analysing texts while understanding the importance of interpretation. Unit 12 engages us in "Online Ethnography," examining qualitative and quantitative methods in the digital realm into ethical considerations and responsibilities essential for navigating this intricate field. Lastly, in Unit 13, "Narrative Analysis," we decode the layers of meaning within personal and collective stories. Explore methods for collection and analysis, unravelling the social dimensions of narratives.

Block 4 Analysing and Interpreting Data provides a better understanding of data analysis and interpretation. In Unit 14, "Fundamentals of Data Analysis," we explore the core concepts shaping modern research methodologies. From handling quantitative data to qualitative analysis techniques, we unravel the intricacies of organising, summarising, and testing hypotheses. Moving forward to Unit 15, "Digital Data Analysis," we will understand through data retrieval and harvesting techniques within Digital Media Research Methods. Learn to refine search results tailored to your research goals amidst the vast expanse of digital information. Unit 16 focuses on data interpretation, empowering you with statistical principles to interpret data confidently. From frequency distribution to measures of central tendency, grasp the tools needed for precise analysis. In Unit 17, we will explore the report writing essentials, guiding you through crafting comprehensive reports from captivating titles to presenting findings effectively. Lastly, Unit 18 addresses "Ethics in Internet Research," examining fundamental principles governing research ethics and the nuances of researching digital platforms.

Despite the changing technological features of the Internet, human communication remains the same. Modes have changed, but the core values of human interactions continue as before. The new tools and technologies of digital data collection will broaden our understanding of the deeper nuances of individual interactions through online spaces.

Block

1

DIGITAL RESEARCH METHODS

UNIT 1

Internet Mediated Research

UNIT 2

Internet Research Traditions

UNIT 3

Identification of Research Problems

UNIT 4

Sources of Data

UNIT 5

Online Sampling Methods



BLOCK 1 INTRODUCTION

In this first Block, you will learn the basic concepts of Internet research traditions and overall trends of Internet research in the recent past. With these understandings, you will learn about the fundamental process of research exercise—identifying a research problem. Once you have identified your problem, you must know the required data and how to collect it. All these areas are discussed in this Block.

Unit 1 Internet Mediated Research, explores Internet research's evolution, significance, and challenges alongside its interdisciplinary nature. This Unit gives an understanding of its pivotal role in journalism and explores methodologies, distinguishing qualitative from quantitative approaches while navigating ethical considerations. It also takes you to the global perspective and integrates technology seamlessly into your research endeavours.

Unit 2 Internet Research Traditions explores into the multifaceted realm of studying the Internet, covering its scope, characteristics, and prevalent issues. It explores the phases of Internet research, from technical challenges to its societal impacts and future advancements. It uncovers methodologies like online surveys, digital ethnography, and data analytics that are crucial for understanding and theorising the Internet's intricate dynamics.

Unit 3 Identification of Research Problems, we will learn the critical process of selecting and formulating research inquiries. We explore distinctions between social and research problems, emphasise the significance of literature reviews, and discuss factors like relevance, feasibility, and achievability. Additionally, we address crafting research questions, objectives, and hypotheses and defining terms of enquiry such as concepts, constructs, variables, and operational definitions.

Unit 4 Sources of Data, explores primary and secondary data, exploring Internet and library searches as secondary data sources. Additionally, we will uncover methods for gathering primary data and discuss essential practices for storing and saving your valuable data.

Unit 5 Online Sampling Methods takes you into probability and non-probability sampling techniques, explores the steps involved in online sampling, dissects various methods, analyses the characteristics of effective online surveys, and discusses sampling for interviews, focus group discussions and social networks. We'll also touch on mixed-mode surveys and the advantages of online data collection and address the ethical considerations and challenges in Internet-based survey sampling.

We hope you will better understand Internet research by learning about Internet research traditions and basic ideas for research activities.

UNIT 1 INTERNET MEDIATED RESEARCH

Structure

- 1.0 Introduction
- 1.1 Learning Outcomes
- 1.2 Overview of Internet Research
 - 1.2.1 Definition and Evolution of Internet Research
 - 1.2.2 Milestones in Internet Research
 - 1.2.3 Contemporary Significance
 - 1.2.4 Challenges and Opportunities
- 1.3 Interdisciplinary Nature of Internet Research
 - 1.3.1 Global Perspectives
 - 1.3.2 Integration of Technology and Research
- 1.4 Role of Internet Research in Journalism
- 1.5 Internet Research Methodologies
 - 1.5.1 Qualitative vs. Quantitative Approaches
 - 1.5.2 Choosing the Right Methodology
 - 1.5.3 Cross-Cultural Methodological Considerations
- 1.6 Ethical Guidelines in Methodology
- 1.7 Let Us Sum Up
- 1.8 Keywords
- 1.9 Further Readings
- 1.10 Check Your Progress: Possible Answer

1.0 INTRODUCTION

The Internet has become an integral part of human lives. It is used to remain in touch with friends and family through social networking sites and plan one's health through health-specific Internet applications. This widespread integration of the Internet is still in process and has primarily started few decades back, though the Internet was first used in the 1990s. Moreover, this Internet usage is not limited to any particular age group or socio-economic group; almost everyone uses it for varied purposes.

Being online has become a necessity in the present as most information and services are available online. This online interaction of numerous users has given rise to online communities and an online world, which is being termed a 'virtual world'. Since these online communities and virtual worlds have a direct effect on human lives, research in this area has also become an essential part of communication and media research.

Internet research is emerging as a new area of research for academicians and researchers in the field of media and communication. All aspects related to the Internet, whether usage patterns, the kind of content generated online, or the different types of users themselves, have become subjects for Internet

research. This unit aims to provide an overview of Internet research, its evolution, and the present trends in this area of research. It also attempts to underline methodological debates and ethical concerns discussed in Internet research.

1.1 LEARNING OUTCOMES

After completing this unit, you should be able to:

- Understand the meaning and evolution of Internet research;
- Analyse research trends in Internet research;
- Comprehend the nature of Internet research; and
- Gain an insight into methodological debates and ethical concerns in Internet research.

1.2 OVERVIEW OF INTERNET RESEARCH

Internet research is an emerging area in communication and media research at present. Under Internet research, researchers are exploring all aspects of Internet usage, their users, and the audio-video content generated by users. Even the political, economic, and social issues that affect Internet usage and emerge from Internet usage are being studied under internet research as these issues play an important role in overall Internet usage. For instance, under social issues, it is being studied which group, according to age, gender or race, is accessing the Internet, whereas, under political issues, the actions or policies of the government in different countries are being analysed to understand the repercussions of these actions or policies on the usage pattern of users in a specific country. The following sections will discuss the evolution of Internet research and its significance in detail:-

1.2.1 Definition and Evolution of Internet Research

As the name suggests, Internet research deals with understanding the medium of the Internet, encompassing all its aspects. Since its development in the 1990s among the public, the Internet as a medium has been evolving, and hence, research on the Internet is also evolving with the medium. One of the initial fundamental questions of concern before researchers was understanding the nature of the Internet as a medium, whether it was an interpersonal medium or a mass medium. Usage of the Internet has shown that the answer is not either, but both, as the Internet can be used as an interpersonal and mass medium. Other main aspects of the Internet are its varied usage, different types of users, the content generated on it and the social, political and economic issues related to it.

Regarding research approaches in Internet research, Hewson, Vogel, and Laurent (2016) mention that quantitative approaches dominated the early studies in Internet research, while qualitative approaches gained momentum after some time in the field. A mixed-methods approach is also being followed in some research studies to follow breadth and depth in the area.

One of the points that needs to be noted is that in the beginning, the Internet was seen just as a tool for data collection rather than as a site for research. Initially, the Internet was considered just one media used to collect data through online questionnaires or interview schedules. This kind of research is also termed Internet-mediated research. Not only for primary data collection, but news reports and other documents, which were initially offline, were also searched on the Internet as an extension of newspapers, magazines, and other media, but no new data was looked at on the Internet. The reliability of the data collected through the Internet, primarily through emails, was also questioned as, in the beginning, researchers did not rely on online mode (though concerns remain even now). It was only in the last decade that the Internet is seen as having much more to offer than just being a tool for data collection. With the increasing use of the Internet, it has emerged as a critical area of research in the field of media and communication. Hence, now all its aspects, from its use to users to its effects, are being studied under Internet research.

1.2.2 Milestones in Internet Research

As mentioned earlier, Internet research is emerging as an area of research with the development and evolution of the Internet. Having publicly announced this for the first time in the 1990s, research started after this period. Further, as the medium evolved, researchers also got many aspects to research as many Internet-based applications emerged with specific target audiences. The following table shows the milestones in Internet research:-

	Developments in Internet technologies	Developments in online social research methods
1960s	Development of early forms of computer-mediated communication. 1962: Licklider sets out a vision for a 'Galactic' or 'Global' 'Information Network'.	
1970s	1973: The first mobile phone was demonstrated (Dr Martin Cooper of Motorola). 1979: First commercially automated cellular network (Tokyo, Japan).	
1980s	1983: first commercial handheld (not linked to a car) mobile phone (the DynaTAC 8000x). 1989: Demonstration of the World Wide Web by Tim Berners-Lee.	1986: Kiesler and Sproull undertake an online survey (Kiesler & Sproull, 1986). The invention of the first computer packages for the analysis of qualitative data.

<p>1990– 1994</p>	<p>1990: public release of the World Wide Web. 1990: The first search tool for the web (Archie) was created. 1993: The first web crawler (Wanderer) was created. 1993: First graphical browser (Mosaic). 1994: Netscape browser launched. 1994: Development of the first popular search engines (Alta Vista, Lycos, Excite and Yahoo).</p>	<p>1993: Rheingold uses the term ‘cyberspace’. 1994: Foster conducts online asynchronous interviews using email. 1994: First methodological discussion of online interviewing (Brotherson, 1994). 1995: <i>Journal of Computer-Mediated Communication</i> began publication.</p>
<p>1995– 1999</p>	<p>1995: Internet Explorer launched. 1995: The first public video conference took place. 1996: Instant messaging services launched. 1997: Google released the 1997 first weblog (blog), which is attributed to Jorn Barger's Robot Wisdom website. 1997: SixDegrees.com is launched. Often seen as the first social networking site.</p>	<p>1995: Correll writes about ‘Internet ethnographies’. 1995: First online web experiment conducted. 1995: First comprehensive list of online psychological experiments published on the web. 1996: The explosion of debate around online research ethics with papers published by Allen, Boehlefeld, King, Reid and Thomas (Allen 1996; Boehlefeld 1996; King 1996; Reid 1996; Thomas 1996). 1997: Krantz et al. published an online psychology experiment in an academic journal (Krantz et al., 1997). 1997: The World Wide Web and Contemporary Cultural Theory conference occurred at Drake University.</p>
<p>2000– 2004</p>	<p>2000: 400 million people across the globe use the Internet. 2001: First commercial launch of 3G (Third Generation) mobile phones. 2001: The first Access Grid was developed at the University of Manchester. 2003: Myspace launched. 2004: Development of Voice Over Internet Protocol (VOIP) telephone service. 2004: Mozilla Firefox web browser released (the 2nd most popular current browser after Internet Explorer). 2004: O’Reilly uses the term ‘web 2.0’. 2004: Facebook launched.</p>	<p>2000: Launch of the Association of Internet Researchers. 2000: Publication of a number of key texts in the field, such as <i>Mail and Internet Surveys</i>. The <i>Tailored Design Method</i> (Dillman, 2000); <i>Virtual Ethnography</i> (Hine, 2000); <i>Internet communication and qualitative research: A handbook for researching online</i> (Mann & Stewart, 2000). 2001: <i>The Internet Research Handbook</i> published (O’Dochartaigh, 2001). 2002: <i>Publication of Online Social Sciences</i> (Batinic et al., 2002); <i>Internet Research Methods</i>, Hewson, Yule, Laurent and Vogel (2003); <i>Standards for Internet-Based Experimenting</i> (Reips, 2002a). 2002: <i>Publication of the Association of Internet Researchers guidelines on</i></p>

		online research ethics (Ess, 2002).
2005– 2009	2006: Twitter launched. 2007: iPhone launched. 2008: Google Chrome browser launched.	2005: Virtual Methods (Hine, 2005) 2006: International Journal of Internet Science begins publication. 2008: The Handbook of Online Research Methods (Fielding et al., 2008).
2009– 2011	2010: iPad launched. 2011: The number of Internet users is estimated at 2 billion worldwide.	2009: Publication of Netnography: Doing Ethnographic Research Online (Kozinets, 2009) and Online Research Essentials (Russell & Purcell, 2009).

Table 1:- Milestones in Internet research (Source: Hooley, T., Marriott, J. & Wallens, J. (2012). What is online research?: using the Internet for social science research. Bloomsbury Academic.)

Recent developments in the period of 2011-2024:

Period	Developments in Internet technologies	Developments in online social research methods
2011- 2014	2012: Introduction of Google Glass wearable technology.	2011: Introduction of new online research methodologies to adapt to the changing landscape of social media platforms.
	2012: Introduction of 4G LTE (Long-Term Evolution) mobile networks.	2012: Integration of social media data into traditional research methodologies for a more comprehensive understanding of online behaviour.
	2013: Introduction of Snapchat, a multimedia messaging app.	2013: Emergence of advanced data analytics tools for processing large volumes of social media data, enabling more in-depth analysis and insights.
	2014: Introduction of Amazon Echo, a smart speaker with voice assistant technology.	2014: Implementation of mixed-methods approaches combining online surveys, interviews, and social media analysis for a holistic research perspective.
2015- 2019	2015: Introduction of 5G networks, promising faster internet speeds and lower latency.	2015: Expansion of qualitative research methodologies to include virtual ethnography and netnography, leveraging the immersive nature of online environments.
	2016: Introduction of Pokémon Go, a mobile augmented reality game.	2016: Increased focus on ethical considerations in online social research, with the development of guidelines and frameworks to ensure participant privacy and data security.
	2017: Introduction of Bitcoin and blockchain technology for decentralised transactions.	2017: Integration of machine learning and natural language processing techniques in online social research for automated sentiment analysis and topic modelling.

	2018: Introduction of GDPR (General Data Protection Regulation) for data privacy.	2018: Adoption of longitudinal studies to track changes in online behaviour and attitudes over time, providing insights into evolving social dynamics.
	2019: Introduction of 5G-enabled smartphones with enhanced connectivity.	2019: Collaboration between academia and industry to develop innovative research methods utilising social media APIs and big data analytics platforms.
2020-2024	2020: Accelerated digital transformation due to the COVID-19 pandemic.	2020: Rapid adoption of remote research methodologies such as online focus groups and virtual interviews to comply with social distancing measures.
	2021: Introduction of AI-powered voice assistants for smart home devices.	2021: Advancements in mixed-methods research designs combining qualitative and quantitative approaches to capture nuanced insights from online interactions.
	2022: Introduction of 6G technology in select regions, promising even faster speeds.	2022: Exploration of novel research avenues such as digital ethnography and participatory action research in online communities to engage with diverse perspectives.
	2023: Expansion of the Internet of Things (IoT) ecosystem with interconnected devices.	2023: Implementation of cross-disciplinary collaborations integrating social sciences, computer science, and data analytics to tackle complex research questions in online environments.
	2024: Emergence of decentralised social media platforms based on blockchain technology.	2024: Emphasis on transparency and reproducibility in online social research, with initiatives promoting open data sharing and methodological documentation for scholarly rigour and accountability.

1.2.3 Contemporary Significance

The Internet has become such an integral part of human lives that the future of society can be delineated from it. Almost all everyday activities of its users involve the use of the Internet. Hence, the significance of Internet research is bound to only increase.

After the announcement of the Internet public in the early 1990s, the generation born after and growing with it can not imagine living without the Internet as it has become such an intrinsic part of their lives. Therefore, if we want to understand the development of human society in the present times, we need to understand Internet usage among its users. The generation that came after the Internet and the generation that may be termed the elderly group have also started using the Internet to a large extent as they need to interact with their children through the Internet. Even the issues and concerns that used to exist in the larger communities have also been transferred into the online sphere. We may take the example of the digital divide, where some

people can access the Internet while others cannot, owing to social and economic factors. It may be seen as an extension of the depravity of one more resource among some people while others are using and enjoying the resources. Similarly, we may see that bullying, which used to happen in the real world, has gone online, with 'cyber-bullying' emerging as a concern in the virtual world. Hence, Internet research is the need of the hour as the Internet has almost become one of the most important ways to remain connected with immediate family members, friends, and the outside world.

1.2.4 Challenges and Opportunities

Since Internet research is still a comparatively new field of research in communication and media, some challenges are bound to exist for the researchers. One of the most critical challenges is the representation of diversity on the Internet. The reach and access to the Internet have increased very fast in the last decade, but the spread of Internet access has not been uniform. Most of the countries in the eastern developing world are still struggling to provide Internet access to all in their countries, whereas the western world of developed countries are dominating in access to digital technologies. Not only in the case of countries, the Internet is not accessible even now across various categories, including gender, class, caste or race. This digital divide, the gap between the 'haves' and 'have-nots', has made it challenging to have representational samples through the Internet or to take up the issues being discussed on public online forums as issues of the general people. Hence, researchers in this field need to consider the socio-economic profile of Internet users and decide who is included in the studies or the samples collected through the Internet. Kaufmann and Tzanetakis (2020) say that it is not only the researcher who needs to take care of fair representation or social barriers of respondents but also the environment in which researchers and respondents create, and the online tools that are used all contribute to making Internet research a credible one.

Another essential aspect emerging as a challenging issue is the reliability of the data collected online. In Internet research, data collection is mostly done online (not to say that data collection does not happen offline in Internet research). Hence, the reliability of data collected, primarily through asynchronous mode, is raising a concern. In asynchronous mode, a questionnaire or interview schedule is generally shared with the respondent on one's email account, and the respondent would usually respond to it whenever one is free. However, since the researcher and respondent do not come face-to-face like in synchronous mediated communication, how much the data can be relied on might be questioned as somebody in place of the original respondent has filled out the questionnaire or under which circumstances the respondent has filled out the questionnaire, the researcher does not get to know. Hence, the researcher needs to devise a strategy to check the reliability of data collected online.

Internet research has opened up many opportunities, too, as it has connected the whole world. Researchers conducting Internet research and through the Internet can reach respondents sitting in far-off areas. For instance, a researcher sitting in India is no longer limited to doing research on topics

related to India or people living in India; one can explore all the topics which are of one's interest without caring to think much about reaching the respondents as the respondents who are living in far-off geographical areas can be reached through online platforms. The usage of the Internet has not only blurred geographical boundaries but also has enabled reach among people with certain specific socio-psychological variables. These people can be reached through social media platforms as people with specific common interests or profiles have their online communities. Reaching respondents online also rationalises the use of resources for researchers as they usually have a shortage of resources in terms of both money and time. Along similar lines, even archives are also available online to access different documents for research related to secondary data. Even for literature reviews, most journals are accessible online. E-books, too, have come for researchers to read without needing to reach out offline every time. Similarly, respondents sometimes do not want to open up about specific issues in face-to-face interactions, but they feel comfortable with mediated communication. Hence, it becomes easy to reach such a population through the Internet.

Most of the time, data in Internet research is collected online using different software. This online software enables the transcription of data independently rather than spending money or time again after the data collection. For instance, when an interview with a respondent is conducted online, either in synchronous or asynchronous mode, it would be recorded independently (the respondent needs to be informed about it under ethical guidelines). Even many online tools enable the calculations/analysis of data with just a few commands on the software. Hence, processing large data sets in Internet research becomes more accessible and quicker. Therefore, the Internet has opened up a plethora of opportunities for researchers.

Check Your Progress: 1

Notes: 1) Use the space below for your answer.

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

1. What are the challenges being faced by researchers in Internet research?

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

2. Discuss some points that may be considered as opportunities in Internet research.

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

Activity: 1

1. Read any one research article related to Internet Research and analyse it.

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

3. Try to identify which aspect of the Internet is studied in the article.

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

1.3 INTERDISCIPLINARY NATURE OF INTERNET RESEARCH

Internet research is truly interdisciplinary. Not only can fields of social sciences like psychology, sociology, and economics be combined to understand Internet research, but technology-oriented fields like computer science and computational sciences can also be used to comprehend the field of Internet research.

1.3.1 Global Perspectives

The Internet is a medium which makes the world a real global one, as almost everyone seems to be connected to others through it. Since the Internet, among other digital technologies, emerged in the developed world, perspectives and research studies have also been brought in from Western countries. However, as the Internet is becoming more accessible in developing countries, research studies are also being conducted in these countries as users are becoming more frequent, and their uses vary in nature and purpose.

Listing different areas which are being explored under the interdisciplinary nature of Internet research, Kim & Weaver (2002) write about how individuals and organisations are using the Internet, how the usage of the Internet affects traditional and other media, how the Internet is being used in politics, education and marketing, legal and economic issues and advancements in technology. Further, Kim & Weaver (2002) note that initial research studies in Internet research focused on the medium itself and historical and philosophical issues of development of the medium, whereas more recent studies attempt to understand the use of the Internet and its effects. Observing that research journals play an important role in bringing researchers related to Internet research together at common platforms.

1.3.2 Integration of Technology and Research

The Internet is essentially a technological medium that has connected millions of people quickly, unlike any other technology of earlier times. In the field of media and communication, technological advancements have always remained important as these developments decide the content of the medium to a large extent. For instance, printing newspapers has undergone many changes over time to make printing much easier and quicker. This printing technology has affected the content of newspapers too, as with advanced technology, more updates till late at night could be accommodated with ease, and more colourful pages could be made, resulting in more photographs from earlier times. Similarly, technological advancements in the Internet have provided innumerable opportunities for users to interact with each other; they can access information and actively create content on the Internet. The Internet is emerging as a technological medium in which there is no standardised content like newspapers or television, but there is such a diversity in the content available on the Internet that various aspects can be researched. From the time spent on the Internet by users to the kind of websites they browse, every move of users on the Internet has become a potential topic of interest and research for researchers.

Along with a range of topics in Internet research, research methods have evolved with the coming of the Internet. Online ethnography can be seen as one example of the integration of technology and research. Traditional ethnography initiated by anthropologists requires researchers to immerse themselves in the field by living in the community for a certain time. In online ethnography, a researcher immerses themselves in online communities by observing the interaction of members in these online communities and participating in the interaction themselves.

1.4 ROLE OF INTERNET RESEARCH IN JOURNALISM

Journalism has remained one of the oldest fields of media and communication, though the same cannot be said about research in the field. With the coming of the Internet, journalism as a field is seeing many changes as most newspapers have started their e-papers and web portals. Further, influential public figures, including political leaders, have posted their statements online. Hence, journalists have also begun considering verified social media accounts as one of their sources. Many aspects of the Internet have started affecting the field of journalism. Research studies on usage patterns of online media by journalists, changes in coverage of newspapers due to emerging social media, and digital skills of journalists are some of the topics on which research is being conducted in the broad area of Internet research in journalism.

Citizen journalism has become very common, as with easy access to digital technologies, citizens can post their concerns and issues directly on different online platforms. Although citizens post their content online with quick and easy access to digital technologies, almost everyone's concerns regarding misinformation and disinformation have also emerged. Misinformation is

false information that is spread, whereas disinformation is false information that is spread or forwarded with intent. Hence, the spread of this kind of information has also been studied many times. People spreading this kind of information do it in the garb of journalism, calling in twin challenges as they raise the trust and credibility of journalism among readers or audiences, and another is that this kind of information may harm the society by spreading false information.

1.5 INTERNET RESEARCH METHODOLOGIES

Deciding on research methodology is one of the most critical steps in any research process. In that case, if the researcher does not use the appropriate research methodology, the whole study may be a failure, as the aim of the study would not be fulfilled without choosing the appropriate research methodology. Hence, the following sections discuss the different research approaches and some strategies for choosing the most appropriate approach for your study:-

1.5.1 Qualitative vs. Quantitative Approaches

A quantitative approach to research in social sciences is taken from physical sciences and aims to measure the concepts under study in numbers. Under this approach, the most used research method is a survey, in which a questionnaire is generally used as a research tool to collect data. Initial studies in Internet research followed quantitative approaches as many researchers from the field of technology-sciences took up research studies on the Internet. Experiments are also conducted using this approach to study different variables.

Under qualitative approaches, generally, in-depth interviews are conducted to collect data. In-depth interviews can be conducted through emails or in real-time through audio-video applications. This difference between the two can be termed synchronous and asynchronous communication. In synchronous, the researcher and respondent would be present on an online platform simultaneously, while in asynchronous, the respondent can send one's response easily through email. Both types of communication have some advantages and disadvantages. For instance, conducting in-depth interviews in synchronous mode would have no concerns regarding the reliability of data as the respondent would be present with the researcher, though, at the same time, we need to consider that sometimes respondents in synchronous communication might not be able to open up on some sensitive issues as they would do in asynchronous communication owing to having more time and anonymity to express themselves. However, there might be specific concerns regarding data reliability in asynchronous communication as the researcher cannot ascertain under what circumstances the respondent has sent the response.

1.5.2 Choosing the Right Methodology

The process of research is a long process which starts with identifying a broader research problem, reading literature related to it, identifying the research gap, making specific research objectives or research questions to fill

the identified research gap, then deciding on the appropriate research methodology, collecting data as per the decided research methodology and finally analysing the data and presenting the results and coming to a conclusion. Hence, the right way to choose the most appropriate methodology is to consider the study's objectives. Depending on the research's purpose and the study's objectives, an appropriate methodology needs to be chosen. For instance, if your research study aims to encapsulate the width (broader in perspective), then quantitative research methods should be deployed, while if the purpose of your research study is to gain depth on the topic, then qualitative research methods like in-depth interviews should be used. However, in the present times, most studies are taking up a mixed-methods approach to present a balanced perspective on the chosen research topics.

In the mixed-methods approach, both quantitative and qualitative research methods are used. These methods may be used sequentially or concurrently. In the sequential type of research, one method would be followed by another, like the quantitative research method of the survey, followed by the qualitative method of in-depth interviews, or vice versa. In concurrent research, both methods are deployed simultaneously, such as asking a respondent to fill out the questionnaire (with close-ended questions) and then asking some open-ended questions in an interview. However, the primary way to choose the correct research method is to keep the research aim in mind and choose the most appropriate methodology to fulfil the aim of the study.

1.5.3 Cross-Cultural Methodological Considerations

Owing to its nature, the Internet has allowed researchers to conduct research across cultures. As discussed in opportunities of Internet research, it may be noted that as the Internet has removed geographical barriers in reaching different respondents, researchers need not limit themselves to their immediate or neighbouring areas. They may explore their research interests across different countries and cultures. However, research across cultures needs to keep specific methodological considerations in mind. Since every culture has its code of conduct and norms, it is always advisable for the researcher to understand the cultural nuances in detail before interacting with the respondents, even if it is through the Internet. At the start of the study, a simple gesture of greetings might make or mar the data collection process as this simple gesture establishes rapport with the respondent and makes them comfortable to share their responses.

Similarly, gender norms also play a vital role in research. As an outsider to a culture, the researcher should know how to greet a person, especially of the opposite gender, and even culturally appropriate dress sometimes helps the respondent to open up more quickly. Language also plays a role as communicating in one's tongue becomes comfortable for the respondent and the researcher. Also, it becomes easier to understand non-verbal communication. Hence, the researcher should always attempt to get familiar with the culture of the community before conducting the study among them.

Activity:2

1. Choose any one research article on Internet research and analyse its research methodology.

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

2. Identify which type of methodology is used in the article.

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

1.6 ETHICAL GUIDELINES IN METHODOLOGIES

(Refer to Unit 18 for the detailed discussion on ethics in online research)

Like any kind of research, ethical guidelines are also an important part of Internet research. Discussion on ethical guidelines in Internet research becomes even more relevant as guidelines are still being debated since the Internet is a comparatively new medium and its usage is evolving.

One of the most important aspects of ethics in Internet research remains obtaining informed consent from respondents. In any kind of survey interview or even experimental study, respondents need to be informed before the start of the study about the purpose of the study and give their consent for participation in the study. Respondents also need to be told that they may leave the study at any time if they want to leave during data collection. This informed consent becomes comparatively easier to take in face-to-face communication as sometimes, in mediated communication, especially in asynchronous communication, respondents might not fully understand the purpose of the study. In the case of content analysis, where the respondents have decided to share the content of their private social media accounts, the researcher needs to obtain informed consent beforehand and discuss the purpose of the study in detail with the respondents.

Another important aspect is maintaining the anonymity of the respondents and the confidentiality of data collected from the respondents. In most of the studies, we ensure the respondents that their identity and the data collected from them would be used only for academic purposes and that both their identity and the data collected would be kept confidential, especially in areas of a sensitive nature. In Internet research, maintaining this confidentiality might be a bit of a concern as data is stored digitally, and there might be apprehensions about their leak or access by somebody other than the researcher.

Analysing content available on online public forums has emerged as an essential area of Internet research. However, some academicians are discussing ethical guidelines for using such content as though the content is available on open online public forums. However, still, since the medium of the Internet is new, many of the users are not aware of its nature. They might have expressed themselves without knowing that the content is available in the public domain and can be used for research purposes. Here, the dilemma can be solved by not revealing the content creator's identity, particularly in sensitive matters.

Check Your Progress: 2

Notes: 1) Use the space below for your answer.

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

1. What ethical guidelines do you keep in mind when doing Internet research?

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

1.7 LET US SUM UP

Internet research has emerged as an important area of research in the field of media and communication. With the increasing reach and accessibility of the Internet among different populations, it is becoming even more relevant in the present times. In every aspect, from the time spent on the Internet by users to the websites being visited by them to the kind of content created by them, every move of users on the Internet has become a potential research topic. As the use of the Internet is still evolving, the area of Internet research is too developing. Questions regarding choosing the most appropriate research methodology or ethical considerations are still being debated. However, the field of Internet research is bound to become even more interesting as both users of the Internet and researchers in Internet research are exploring the medium and innumerable possibilities.

1.8 KEYWORDS

Online Ethnography: It is a research method in which the researcher immerses himself in online communities by observing the interaction of members in these online communities and participating in the interaction oneself.

Synchronous mode: The mode of data collection in which both the researcher and respondent are present online in real time.

Asynchronous mode: This is the mode of data collection in which the researcher and respondent are not present online in real-time.

1.9 FURTHER READINGS

1. Bryman, A. (2012). *Social research methods*. Fourth edition. Oxford University Press.
2. Fricker Jr.,R..D. and Rand,M.S.(2002). Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods*, 14(4), 347–367.
3. Hewson, C., Vogel, C. & Laurent, D. (2016). *Internet research methods*. Second edition. Sage.
4. Hooley, T., Marriott, J. & Wallens, J. (2012). What is online research?: Using the Internet for social science research. Bloomsbury Academic.
5. Kaufmann, M. and Tzanetakis,M. (2020). Doing Internet research with hard-to-reach communities: Methodological reflections on gaining meaningful access. *Qualitative Research*, 20(6), 927–944.
6. Kim, S.T. & Weaver, D. (2002). Communication research about the Internet: a Thematic meta-analysis. *New media & society*, 4(4), 518–538.
7. Tomasello,T.K. (2001). The status of Internet-based research in five leading communication journals, 1994-1999. *J&MC Quarterly*, 78 (4), 659-674.

1.10 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Most of the countries in the eastern developing world are still struggling to provide Internet access to all in their countries, whereas the western world of developed countries are dominating in access to digital technologies. Not only in the case of countries, but the Internet is also not accessible even now across various categories, including gender, class, caste or race. This digital divide, the gap between the 'haves' and 'have-nots, has made it challenging to have representational samples through the Internet or to take up the issues being discussed on public online forums as issues of the general people. Hence, researchers in this field need to consider the socio-economic profile of Internet users and decide who is included in the studies or the samples collected through the Internet.
2. Researchers conducting Internet research and through the Internet can reach respondents sitting in far-off areas. The usage of the Internet has not only blurred geographical boundaries but also has enabled reach among people with certain specific socio-psychological variables. These people can be reached through social media platforms as people with specific common interests or profiles have their online communities. Along similar lines, even archives are also available online to access different documents for research related to secondary data. Even for literature reviews, most journals are accessible online. E-books, too, have

come for researchers to read without needing to reach out offline every time. Similarly, respondents sometimes do not want to open up about specific issues in face-to-face interactions, but they feel comfortable with mediated communication.

Check Your Progress: 2

1. Ensuring data reliability and ascertaining diversity in the sample are being seen as challenges in Internet research.
2. Convenience in reaching respondents, the requirement of fewer resources in terms of money and time, and quicker or automatic data transcription with the help of software are some of the opportunities that researchers can explore in Internet research.
3. It is a research method in which the researcher immerses oneself in online communities by observing the interaction of members in these online communities and participating in the interaction oneself.
4. Getting informed consent and ensuring the reliability and credibility of data are some of the ethical guidelines that need to be considered while conducting Internet research.



ignou
THE PEOPLE'S
UNIVERSITY

UNIT 2 INTERNET RESEARCH TRADITIONS

Structure

- 2.0 Introduction
- 2.1 Learning Outcomes
- 2.2 Studying the Internet
 - 2.2.1 Categories of Internet Research
 - 2.2.2 Characteristics of the Internet
 - 2.2.3 Issues and Concerns
- 2.3 Phases of Internet Research
 - 2.3.1 Technical Issues with the Internet
 - 2.3.2 Uses and Users of the Internet
 - 2.3.3 Effects of the Internet
 - 2.3.4 Future Improvements
- 2.4 Internet Research Methodologies
 - 2.4.1 Online Surveys
 - 2.4.2 Online Interviews
 - 2.4.3 Web Content Analysis
 - 2.4.4 Social Network Analysis
 - 2.4.5 Digital Ethnography
 - 2.4.6 Data Analytics
 - 2.4.7 Online Experiments
- 2.5 Theorising the Internet
- 2.6 Let Us Sum Up
- 2.7 Keywords
- 2.8 Further Readings
- 2.9 Check Your Progress: Possible Answers

2.0 INTRODUCTION

In the previous Unit, we discussed the generic overview of the Internet or online research. This Unit continues that by covering the actual academic activities/practices and traditions of Internet/Online research in the recent past.

This Unit will introduce you to the vast and unprecedented field of Internet research. Starting by establishing its origins, we will cover various aspects of study while studying the Internet. Moving forward, we will discuss the multiple paradigms covered by the broad discipline of Internet research. Next, you will be presented with traditions and practices adopted by researchers in different phases of the research in this field. We will also discuss different methods to study the Internet over the years. You will also learn to appreciate factors influencing the Internet and those which are, in

turn, affected by it. Lastly, discussing the vistas opened by the discipline, from psychology to governance to education and even health, we will try to chart its boundaries.

2.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Describe various paradigms covered by Internet Research;
- Identify traditions and practices in different phases of Internet Research;
- Evaluate different methods adopted for Internet Research; and
- Recognise factors influencing the Internet and vice versa.

2.2 STUDYING THE INTERNET

The Internet is simply understood as a global network of connected computer systems that can communicate. Over 5 billion people worldwide, an estimated 65 per cent of the current world population, are using the Internet. The uses range from tasks related to everyday life to researching and contributing to local and global issues. The Internet has not only expanded geographically, but it has also penetrated the social strata of society, making the user base a vast and heterogeneous body. The diversity in usage patterns, user base and applications has led to the emergence of several issues for inventors, users, and researchers alike. These issues range from privacy concerns, the digital divide and the quality of content being hosted online.

The Internet emerges not only as a subject but also as a tool for research and a larger social phenomenon affecting many walks of academia. Internet studies is a comprehensive field of study that encompasses all issues pertaining to the Internet, be it its design, its use, or its impacts on various segments of society. Internet Research Traditions pertain to the methods acquired and applied by scholars to study the Internet. But what do we exactly study when we study the Internet?

2.2.1 Categories of Internet Research

Recall your Internet usage since this morning. How many apps have you accessed, how many social media profiles have you checked? How much content have you consumed, be it checking the weather update, watching an episode of your favourite show, or searching for your favourite recipe? This brings us to the question - what do you exactly study when you study the Internet? Is it the content? Is it the user? Is it the traffic on a webpage? As a user, you do not have access to a lot of data, like the code and the traffic analytics of websites. The amount of accessible data is also vast. For example, social media platforms generate content in uncountable numbers and create billions of comments every day. So, how does one even start quantifying the units of study?

Internet Researchers have explored varied topics to study, analyse, understand, and define. These range from qualitative aspects to quantitative as well. They can be broadly divided into the following categories:

1. **Technology:** The Internet is a technological innovation first and makes use of technology furthermore in designing the websites, browsers, networks, applications and even infrastructure of the World Wide Web. Marshall McLuhan has famously said, "Medium is the message", and being technologically driven, innovations for the Internet are also tied with technological innovations. This is why the Internet of Things, Artificial Intelligence and Virtual Reality are currently emerging as hotcakes in Internet research. Thus, researchers studying technology can observe how it is evolving and affecting the structure and applicability of the Internet.
2. **Usage:** The applicability of the technology emerges as the second research category. The Internet has found several ways to reach the public, from ICT (Information and Communication Technologies) to social media to e-commerce. In a networked society, the offline and online populations become the same, thus diversifying usage. On the other hand, unlike researchers of different mass media, Internet researchers acknowledge that the audience on the web is not just a consumer but a producer. Thus, the researchers examining the web usage would explore the user base and the content being hosted and consumed on the web. This can encompass qualitative and quantitative study methods as well.
3. **Policy:** As Internet penetration and its usage increase, it becomes necessary to streamline and monitor it. Without central governance, Internet policies face difficulty in formulation and application. Policy research focuses on how technology governs human interaction and, in turn, affects national and international information infrastructures. The researchers of the Internet policy analyse the existing guidelines and framework regarding Internet usage, suggest future courses for policymakers, and conclude their research.
4. **Impact:** Overlapping with the field of psychology and sociology, the researchers of this paradigm study the effects of Internet usage on individuals and groups. Whether through online education or e-commerce, apart from creating new virtual avenues for communication, the Internet also modifies human interaction in the offline world. Several disciplines associated with the study of humans have started acknowledging the Internet as a catalyst in human development and are probing into this further.

Emerging as a multidisciplinary field, Internet Research broadly adheres to these four categories. The research methods and theoretical frameworks have also been derived from the disciplines. For a long time, the Internet was termed and treated as 'New' media, and thus, the dichotomy between old and new media existed for researchers. Today, the line has been erased as the Internet generates its own line of methods, theories, sampling techniques and theoretical frameworks.

2.2.2 Characteristics of the Web

While charting a journey to study the web, there is a need to distinguish the fundamental qualities that make it different from any other traditional

communication media format. Following are a few characteristics of the web that make studying it more challenging and unprecedented for Internet scholars:

Intertextuality: While reading a book, you might have got stuck on a word naming a phenomenon. You would have wondered about it and then proceeded with the text with a mental note to check more about it later. Later, you might have even forgotten about it, but knowing it at the time you came across it would have also added to your reading. On the web, jumping to a subtopic and returning within a short time is easier. Think of having a dictionary or encyclopaedia on the side, open at the right page all the time.

The text on the web is interconnected with the sub-topics it holds; it does not exist in isolation. Think of a tree branching out to meet other trees. Hyperlinks connect the topics with each other, sometimes leading to a section on the same page or sometimes to a different page altogether. These links allow the reader to jump to any definitions, explanations, and data verification and then return to the topic immediately. This enriches the reading and allows the reader to branch out as per their own interests.

Nonlinearity: The traditional text layout is linear and has a specific beginning and end. Look at the example of this Unit; while you may choose to skip the beginning or start from the end, it has been designed to be read in the direction of beginning to the end. Compared to this, the text on the web is non-linear, and the reader can jump through the hypertexts, as discussed earlier.

The function of nonlinearity allows the reader to go as deep into the text as they want, thus making the reading process highly subjective. The reader may also choose to continue reading the text further from different sources for as long as possible. In this case, the text becomes endless and depthless. This feature implies that each reader on the web will derive a different meaning from the same text.

Prosumerism: Traditionally speaking, the reader has remained at the receiving end of the content, rendering them passive. Look at this Unit again; the writer wrote it sometime back, and you are reading it today. We may not have any other communication apart from this text. You do not have a way to reach out to me to share your feedback or ask any questions. Some methods may have feedback options, like classroom teaching, writing a letter or a telephonic conversation. Feedback from the audience is also a relatively new concept in communication studies.

In the case of the web, the reader can get in touch with the writer, edit the existing text or even create their own text. He or she may also dissect it and disagree with it. You might have seen fan-made trailers of films or have read fan fiction. Wikipedia is editable, and the readers can alter the text. This feature of the web not only raises questions about the responsibility of the reader but also about the text's credibility. On the other hand, prosumerism also helps establish the Internet as a medium, breaking away from the hegemony of mainstream media by challenging the traditional producer-consumer hierarchy and establishing the Internet as an alternate medium.

These are a few of the features of the Internet that affect how scholars study it.

2.2.3 Issues and Concerns

Like any other field of study, the Internet has issues and concerns about research. Let us discuss a few of them in detail:

1. **Digital Divide:** While increasing daily, Internet access is still not available to all populations in the world. Many researchers, specifically in third-world countries, struggle with limited Internet access. Their research also does not garner as much exposure, thus tying the fulcrum in favour of the Western lens and literature in research. This disparity further causes an imbalance and bias in data, affecting the research conclusions.
2. **Sampling:** Content worth more than 320 Million TB of data is generated daily on the Internet. Vast swathes of the population are active on the web daily, in varying degrees and for multiple purposes. With a vast universe of research, sampling can be problematic for researchers. Identifying not only a representative number but also a sample representative of the vast body is a task in itself.
3. **Reliability and validity:** In offline research, data reliability and validity can be assessed using multiple tools. But on the web, one does not have a fool proof method to verify whether they are interacting with a bot or a human being across any platform. Similarly, concerns with individuals' identities and the validity of data obtained are major issues for Internet researchers.
4. **The impermanence of the text:** As readers can also alter the text to innumerable iterations, studying the text becomes difficult, if not obsolete. The text on the web can be edited and, at times, completely removed. The data on the web is dynamic, so the same test can yield different results every time it is run.
5. **Security concerns:** Internet researchers must also tackle cyber threats and other cyber security issues, such as hacking, phishing, etc. while conducting the research. If mishandled, the data gathered for the study can be accessed without authority and pose severe threats to all parties involved.
6. **Ethical concerns:** The researchers also struggle to distinguish between personal and public online. Accessing and using data for research can breach an individual's privacy, while informed consent can be difficult to obtain from a large sample. One may also get entrapped in legal issues if one is not careful about copyright laws and intellectual property rights.

Activity: 1

Chart the Internet usage of your family members. What purpose do they use the Internet for? How long are they exposed to it? What devices do they use to log into the Internet? Do you see any patterns in your records based on age, gender, or employment?

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. How are Internet research topics categorised?
.....
.....
.....
- 2. What are a few ethical concerns related to Internet Research?
.....
.....
.....

2.3 PHASES OF INTERNET RESEARCH

Although predated by the actual Internet itself, Internet research began when researchers started exploring or discovering it. The field, now well established, can be divided into broad chronological phases. The phases rolled out with the development of the Internet, its increasing usage and resulting critical analysis by the researchers, which has helped tremendously in further development. This pattern highlights the circular relationship between the developer, the user, and the researcher.

While academia has had the task of researching appropriate subjects at the needed time, the Internet has made access to research possible across the globe, thus accelerating it. As discussed earlier, the Internet has emerged as an object of study and a tool, archive, and publisher. But this has not happened overnight; it has taken decades and can be categorised now in hindsight. The four phases of Internet Research can be broadly categorised in the following manner:

2.3.1 Technical Issues with the Internet

As the Internet was in its infancy, the first phase of Internet research dealt with philosophical and technical issues. One school of thought that emerged analysed the Internet as a stand-alone invention and studied its phenomenon in isolation. This school believes the Internet will/is cutting off humans from society. In a highly dystopian approach, the researchers in this group thought that the Internet would cut off civilisations from each other. This group of researchers exists even to date.

Another significant perspective that emerged was looking at the Internet as a continuum with society, as with any other scientific invention and its application. The researchers of this group analysed the functions and services provided by the Internet and how they compared with other media and communication platforms. The researchers approached the Internet as a tool and looked for ways to sharpen it for better use. The extreme ideological perspective of this group believed that the Internet would create a super smooth network across the globe, connecting all citizens and empowering all sections of society.

Still an invention, the majority of the research in this phase explored the scientific and technical issues emerging. Connectivity, speed, communication technologies, data archival, and access were the first popular subjects of research, focusing on the functionality of the World Wide Web. Due to its utility, a major focus was also laid on the economic repercussions of the Internet in the workplace or commercial sectors. This phase rolled into the second phase as the focus shifted to utility from mere function.

2.3.2 Uses and Users of the Internet

The next phase of Internet research explored various Internet applications and how users interacted with them. The academia had accepted and experimented with the Internet as the browsers had made it easy for the general user to access it, and websites became navigable for laymen. Marketing firms became the first organisations to study the consumer base of Internet platforms to suit their purpose. User perceptions and attitudes were surveyed with the help of questionnaires and explored with the help of interviews and focus group discussions held online and offline for triangulation. This phase also saw collaborations between social and computer scientists.

Comparing the differences between the demographics of offline and online populations, these studies worked out the usage patterns and traffic on the web. This helped identify the socioeconomic strata on the web and thus quantify the penetration of the Internet. This phase also helped reduce the gap between the utopian and dystopian views adopted by the scholars in the first phase. While the world has become a global village, scholars have worked on the idea that equality and equity are still long-term targets.

2.3.3 Effects of the Internet

As the Internet progressed from a novel invention to a utility for the masses, the research focus also shifted from merely studying the Internet as a phenomenon to studying its impact. Its involvement in daily life had increased manifold, and the transformation of users from consumers to prosumers was initiated in this phase. Many interdisciplinary researchers had forayed into Internet research by this time, enriching it with theories, methods, and literature from their respective disciplines.

The Internet's various psychological, sociological, economic, and political impacts were studied during this time. With social networking, online communities have started mushrooming. The online and offline personas

were bifurcated and provided material to study for the researchers. Online communities started formulating cultural practices, while psychographics tracking helped predict online behaviours. From political parties holding election campaigns online to e-commerce, entire countries and economies began running with the help of the Internet. This research phase is still on the rise in third-world economies as more and more disciplines are associated with the Internet. A more considerable discipline of Digital Humanities uses Internet penetration and other digital tools in the humanities paradigm.

2.3.4 Future Improvements

In the running phase, academia has moved forward from accepting, exploring, and studying the Internet. Researchers are now exploring ways to incorporate the Internet into their disciplines while suggesting ways to improve existing applications. Various concepts and models related to the Internet, directly or indirectly, are structured and defined. Newer technological developments such as augmented and virtual reality, machine learning, and artificial intelligence have paved the way for future research. As the world finally becomes a global village with comparable offline and online populations, cultural exchange and knowledge sharing globally becomes a thing of discussion. The research lens shifts to a critical analysis of the Internet and its applications, and the fulcrum shifts to refining its scope for better applications and usage. The future will tell what this phase of research summarises.

Activity: 2

Consider the four phases of Internet research. Which one do you most relate to? Do you believe in the dystopian view or the utopian one? What phase of research do you relate to most?

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. What are the four phases of Internet Research?

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

2. What were the utopian and dystopian views in the initial Internet Research?

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

2.4 INTERNET RESEARCH METHODOLOGIES

Due to the vastness of the subject, the researchers have devised and employed many methodologies for the study of the Internet. Few methods have been adapted from offline research methods already in practice, while few have been designed specifically for the web. Let us discuss a few of them in brief:

2.4.1 Online Surveys

Akin to the offline format, online surveys make use of a questionnaire distributed amongst the targeted sample to extract specific information. The questionnaires are created in the form of a digital form, which can be easily circulated using Internet-based communication tools such as mail, social networks, etc. The answers gathered through it can also be automatically coded in desired formats, saving labour on the part of the researcher. Targeting the sample online also makes it easier to reach people even in remote locations as well, saving the researcher the effort of travelling as well.

2.4.2 Online Interviews

Online interviews are conducted using digital tools in an online space. The interviewer and interviewee meet virtually via telephonic tools, text-based chat, or video calling. In theory, these interviews are like the offline counterparts only, but they have the opinion of being recorded. Moreover, these interviews are conducted without either of the parties moving outside their comfort zones.

2.4.3 Web Content Analysis

As discussed before, the content on the web ranges from text to images to graphics and even audio and video. This content is extracted, sampled, and studied extensively by researchers using this method. In the initial days of Internet research, the researchers used to code the content manually. With the advancement in research tools, automated programs today can extract and code the content for easier analysis.

2.4.4 Social Network Analysis

This methodology studies the relationships and interactions between individuals or groups within social networks. A sociogram is used to depict the connections between people, represented through nodes and the connections are depicted through routes. These diagrams help the researchers identify the information flow and study the community's anchor points. This further helps identify the opinion leaders in the group and checks how information can be filtered and disseminated.

2.4.5 Digital Ethnography

This studies online populations aggregating in an online group or platform. The virtual group usually starts cultivating its values and traditions, becoming a subject for observation and study. Like offline ethnography, the researchers join a community, observe their traditions and practices, and develop a 'thick description' of the group, albeit in the virtual space.

2.4.6 Data Analytics

This method extracts data sets and employs algorithms and statistics to analyse them. The data could range from user profiles on a web page to traffic patterns. The analysis helps identify emerging patterns and generate insights within the data for practical applicability. The data sets are usually large, ranging in millions and billions, thus called 'Big Data'.

2.4.7 Online Experiments

Experiments are methods of research where a group of participants are studied before and after introducing a stimulus. They begin with a hypothesis, and the procedure helps prove or disprove the hypothesis. In an online format, the researcher changes the variables on a particular platform for a specific group, and the impact is studied to conclude.

These are a few of the research methods employed to study the Internet. Researchers often combine these methods to approach complex problems, uncharted territories, and new research paradigms. The choice of method also depends upon the objective of the research, the target sample, and the availability of resources.

Activity: 3

You may remember filling out an online consumer survey for a brand. What sort of questions do they ask? What do you think they do with the data? How do you think the brand benefits from this research? Would you be more comfortable if a representative came to your house to ask these questions?

2.5 THEORISING THE INTERNET

Internet Studies have not emerged as a school of thought, or an academic discipline structured under institutional boundaries. It has emerged organically as individuals across the globe started working on it simultaneously. This is why the research literature overlaps and is ambiguous, as many terms remain differently defined by scholars. For e.g., Online Ethnography is also known as Cyber Ethnography, Virtual Ethnography and even Netnography, all of which refer to the study of communities in an online sphere.

Starting as a network to facilitate the military, the Internet became a research-worthy subject due to its expansion, penetration, and popularity. Email changed the landscape of human communication, catalysed further by social media platforms. The Internet of Things has further connected machines and humans, blurring the boundaries between man and data.

AoIR, or Association of Internet Researchers, is one of its first kinds. It is an early association of researchers that was gathered with the help of a mailing list. Its first congregation is associated with the formal emergence of Internet Studies as a discipline, although the Internet predates it by decades. While some believe that Internet studies are an overlap of many fields and not a

field in itself, others believe that it has fragmented and differentiated itself, such as communication researchers differentiated themselves from sociology. Many scientific and peer-reviewed journals also started being published during the 2000s to facilitate Internet research in academia.

As the psychographics of the users are mapped to feed into the algorithm and change the landscape of the Internet for each user, the fact remains that the Internet is customisable and thus appears different for everybody. It becomes even more difficult to universally document something that can be labelled as a 'shapeshifter' at best. Echo chambers not only exist for Internet users but also for researchers, for the Internet is not just a subject but a popular medium for conducting research.

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. What are different methods of research on the Internet?

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

2. What is AoIR, and what is its importance?

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

2.6 LET US SUM UP

In this Unit, you were introduced to the discipline of Internet research, starting from its origins to exploring various study topics under its aegis. We discussed various phases of Internet research and their different traditions. We further explored various methods of research that have been adapted for the Internet and discussed various issues and concerns that arise while studying it. Lastly, we identified factors influencing the Internet and those governed by it.

2.7 KEYWORDS

Avatar: Visual identifiers of the individual identity in the digital world. Based on the idea of changing shape and identity as in Hindu mythology.

Chatbot: A pre-programmed software that can emulate a human conversation.

Cyberculture: Practices and traditions developing in a community congregating on the Internet using digital tools.

Emoticon: It is an icon which conveys a particular emotion. Used in text-based chatting on digital platforms.

Netiquette – A set of guidelines for online behaviour.

2.8 FURTHER READINGS

1. Burnett, R. (2011). The handbook of Internet studies. M. Consalvo, & C. Ess (Eds.). Oxford: Wiley-Blackwell.
2. Dawson, C. (2019). AZ of digital research methods. Routledge.
3. Graham, M., & Dutton, W. H. (Eds.). (2019). Society and the Internet: How information and communication networks are changing our lives. Oxford University Press.
4. Hunsinger, J., Klastrup, L., & Allen, M. (Eds.). (2010). International handbook of Internet research. Springer Science & Business Media.
5. Paulus, T. M., & Lester, J. N. (2021). Doing qualitative research in a digital world. Sage Publications, Incorporated.
6. Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, culture & society*, 30(4), 22-46.
<https://research.gold.ac.uk/id/eprint/7978/1/DigitalDevicesDistCopyJSept2013.pdf>
7. Snee, H., Hine, C., Morey, Y., Roberts, S., & Watson, H. (Eds.). (2016). Digital methods for social science: An interdisciplinary guide to research innovation. London: Palgrave Macmillan.
8. Tsatsou, P. (2016). Internet studies: Past, present, and future directions. Routledge.

2.9 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Internet research is broadly categorised into the following segments:
 - a. The technology used for the Internet
 - b. Usage of Internet
 - c. Policies for Internet
 - d. Impact of Internet
2. Internet research has several ethical concerns. First, researchers struggle to draw a line between personal and public on Internet platforms. Accessing and using data for research can breach an individual's privacy, while informed consent can be difficult to obtain from a large sample. One may also get entrapped in legal issues if one is not careful about copyright laws and intellectual property rights.

Check Your Progress: 2

1. The four phases of Internet research are:
 - a. Technical Issues with Internet
 - b. Uses and Users of Internet
 - c. Effects of Internet
 - d. Future Improvements
2. The ideological perspective of Internet research believed that the Internet would connect a super smooth network across the globe and would empower all sections of society. On the contrary, the researchers with the dystopian approach believed that the Internet would cut off civilisations from each other.

Check Your Progress: 3

1. There are many ways to research on the Internet. A few of the following are Online Surveys, Online Interviews, Web Content Analysis, Social Network Analysis, Digital Ethnography, Data Analytics, and Online Experiments.
2. AoIR, or Association of Internet Researchers, is one of its first kinds. It is an early association of researchers that was gathered with the help of a mailing list. Its first congregation is associated with the emergence of Internet Studies, although the Internet itself predates it by decades.

UNIT 3 DEFINING AND FORMULATING RESEARCH PROBLEMS

Structure

- 3.0 Introduction
- 3.1 Learning Outcomes
- 3.2 Selecting the Research Problem
 - 3.2.1 Difference between a Social Problem and a Research Problem
 - 3.2.2 Importance of Review of Literature
 - 3.2.3 Questions of relevance, feasibility and achievability
- 3.3 Formulating the Research Problem
 - 3.3.1 Research Questions
 - 3.3.2 Research Objectives
 - 3.3.3 Research Hypotheses
- 3.4 Defining the Terms of Enquiry
 - 3.4.1 Concepts
 - 3.4.2 Constructs
 - 3.4.3 Variables
 - 3.4.4 Establishing Operational Definitions
- 3.5 Let Us Sum Up
- 3.6 Further Readings
- 3.7 Check Your Progress: Possible Answers

3.0 INTRODUCTION

In this Unit, we shall discuss the process of selecting and formulating a research problem. In this regard, surveying the available literature is the most crucial step, though peers, sponsoring agencies, and the researcher's own observation of the environment can also suggest a research problem. Once the broad area has been selected, it is then necessary to formulate the research problem in terms of specifics such as research questions, objectives and hypotheses. For this, reviewing existing literature in the concerned discipline becomes essential. The more specific and clear you are, the better, since well begun is half done. If you are clear about what you want to research, then how you will conduct your research is merely its logical extension.

It is also important to understand the constituents of the terms of enquiry, i.e., concepts, constructs and variables. Their varying values in terms of occurrence, degree of intensity, nature of direction, etc., enable us to draw conclusions about the associations, correlations and causality that exist in human and social phenomena at any given point in time and space.

3.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Differentiate between a social and a research problem;
- Identify the factors that determine the selection of a research problem;
- Discuss the importance of review of literature;
- Define hypotheses and explain their various types;
- Discuss concepts, constructs and variables; and
- Explain the need for and importance of operational definitions in research.

3.2 SELECTING THE RESEARCH PROBLEM

This section will focus on the difference between a social problem and a research problem. We will also explain the importance of reviewing what has already been studied so that you do not end up 'rediscovering the wheel' unless you are interested in replicating previous studies to test or strengthen existing theories. We will finish this section by outlining various factors that go into making the final decision regarding the selection of the research problem.

3.2.1 Difference between a Social Problem and a Research Problem

Therefore, a social problem involves society and has a broader framework than a research problem. A research problem has a limited context, primarily involving academia or industry. Social problems also deal with various issues, including politics, culture, and religion, while research problems are often limited in scope and well-defined.

A social problem has the element of public concern, while a research problem can exist at the level of curiosity alone, i.e. when a scientist does research just because s/he wants an answer to a question. It can also exist at the profit level, for example, when a manufacturer seeks to identify the optimum market for a new product.

A social problem (riots, global warming, female foeticide, etc.) exists irrespective of the researcher. The researcher, on the other hand, formulates a research problem. For instance, the researcher could be interested in the role of media in covering communal riots or awareness of female foeticide.

A social problem deals with mandatory (e.g., the legal age of marriage is 18 years) or normative propositions (e.g., literacy should serve as the foundation of a nation) propositions. A research problem deals with propositions that can be proved or disproved through empirical determination. For example, early marriage leads to health problems, or literacy leads to social awareness. The validity of both these statements can be tested through research.

A social problem is a trigger or a launch pad for formulating research problems. Research findings, on the other hand, provide descriptions, explanations, predictions and solutions to social problems. For instance,

research could establish that motherhood at a tender age is associated with the death of the mother during childbirth or a higher risk of infant mortality. Therefore, we can say that the findings of research studies enable the framing of mandatory propositions, laws, and policies.

A social problem is a one-time problem. It finishes once it is tackled or until it emerges again. The researcher formulates a research problem that can be repeated per his/her requirement. In this regard, experimental research studies set up in laboratories are a case in point.

A social problem disrupts the smooth running of the social order, while a research study causes no such disruption other than minor intrusions into the respondents' lives.

A social problem can be dealt with using intuitive thinking, common sense, etc., while a research problem uses scientific thinking based on empirical logic and reason. Moreover, a social problem can suddenly arise, necessitating spur-of-the-moment decisions. On the other hand, a researcher systematically proceeds to identify, formulate and lay down a procedure to study a research problem, and only then does the researcher start data collection and analysis.

3.2.2 Importance of Review of Literature

Certain relevant factors must be considered after identifying the broad area of study and before finalising the research topic. This is where the literature review comes in. We have reviewed the existing literature on our chosen research topic to ensure it is relevant, feasible, and achievable within a particular time frame, budget, and infrastructure.

Review of literature economises efforts and maximises profit. It gives you an idea of what research has already been done, by whom, where, when and how. As an extension, it also tells you which aspects have been largely ignored or which methodologies have not been used - in effect, it allows you to identify the research gaps in subject matter, theoretical perspectives, methodologies, etc. You can then decide to either replicate previous research and strengthen existing theory or take up new areas and thereby revise existing theory or formulate new explanations for existing phenomena. A review of the literature gives you an idea of the practical problems faced by previous researchers as well as takes up their suggestions for further research. A literature review also gives inputs regarding methodology and terms of inquiry and insight into defining the variables you intend to study in your research.

Literature refers to all information in the printed or oral form available on your research interest topic. For a review of literature, one can consult: Indices—these contain alphabetically organised listings of research articles under subject, title and author headings as well as information regarding where the article can be sourced, e.g., the Social Sciences Citation Index, the Humanities Index, etc. Unit 6 deals extensively with this topic.

Bibliographies are akin to indexes. Other than that, they concentrate more on periodical/journal literature. A bibliography can also be limited to books

published in a particular discipline during a specific period.

Compilations of Abstracts can also be consulted. These are typically arranged according to subject matter and author and contain summaries of specific published articles. In that sense, they are more beneficial since one can decide to consult the full study based on whether the abstract is relevant to one's study, e.g., Dissertations, Abstracts, International Journals and Periodicals are other sources of literature. They are publications of a fixed periodicity and discipline that contain full papers, articles, reviews, etc.

Books and Encyclopaedi are also consulted for an exhaustive analysis and overview of one's interest area. An encyclopaedia can be particularly useful for framing operational definitions.

General and professional/trade publications are other sources. Articles for laymen appearing in newspapers and magazines can provide interesting perspectives on complex problems in Communication and media for researchers to pursue. Many trade publications are published in films, advertising, public relations, etc., which can provide case studies in respective areas.

The Internet is an immense source of information. The only drawback is that one must be careful regarding the credibility of the information source online. It is advisable to refer to the websites of recognised journals and research institutions.

3.2.3 Questions of relevance, feasibility and achievability

A good research topic can be related to a theoretical frame of reference. For instance, if you are interested in doing research in advertising, you can refer to the Uses and Gratifications theory, Diffusion of Innovations theory, etc. or any other theory discussed in detail in the Course on Communication and Media Studies. This will root your research in an existing body of theory and make it easier to identify potentially fruitful questions. This theoretical base will also enable your research to find its niche in the existing knowledge since your findings will either support or help revise the existing theory. Even if you are researching a completely new area, you can still consult theories from related disciplines and develop a better understanding of your subject. For instance, if you are researching social media addiction, you can always draw upon theories from psychology and link them to the study of a mass medium. A review of the literature of existing research studies helps to determine how relevant your research would be concerning the existing knowledge - will it fill a gap, or will it strengthen it - in other words, is your chosen research problem significant?

One also needs to determine the feasibility of the topic before starting work on it. A topic might prove unsuitable for investigation simply because the question being asked cannot be answered with the facilities and information currently available. For instance, there is no point in taking up a topic that requires classified information for analysis or any other information that cannot be legally obtained. You may not have the necessary budget or infrastructure to use the technology needed to investigate the problem. Often,

the essential technology may not even exist. For instance, we only now have technologies like brain mapping to determine how individuals react to message stimuli.

Moreover, a topic does not lend itself to productive research if it requires data collection that cannot be observed or measured with validity or reliability. If controlling for error is essential for you, then such a topic may not be feasible. Having an obvious idea of one's capabilities is equally important. If one is not a team member, it would not be feasible to discuss a topic requiring large-scale collaboration. Similarly, taking up a qualitative study is better if one is uncomfortable with numbers or statistical analysis.

Before finalising one's research topic, one also needs to address the question of achievability. You might be engaging in research to fulfil the requirements of a degree, such as a dissertation. Naturally, you will be working on a deadline if you have a research grant or a fellowship for a specific period. Therefore, ensuring that your research topic is not too broad and unwieldy is essential. One tip is to write down your proposed title and then attempt to dissect the topic into small questions. For instance, Computer-mediated Communication as a tool for social relations - this working title will immediately ensure that you draw up boundaries and concentrate on this aspect to arrive at closure within a specified time period.

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. List two differences between a Social problem and a Research problem.

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

2. List the sources of information for conducting a review of the literature.

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

3. How can you determine the relevance of your topic?

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

3.3 FORMULATING THE RESEARCH PROBLEM

In this section, we will focus on formulating the enquiry terms for your research. These primarily constitute research questions, objectives and hypotheses. All three don't need to be included in every research study. For instance, you may just have some research questions and objectives when

conducting exploratory research. You may have research objectives and hypotheses in more structured studies that address association, correlation and causality issues. A study does not necessarily become weaker in the absence of hypotheses. The importance of formulating the terms of enquiry lies in the fact that it narrows down the research problem to a manageable specificity, lays down the extent and limits, and identifies the relationships proposed to be significant - in short, all that you hope to achieve through your study.

The critical thing to remember is that there is coordination when formulating the enquiry terms. For instance, if your research question is 'whether demographic variables impact Internet usage', then the corresponding objective should be to collect data about the demographics of the Internet user and his/her Internet usage patterns. Only when you have this data can you answer your research question. How the question is answered will prove or disprove your research hypothesis that perhaps could be that 'socioeconomic status and frequency of Internet use is positively co-related'.

3.3.1 Research Questions

Research questions are usually asked in exploratory research when not enough is known about the research area to justify framing categorical hypotheses at the outset. Since the researcher is working in an area that has only been studied marginally or not at all, s/he may be unsure of the existence of the determining factors or the interrelationships. At this point, research is a knowledge-gathering exercise rather than verification. So, s/he asks a series of questions, gets many answers, and looks for patterns, and it is only at the end of research that s/he may be able to frame hypotheses to be taken up later in a more structured study.

3.3.2 Research Objectives

When a researcher is working in an area for which extensive literature is already available, it acts as a facilitating mechanism for him/her to frame the research objectives straightaway, followed by research hypotheses. The latter is like an answer- a tentative proposition whose validity remains to be tested. For instance, s/he can directly list a hypothesis - 'media consumption and need gratification is positively correlated' - without first framing it as a question. While framing the objectives, you must remember that they should be exhaustive and logically sequenced. For instance, an objective to determine media usage patterns should be listed before an objective to study the reasons for those patterns. Also, only those objectives that are empirically achievable must be listed, i.e., amenable to being observed or measured.

3.3.3 Research Hypotheses

As mentioned earlier, a hypothesis is a tentative generalisation, the validity of which remains to be tested. Hypotheses enable testing existing theories, suggesting new ones, and helping describe social phenomena with a certain degree of certainty. For instance, a hypothesis such as 'the more the exposure to televised violence, the stronger the belief in a violent world' enables you to test George Gerbner's Cultivation theory. Similarly, if proven, a hypothesis

such as 'communication breakdown between couples is the primary factor for divorce' enables you to describe the reasons behind divorce rates in the country. The importance of hypotheses in a research study lies in their ability to bring direction, specificity and focus. They help us control the intervening variables and enable decision-making and policy formulation.

Hypotheses can be classified based on various criteria: source, level of abstraction, composition, and functionality.

Based on source:

- a. The researcher can frame a hypothesis based on his/her perception or intuition regarding a particular phenomenon. For instance, a researcher may belong to a family where the men only discuss issues regarding their work or business while the women only discuss domestic issues. The researcher here draws upon one's personal experiences and hypothesises that 'conversational engagement is associated with identity perception'. These kinds of hypotheses make an important contribution to the theory and are useful in exploratory research since they bring forth interesting questions.
- b. The researcher can frame a hypothesis based on existing theory. The genesis here lies not in his/her intuition but in the findings of specific research conducted to study a particular phenomenon. For instance, s/he can take up the want-get ratio formulated by Daniel Lerner and seek to replicate it. S/he, too, may find that this ratio regularly operates in underdeveloped countries while not so much in the developed world. S/he may also identify gaps in existing research and frame a hypothesis accordingly. For instance, one may find that loneliness has not been studied as a factor in media consumption, so s/he can frame a hypothesis such as 'social support and media consumption are associated'.

Level of abstraction:

- a. Level of commonplace impression: These kinds of hypotheses are observations about some empirical uniformity with respect to the phenomena under study. The focus here is on the descriptive; for instance, 'male reporters cover hardcore reporting.'
- b. Level of conceptualisation: These hypotheses are based on existing theory where a hypothesis has been more or less proved but is open to falsifiability. For instance, the Two-step flow hypothesis is relatively well-established in understanding how information flow occurs in remote areas. Here, information is routed through the local opinion leaders like the *sarpanch* etc., rather than directly through the mass media.
- c. Level of specification: These hypotheses specify the nature of the relationship between the variables that constitute the hypothesis. A hypothesis may state that something tends to be accompanied by something else, i.e. it states the association between the variables - 'Education is associated with civic awareness. Specifying the degree and direction of the relationship can also state a hypothesis, e.g. 'Increase in leisure time leads to increase in media consumption. A hypothesis

regarding a causal relationship between variables can also be stated: 'Continuous exposure to computer screens causes dry eye syndrome'.

Based on composition/complexity:

- a. A simple hypothesis has only one or two variables. For instance, 'The Times of India is the most circulated newspaper in India', or 'Age and Internet usage are associated'.
- b. A complex hypothesis has more than two variables. For example, an increase in age and socioeconomic status leads to an increase in media consumption. These hypotheses are more difficult to test because the more variables in a single hypothesis, the more difficult it is to assess their interrelations—quantitatively and theoretically.

Based on functionality/testability:

- a. A theoretical hypothesis is stated in terms of abstract concepts and is untestable in any direct sense.
- b. A research hypothesis is derived from the above but is stated so that it can be tested, i.e., it is a functional, testable, or working hypothesis. Most research methodologists recognise only two hypotheses under this category - Research hypothesis and Null hypothesis. The theoretical and statistical hypotheses are assumed to be understood as part of the research hypothesis. Here, the hypothesis regarding some operations to be performed, i.e., construct language, is stated. 'I.Q. levels and comprehensive abilities are positively correlated. To prove or disprove this hypothesis, the constructs of I.Q. and comprehension are measured with the help of test instruments. Any standardised I.Q. test can be used to determine the I.Q. To measure comprehension, we could construct a test instrument such as a reading passage followed by questions - the respondents' test score will rate their comprehension capability.
- c. A null hypothesis can be deemed the reverse of the research hypothesis or its negation. To take up the above example, the null hypothesis would be, 'I.Q. levels and comprehensive abilities are negatively correlated' or 'I.Q. levels and comprehensive abilities are independent of one another. It is the null hypothesis that is tested in research. This practice is followed by convention because testing the null hypothesis first makes the researcher declare his/her objectivity. Since the null hypothesis denies what the researcher holds to be true, according to its priority, s/he declares that s/he is giving all the alternatives the first chance to be proved.

The concept of null hypothesis is based on the law of probability. The research hypothesis and the null hypothesis are alternative statements. Both have the probability of being true or false, so if one is proven, its alternative, by logic, is considered disproven (though it should be noted that neither the research hypothesis nor the null hypothesis can be true or false under any given test of it. That is why we say that the data failed to support the hypothesis or failed to reject the hypothesis). Moreover, it seems easier to prove something false than to prove it is true - one can

say that there could be many reasons for a particular effect, but this one is not the cause.

- d. A statistical hypothesis refers to the technical form in which the hypothesis is stated for purposes of a statistical test, i.e. it is indicated in a manner that makes it amenable to applying the relevant statistical tests of significance. The data is quantified, confidence levels are set, tests are applied, and values are obtained and checked to determine whether they fall in rejection or acceptance. It is important to note the difference between significance and importance in research methodology. The former is mathematical, while the latter is subjective, based on experience and judgment. A hypothesis may be rejected based on not clearing a statistical test or being set to specific parameters, but that does not mean it ceases to be important.

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. When are research questions asked?
.....
.....
.....
- 2. What is the importance of hypotheses?
.....
.....
.....
- 3. Define a Null hypothesis.
.....
.....
.....

3.4 DEFINING THE TERMS OF ENQUIRY

In this section, you will study how hypotheses are framed. The criterion of a good hypothesis is that it should be clear and concise regarding what is to be studied and how. In that sense, it should also be testable. Whether the hypothesis is merely descriptive or analytical, it will always have one or more variables to be observed or measured. In certain instances, this task is easy, while in other cases, procedures have to be put in place so that the observation or measurement can take place. Since we are dealing with human and social phenomena, we must also understand that the same phenomena can have multiple meanings depending upon the context of time, situation and person. Therefore, it becomes important to understand how variables should be defined for a specific research study to ensure that the findings are understood from the right perspective.

3.4.1 Concepts

Concepts can be called observables, i.e., terms with direct empirical referents. A concept is formed by summarising related observations of either directly observable or easily measured characteristics. For instance, income is a concept that can be deduced through observations of a person's lifestyle or directly measured in terms of monthly/annual income. We can say that a concept is a term formed by generalisation from particulars. We can examine the properties of a particular book - size, weight, thickness, paper quality, print, subject matter, the ratio of visual content to verbal content, etc. and slot it in the genre of textbooks, reference books, fiction and so on.

3.4.2 Constructs

Constructs can be referred to as non-observables as they cannot be easily related to the phenomena they are intended to represent. They have no direct empirical referent, i.e. they are not directly part of our empirical world. For instance, anxiety, violence, faith, intelligence, social prestige, attitudes and so on. You cannot point to something and say this is faith. To observe or measure faith, you will have to frame specific parameters or indicators to do so, and depending upon how many of those parameters are met, you will be able to state the quantum of faith possessed by a particular person or group. Therefore, we can say that constructs can be defined as higher-level abstractions constructed from concepts at a lower level of abstraction. Researchers often construct their own indicators to concretise a relatively abstract phenomenon, or they combine concepts to explain a construct. In that sense, its exact meaning relates only to the context in which it is found. For instance, to measure social status, a researcher may combine the concepts of income, education, occupation, etc.

3.4.3 Variables

A variable is the empirical counterpart of a concept or construct. It is capable of taking on different values; thus, in effect, a variable is something that varies. Variables are important because they link the empirical world with the theoretical - they are the phenomena and events that can be measured or manipulated in research. A person's age varies in terms of years, income varies in terms of money earned, and intelligence varies in terms of scores achieved on an I.Q. test and so on. A researcher uses variables to measure how they affect other variables. They can also be manipulated to study the corresponding effect of this manipulation on other variables. Researchers try to test several associated variables to develop an underlying meaning or relationship among them. After suitable analysis, the most important variables are retained. These are marker variables and continue to be built upon as research continues in a particular area. Variables can be classified according to the manner of their usage.

- a. Independent variables are those that are systematically varied or manipulated by the researcher. Dependent variables are those systematically observed for concomitant variation due to the manipulation of the independent variable. For instance, to measure whether the camera angle influences the credibility that a viewer attaches

to a news reader, a researcher may shoot the same news bulletin using three different camera angles - top, low and eye level. S/he then shows this news bulletin to three groups and asks them to rate how credible the newsreader is. S/he may find that the group shown the news bulletin where the news reader is looking directly into the camera found the bulletin to be most credible.

In this example, the camera angle is the independent variable, and the effect of its variation is measured on the dependent variable, i.e. credibility. Since this is an experimental study with active manipulation of the independent variable, this variable can also be referred to as the active variable, the manipulated variable or the experimental variable. The dependent variable here can also be referred to as the passive or measured variable. In non-experimental research where there is no active manipulation of variables, different terms can be used - the variable that is used for making predictions or is deemed to be the cause of something, e.g. if poverty is present, so will illiteracy, this variable can also be termed as the predictor or the antecedent variable. In contrast, the variable predicted to result or is assumed to be affected is called the criterion variable.

- b. When specifying relationships in a hypothesis, it is important to account for or control the extraneous or intervening variables. If this is not done, the findings may be erroneous or invalid. For instance, we may have hypothesised that education leads to modernity. We collect data regarding the two variables and prove or disprove the hypothesis accordingly. But it may so happen that even though education impacts whether a person develops a more modern outlook or not, no education will greatly change a person's outlook unless his/her family is also a progressive one. However, since we have not factored in the family background variable, we have erroneously held education responsible or not responsible for a modern outlook. In contrast, the causal variable had been the family background.

Since many factors combine to create a particular effect in human and social phenomena, the researcher must either include all the variables or control them. For example, if the researcher feels that family background could be an intervening variable, s/he can control it by only including those respondents in the sample who come from similar backgrounds. Doing so will cancel its effect out of the equation since this variable will now be held constant. When such a thing is done, these variables are called 'control variables'.

Intervening variables can also be understood as those without the intervention, for which the relationship between the independent and dependent variables cannot be established. Suppose you want to study the relationship between age and newspaper readership; you assume that older people read the newspapers more. However, without the intervening literacy variable, the relationship between the independent and the dependent variable will not be complete.

3.4.4 Establishing Operational Definitions

It is very important to define the variables under study, as failure to do so impacts the reliability and validity of the study. Generally, any first attempt at a definition applies common sense and everyday life experience. For instance, the most conventional approach to defining age would be to define it in terms of years lived beginning from birth. However, as knowledge grows, concepts evolve, meanings change, and perspectives multiply, it becomes important to categorically state how one will study a particular concept for a specific study. For instance, one researcher may want to study violence purely on physical terms so s/he defines it accordingly. Another researcher wants to study it from the point of view of verbal violence, yet another wants to distinguish between animated and actual violence. Still, another wants to study it only from the perspective of the victim and so on. As complexities increase, so does the need to focus on specifics.

When defining the variables under study, two approaches are followed - convention, i.e. framing a definition in terms of what is most commonly understood, or framing a definition in terms of the researcher's understanding or objectives. In the latter case, the researcher has to ensure that the definition remains within the bounds of logic and reason. Conventional definitions are usually constitutive, nominal, or formal, defining a word by substituting other words. These are traditionally the dictionary meanings. For instance, status is defined as a person's standing in society; anxiety is defined as nervousness, etc.

The stipulated definitions are usually the operational definitions. An operational definition defines a concept in terms of the operations by which it will be represented in a particular study, i.e., it specifies the procedures to be followed when expressing or measuring a concept. In a way, operational definitions are quantifications of the nominal definitions. For instance, violence could be defined as any physical act that results in bodily injury to be measured in time units. Anxiety could be described in terms of the score obtained on a five-item scale or the basis of observation of the occurrence of specific physical actions such as restless gestures, perspiration, stuttering, etc. An operational definition provides an empirical reference to the variable under study.

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. What is the basic difference between concepts and constructs?

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

2. Define a variable.

.....

.....

.....

.....

3. What is an operational definition?

.....

.....

.....

.....

3.5 LET US SUM UP

In this Unit, we discussed how to select a research problem. We started by differentiating between a social and a research problem. We then highlighted the importance of conducting a literature review, which helps identify and select the research problem. We also learned the importance of ensuring that our selected topic is relevant, feasible, and achievable.

We further focused on formulating the terms of enquiry—research questions, objectives, and hypotheses. The various types of hypotheses were categorised. We also described how the hypothesis is framed using variables. The difference between concepts and constructs was highlighted, and the importance of framing clear operational definitions was emphasised.

It is hoped that the above discussion will help you define and formulate your own research problem. In the next Unit, we shall move forward with the discussion and focus on sampling methods.

3.6 FURTHER READINGS

1. Kerlinger, F.N., (1973), *Foundations in Behavioural Research*, Holt, Rhinehart and Winston. N.Y. Kumar, Ranjit (2008), *Research Methodology*, Pearson Education. N.D.
2. Neuman, W.L., (2007), *Social Research Methods- Qualitative and Quantitative Approaches*, Pearson Education. N.D.
3. Wimmer, R.D and Dominick, J.R (2010), *Mass Media Research - An Introduction*, Thomson Wordsworth Publishers, London.

3.7 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. A social problem can suddenly arise while a research problem is

formulated. Secondly, a social problem has the element of public concern, while a research problem can exist at the level of curiosity alone.

2. Consulting indices, bibliographies, abstracts, journals, periodicals, books, encyclopedias, general interest publications, trade publications, and the Internet can help conduct a literature review.
3. The relevance of a research topic can be determined by ensuring that it has a theoretical frame of reference as its foundation, thereby either strengthening existing theory or revising it. Its relevance also exists in its significance—it should fill research gaps in subject matter, make the methodology more precise, or have implications for policy formulation.

Check Your Progress: 2

1. Research questions are usually asked in exploratory research when not enough is known about the research area to justify framing categorical hypotheses at the outset.
2. Hypotheses enable us to test existing theories, suggest new ones, and help us describe social phenomena with a certain degree of certainty. From the point of view of methodology, hypotheses ensure a more precise study; they impose boundaries, clarify concepts, and enable controlling for intervening variables and testing for statistical significance.
3. A null hypothesis is based on the law of probability and is deemed the reverse of the research hypothesis or its negation.

Check Your Progress: 3

1. The basic difference between concepts and constructs is that concepts have a direct empirical referent, which constructs do not have.
2. A variable is a concept or construct that takes on different values; in effect, it is something that varies.
3. An operational definition defines a concept in terms of the operations by which it will be represented in a particular study. It specifies the procedures to be followed in expressing or measuring a concept.

UNIT 4 SOURCES OF DATA

Structure

- 4.0 Introduction
- 4.1 Learning Outcomes
- 4.2 Primary and Secondary Data
- 4.3 Sources of Secondary Data
 - 4.3.1 Internet search
 - 4.3.2 Library search
- 4.4 Sources of Primary Data
- 4.5 How to Store and Save Your Data
- 4.6 Online Data
- 4.7 Let Us Sum Up
- 4.8 Further Readings
- 4.9 Check Your Progress: Possible Answers

4.0 INTRODUCTION

Media research is increasingly important as information is integral to 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 as 'data'. In understanding data for research, it is important to distinguish data by the source from which it is derived and how it is gathered.

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

As you may be aware, the term 'data' is plural. 'Datum' is the term used for a single piece of data. Our focus is on data. **Data collection** is the process where information is gathered and measured on variables of interest. It is important at all stages of a research process. Good data is needed to develop research questions and hypotheses and their subsequent tests.

4.1 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.

4.2 PRIMARY AND SECONDARY DATA

As you know, research is based on data, not opinions and hunches. Therefore, the data collection process must be accurate and appropriate. Sometimes, the approach to a research problem can depend on the collected preliminary information. 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 collected fresh and/or for 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, the government conducts the National Family Health Surveys. Such data collected from the field is primary and analysed before being presented to us in reports. When we read the reports, we look at secondary rather than primary data.

There are several methods of collecting primary data. We may collect the data through surveys, observations, discussions, and interviews with experts and respondents in our selected research area. We may also collect primary data through photographs, audio, and video recordings 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 have not been published anywhere else; they have not been interpreted or translated.
- Data thus collected helps to create original documents because the origin is the research conducted in the field.

Secondary data, on the other hand, are data that are already available and have been collected and analysed by someone else. Such data may have already been analysed through statistical procedures, such as census reports and national readership surveys. In these cases, an agency has collected data from the field, tabulated and analysed it and presented it in tables, 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 several sources, such as publications of central and state governments, annual reports, technical and trade journals, books, and research reports. Some such data may have been published in journals, while others may have been unpublished reports.

The key factors to remember about secondary data are:

- Secondary data are not original.
- Secondary data have already been analysed, so using secondary data means examining already examined data and conclusions.
- Secondary data may have been taken from a source but repackaged for objectives that may differ from our research.
- Secondary data may have been analysed differently.
- Secondary data is a cost-effective method of data collection, but such data must be reliable, valid, and accurate. For example, Wikipedia is not a reliable secondary data source for academic research.

There are some advantages of using secondary data in research. The use of secondary data is cheaper, timesaving, 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.

However, 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 of 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 yardsticks to use for evaluating secondary data.

Most research in mass communication uses a combination of primary and secondary data. The extent to which a data source is used depends on the research undertaken. The first issue that comes to mind is what kind of research we are undertaking. Let us examine this regarding the research classifications presented in Unit 2.

There are four requirements for 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, we would have no choice but to collect them afresh. We may also get several physically disabled persons, but we may not know specifically who they are in a family.
2. **Relevance:** The data must be relevant to the problem under study. Do not use irrelevant data sources.
3. **Accuracy:** Secondary data must be accurate and reliable. There must be clarity on how the data were collected, the methodology used, the margin of error, and whether the sources were dependable.
4. **Sufficiency:** There must be enough data available. If not, 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 critical factors in choosing secondary data?

.....

.....

.....

.....

4.3 SOURCES OF SECONDARY DATA

Let us examine secondary and primary data sources in more detail, especially the processes followed in conducting research studies. Communication is an interdisciplinary subject; therefore, we often research topics 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 look at various 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 the information to be found will guide us in proceeding. 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; secondary data and secondary data sources are most important here.

4.3.1 Internet Search

In today's world, however, an Internet search is the first place to start. 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 keywords 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 social sciences include Google Scholar, the Social Science Citation Index, The Arts and Humanities Citation Index, The Communication Initiative Network, ERIC, JSTOR, Science Direct, and Infilbnet.

Google Scholar

When you search Google for a topic, the search engine will show you what articles are listed and emerge in Google Scholar. Google Scholar is a web search engine focusing on academic 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 websites. Google Scholar will then provide the link, telling you where the book or article exists. Sometimes, Google Books (where you can read the book or article) will give a link and show you how to access it. It is an excellent place to start your hunt for secondary sources. You will also find out if the full article is free of cost and on payment, who the copyright holder is, and how the article will be cited in the bibliographic citation format needed in academic publications.

Social Science Citation Index

The Social Science Citation Index (<http://mjl.clarivate.com/cgi-bin/Ernest/options.cgi?PC=SS>) covers about 3,000 journals across 50 social sciences disciplines. You can search the index by title, author, subject,

journal, or ISSN number, which is the international journal classification. You can view a list of 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 keywords.

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, including indexing and abstracting. You can search by topic 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 100,000 individuals and organisations working in communication and development. Search their database for research papers, case studies, and 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 weekly. The ERIC Collection, begun in 1966, contains records for various 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 to 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 developed countries through partnerships with philanthropic institutions. JSTOR has more than 10 million academic journal articles, 50000 books, and 2 million primary source documents in its digital collection. 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 massive 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 there are no mass communication materials. However, you must understand that because mass communication is interdisciplinary, secondary data sources in other subjects, like Psychology,

Education, etc., are also important. Also, suppose you are looking for research on theme areas in development, such as health communication. In that case, a wealth of articles will be 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 support academic research and learning through cooperative facilities among universities and other academic institutions. Inflibnet has many activities, including 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 several bibliographic, citation and factual databases in different disciplines from many 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, 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 helpful for students and researchers as it informs you 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, paying US\$35 just to view an article can be very prohibitive. This is where access to a library becomes important.

4.3.2 Library search

Before the Internet age, the library was where a researcher would go to collect data. The library stores books, journals, indexes, reports, newspapers, magazines, and unpublished documents. Not only do libraries hold published materials, but many also hold collections of old manuscripts and scrolls in hard documents and within their microfilm, microfiche, and electronic resources sections. Libraries can also access electronic 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 online searches and drawers full of alphabetically organised library cards. 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 are housed, such as Encyclopaedia, Yearbooks, and even the hard copies of the Citation Indexes mentioned in the earlier section. Once you find the journal article reference, go to the journals section of the library. This is where journals are kept; you may find the journal's issue there. Newspaper articles are usually kept in the library's microfilm or microfiche collection 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 provides documentation and information support to researchers in the social sciences and contains a huge collection of documents. Scholars working in autonomous research organisations, universities, and government agencies can easily use this facility to obtain data.

Newspaper and Magazines

In media research studies, newspapers and magazines can be important in collecting secondary data. With the development of the 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, 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 (not peer-reviewed), Wikipedia sources, and individual blogs are not recognised as valid sources for academic research.

Much of the secondary data sources are consulted at the formative stage of a research project, much before the researcher goes to the field. Secondary research sources form the heart of the Review of Literature discussed in Unit 5. The literature review forms the basis for developing research questions and hypotheses, which later link the current research to a body of theory on the subject. Identifying, sifting the useful from the irrelevant, and finding knowledge gaps that need to be filled requires understanding how to find secondary data and how to use it. And finally, because our secondary data sources must be reliable, we must 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 Inflight useful for Indian academics and researchers.

4.4 SOURCES OF PRIMARY DATA

Primary Data are original research data collected by an organisation or an individual, i.e., first-hand data. 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, reliable, and relevant primary data is discussed in various units of this course. At this stage, we will simply discuss the multiple ways in which primary data can be collected. Collecting primary data is often time-consuming and 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, they become secondary data for another group.

What are the characteristics of primary data?

- Data is collected in its original form.
- Such data have not been analysed or published anywhere else. They may be fresh data collected in the field, relevant only to 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, which can also be original data. We will discuss different kinds of primary data in mass communication regarding the different objectives of a given research effort.

Suppose your research aims to understand the role of the press in the Independence struggle in India. In that case, you will look for books and

articles about the topic. This is **historical research**. You may also go to the original newspapers' pages and original audio recordings. For example, in pioneering research on the role of Gandhi and Mao as communicators, Dr Kusum J. Singh (1980) used the originally 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 by leaders and historical figures. In all cases, you will use original materials. They include letters, life histories, memoirs, diaries, biographies, autobiographies, etc. These help get information from the person himself and thus are reliable enough to be considered primary data. If you are doing legal research, you will study case files, legal decisions and previous case histories and decisions. These documents give information regarding events in the past and are written by an individual to narrate incidents in his/her life.

Suppose your research aims to analyse the gender-based communication patterns in a village, especially the access and use of the Internet. However, you have little readily available data to base your hypothesis. In this case, you would carry out **exploratory research** in the village, using various primary data tools to collect primary data. We will quickly look at various primary data collection methods without going into detail because these are discussed in subsequent units.

Survey

As a first step, you may carry out a household door-to-door Survey to determine your sampling frame accurately. This survey would give you information about each household's members, gender, age, educational level, etc. Such information will help you 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

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, cell phone towers, and routers. You may take some photographs and recordings.

You may participate in the village activities to understand the dynamics. You would then use participant observation, which is also part of ethnographic research. You would record your observations of the participation in a diary or journal. The diary or journal notes become part of the 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. You are conducting **Interviews** with individuals or women in groups by talking to them. Collecting data through Interviews is discussed in detail in Unit 10.

Experiments

A significant amount of primary data can be collected through experiments. As discussed in Unit 9, experimentation is 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 help you frame your research questions, hypotheses, sample sizes and samples.

Primary data are not only collected at a preliminary stage. Primary data will also be collected during the fieldwork part of your research. All the data collected initially as part of your fieldwork 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 necessary for research. Without secondary data, one cannot begin to sharpen the research focus, define a problem in addressable terms, frame questions and hypotheses, know what information must be collected and what is not needed, and make reliable judgments. When primary data are collected and analysed, findings must be presented regarding prior research and evidence collected from secondary sources. Without such a connection, the continuity of the research tradition is 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 gathered in a tea shop while women gathered at the temple. What kind of data have I collected?

.....

.....

.....

4.5 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, 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 video tapes, 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, the examiner may want information about your source if you are writing your thesis. You should have it readily available for scrutiny when questioned. You may also have to add 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 initially, you may land 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/>) to reference a book, a journal article, and/or a web page. Save those details in a folder or file with each citation on your computer or library cards.

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

4.6 ONLINE DATA

In this section, you will learn about the nature and application of Digital Media Research Methods (henceforth DMRMs). This understanding of DMRM (also called new media research methods) will help you understand the concept of new methods developed due to the advent of new communication technologies. The information delivered through digital media platforms needs a different set of research approaches, some entirely new methods, and, in a few cases, old and conventional research methods that have been tweaked and adapted to the new digital media environment.

Digital methods have previously been described as 'approaches to the web as a data set' (Hutchinson).

You will learn the concept of big data and methods to spot the relevant digital data while simultaneously applying the DMRMs principles. This would help you to have insight into the data from various publicly available datasets and your customised data in the form of webpages, scrapped information, RSS feeds, etc. DMRMs tend to alter the monotony associated with conventional research methods while ushering into a world of endless "fun-laden learning". You would learn the "computational aspect" of the data, which is widely applied in all disciplines like pure sciences, humanities and social sciences, arts, etc. For systematic and easy learning, this DMRMs chapter has been divided into data movement types, online places for data availability, benefits of open data and different data characteristics, which cover the whole gamut of topics. The Internet itself is a huge graveyard of information where the data, once uploaded, remains in the servers of services like "Internet archives" (www.archives.org), which aim to give free "universal access to all knowledge" with access to digitised materials on software, websites, databases, books, images etc. The archived data relating to news, online videos, discussions, debates, discussions, blogs, Facebook posts, Twitter, YouTube comments, etc., presents a big challenge and opportunity for the researchers.

The modern world is data driven. The parameters of volume, velocity, format, type, and authentication characterise it. Data is regarded as the most sacrosanct component of any research study. Data (datum singular) is regarded as a scientifically recorded factual piece of information which can be appropriated for further analysis. This way, the content available in online newspapers, social media platforms, forums, government documents, etc., constitutes big data that can be harvested using relevant insights tools. According to Luo and McKinney, "research methods are defined as specific methods and techniques employed to collect and analyse data in a research study". Hence, any software, retrieval or analysing tools, data formats, techniques or commands, freeware, open databases, data format converters, websites, apps, scripts, browser extensions, etc., form the ecosystem of the DMRMs. In this part, you will acquaint yourself with the types of data and the format available for harvesting. The data online is either structured (excel format, CSV format or Txt format) or unstructured (mail, scanned documents, images, reports, videos, etc). Mostly, it is unstructured, and even if structured, it must be customised as per our research objectives and purposes.

Two types of data movements are traversing the globe – free and open data (also democratising data). The former is free to use (with a few restrictions like non-commercial use, giving credit, etc.). In contrast, the latter is openly distributable for all purposes, such as remixing and onward distribution of the improved dataset. "Open data" is information accessible to everyone, machine-readable, offered online at zero cost and has no limits on reuse and redistribution (Tran & Scholtes, 2015). It is free from copyright, patents, licensing restrictions and other controls. However, terms and conditions may apply. Open data and online repositories gained popularity after 2009, after

the signing of an executive order for open government data by U.S. President Barack Obama. International Open Data Day by Open Data Enthusiasts is an annual event that "promotes awareness and use of open data". "It takes place globally, usually in February or March and typical activities include talks, demonstrations, seminars, hackathons, training or the announcement of open data releases or other milestones in open data" (Wikipedia). "International Open Data Day is a gathering of citizens in cities around the world to liberate data, create open map and write software applications to analyse, visualise and publish the results" (www.opennepal.net/odd2017/). Historically, the open data concept dates to the time of Enlightenment, when a few philosophers criticised the idea of state secrecy. However, the present open data movement has roots in the "open government" movement in Britain in 1800 and in the passage of the Freedom of Information Act in 1966, which allowed proactively the partial or full disclosure of the information held by the government. However, in the USA, the case for open data movement was strengthened with the Amendment to the Electronic Freedom of Information Act of 1996, which allows citizens to request information in electronic form in whatever format they require. A few of the world's open data initiatives have been mentioned below:

www.data.gov.in. As regards the Indian scenario, this open data platform of the Indian government, under the Digital India Initiative, is of considerable value to the researchers, the world over in general and India in particular.



It has high-value datasets on all important government organs, such as legislation, agriculture, transportation, etc. The datasets can be filtered before downloading by department/ministry, state and union territories, and sector type. In addition to the usual features, the site has ready-made visualisations generated for a few datasets. The additional link "open data site" on the website makes it further useful by mentioning authentic databases of governments around the world. Any user can also recommend a particular dataset to the site.

www.data.gov—This is the initiative of the U.S. government's extensive data on climate, agriculture, consumers, ecosystem, energy, finance, health, local government, manufacturing, maritime, public safety, and science and research. It has 1.93 lacs (and growing) structured datasets that are freely

downloadable.

<https://www.archives.gov/open/available-datasets.html>—This website hosts dataset links for all U.S. government information, such as Presidential speeches, proposed amendments to the Constitution, trade agreements, government manuals, etc.

www.data.gov.uk – The U.K. government's open data is available on transport, health, crime and justice, education, society, environment, etc. The data is filterable according to the criterion of relevance of search, most viewed, frequently downloaded, time frame, etc. It has site analytics too.

www.cia.gov/library/publications/the-world-factbook/geos/in.html - This is the world's fact book maintained by the Central Intelligence Agency of the USA, which has detailed information regarding geography, people, transportation, communication, energy, government, territorial disputes, etc., for every country.

www.github.com/planetopendata/awesome-world—This website is the biggest source of open data generated by users and contributors. GitHub is the world's largest platform for sharing the source code of various open-source programs.

www.data.worldbank.org—The World Bank's open access to development data is the most comprehensive and exhaustive for various indicators. This provides datasets on government performances on various development indicators and miniature data on a particular country's performance on various indicators. This list of data topics is far more extensive and very useful.

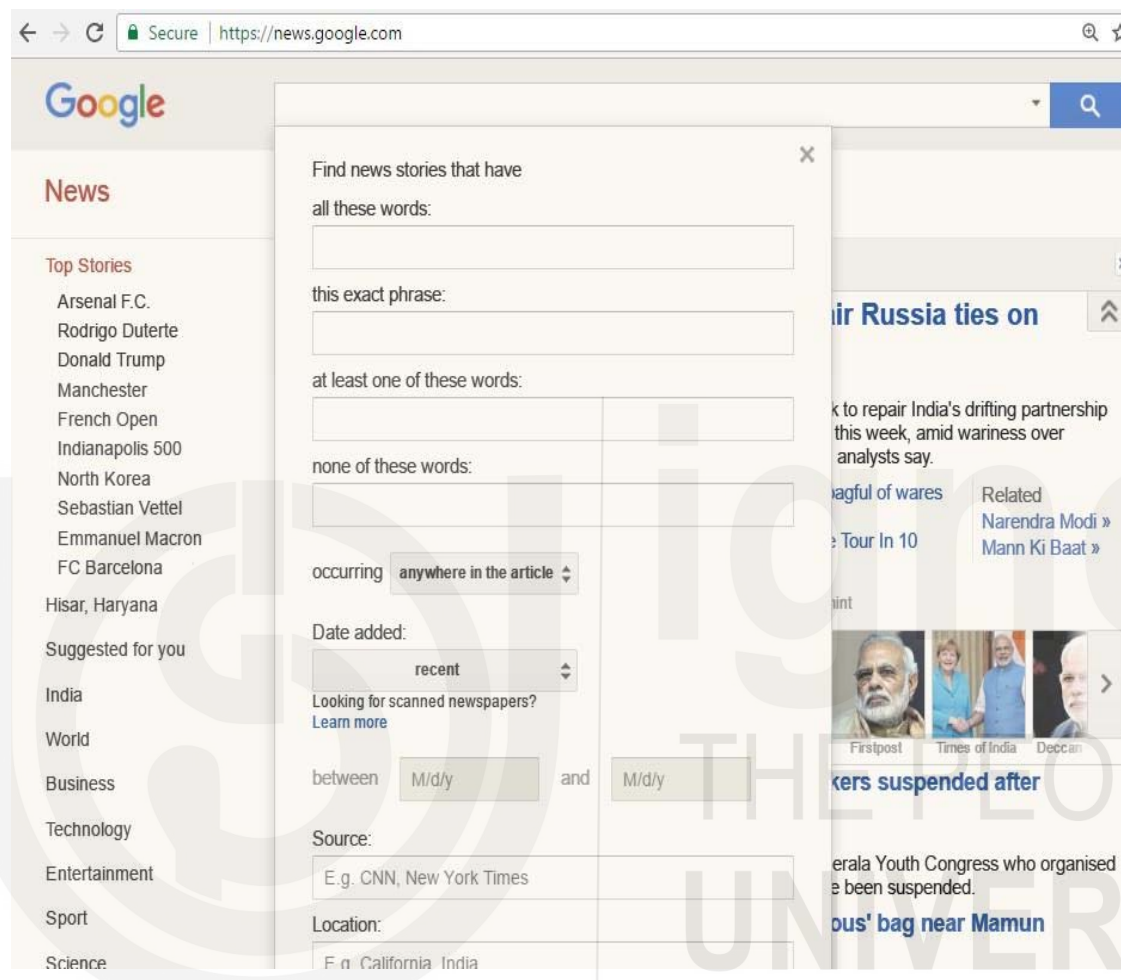
www.data.oecd.org – The database of (the Organization for Economic Cooperation and Development) is another user-friendly free resource equipped with online statistical analysis functions and visualisation. The data can be filtered for various parameters before downloading.

www.familysearch.org/wiki/en/Digital_Historical_Newspapers—Newspapers record any country's significant and historically important aspects; hence, the archive serves the idea of "memory freeze". The above website is a good repository for digitally scanned print newspaper enthusiasts, where all the historic newspapers of the world (sorted continent-wise) are freely available.

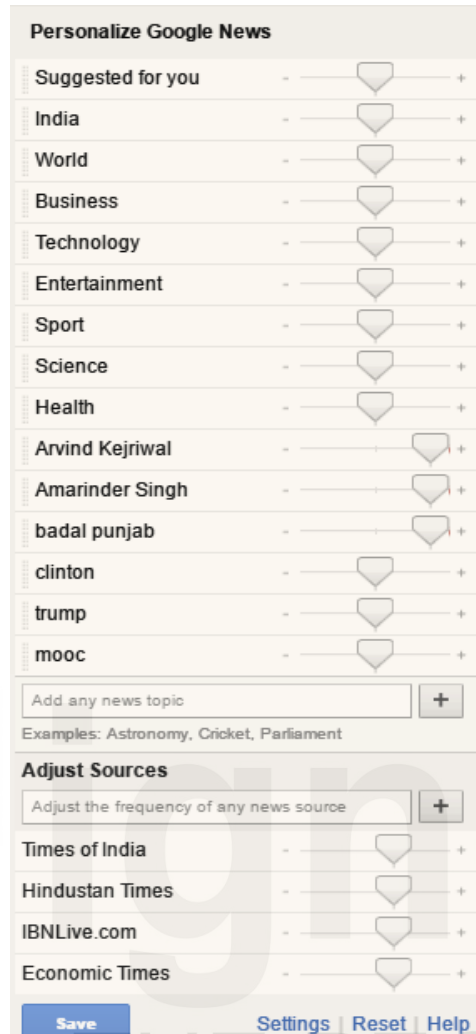
www.en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives—It has a list of online newspaper archives of newspapers from different countries, and the links are hosted here for the third-party websites.

www.news.google.com/newspapers - This is the world's biggest collection of very old newspaper editions from all newspapers, sometimes dating back to 1830. This site scores many good points for content analysts who can search the whole archive of all newspapers for their research query (e.g. economy, crime, finance, culture, etc.). The extremely old content is scanned, and the same is amenable to search. The scanned image cannot be saved, but a good freeware screenshot can serve the researcher's purpose here.

Another good example is Google News (www.news.google.com), which has a real-time and archived collection of world news. It covers virtually all news sources, such as news websites, T.V. channels, financial newspapers, news portals, etc. The trending news topics (e.g., Trump et al. elections, etc.) alongside the page ensure the grouping of all news items on a single topic in various international languages and regional ones, too. The user can search Google News with various filters.



The search further retrieves data on specific dates, such as one-year, last day, last week, etc. Also, the data is further categorised according to its nature, e.g., article, blog, multimedia, editorial, all types, etc. The user can view the news by using a particular word and not by using any other word. Researchers can find the news from a selected source (BBC, CNN, etc.) and decide the news's location. The "exact phrase" search function allows the exact sequencing of more than one word in a search, e.g., the "demonisation story" query would be searched as such.



This facility gives you the remarkable control over the data to be retrieved. The researcher can prioritise the search terms and topics (as shown in the figure) and adjust the news source. Another very useful feature is the RSS (Really et al.) feed function, where the researcher can generate his own RSS link by choosing the nature of the query (e.g. global warming, crime, etc.) and frequency of the news source. All these functions give the researcher a complete command over every word in the news world.

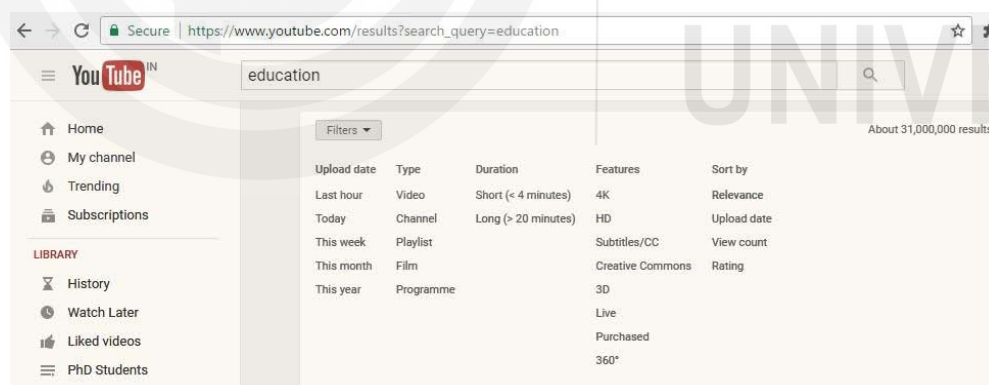
All newspapers have their data live data as archived editions open to data retrieval (also known as data scraping). A few of them have the option of "advanced search" on their sites. The New York Times has had digital copies of all newspapers since 1851 at www.query.nytimes.com, and it allows searching as per the keyword (while showing the associated phrases related to the keyword as an instant result).

This is similar to the case of other big newspapers like The Guardian (www.theguardian.com/gnm-archive). A few universities' websites and non-profit organisations have their newspaper data portal (single gateway to all the newspapers as per country, continent, globe, language.) to narrow focus their results. A British newspaper archive dating back to 1710 is at <http://www.britishnewspaperarchive.co.uk>, with 1.9 million pages and 757 titles. Similarly, The Times of India Historical Archive at www.library.upenn.edu/news/799 has digitised newspapers from 1838 to 2001.

Regarding data accessibility, these online resources and their features save much time and energy for fledgling researchers. The RSS (Really Simple Syndication or Rich Site Summary) feeds, too, is an effortless method of accessing the updated news data. www.newsrack.in/extras/known-indian-feeds is a nice collection of all Indian newspapers' RSS feeds (on many topics in the same newspaper), which can be fed to an RSS reader, and desired content can be read. This data can be easily harvested using free online resources and imported into data analysis software.

Similar web services exist for other forms of data (blogs, videos, podcasts, user comments, discussion forums, groups, etc.). Blogs are the centre of the universe for "citizen journalism". The latter may be regarded as the "fifth estate" after the "fourth estate" (Media). There is documented evidence of citizen journalists pulling down authoritarian governments (Arab Spring in Egypt), and the blog phenomenon is the oldest among all citizen journalism tools. The researcher who wishes to access the data on blogs exclusively can search it on www.blogsearchengine.org. Similar blog search engines exist for Blogger (www.searchblogspot.com), WordPress (www.en.search.wordpress.com), and Tumblr (www.tumblr.com/tagged/search-engine). These blog services provide APIs to researchers to provide data harvesting facilities. Retrieving the blog's content for research without APIs is a cumbersome task. Online apps can read and harvest the blog data for the researchers as per their objectives.

Many useful visual data (videos) are uploaded to the biggest video-sharing website, www.youtube.com, a qualitative and quantitative resource for researchers. It is presumed that the next generation of learners (present too) would be "visual learners" predominantly. The videos and the underlying user's comments also serve a useful purpose.

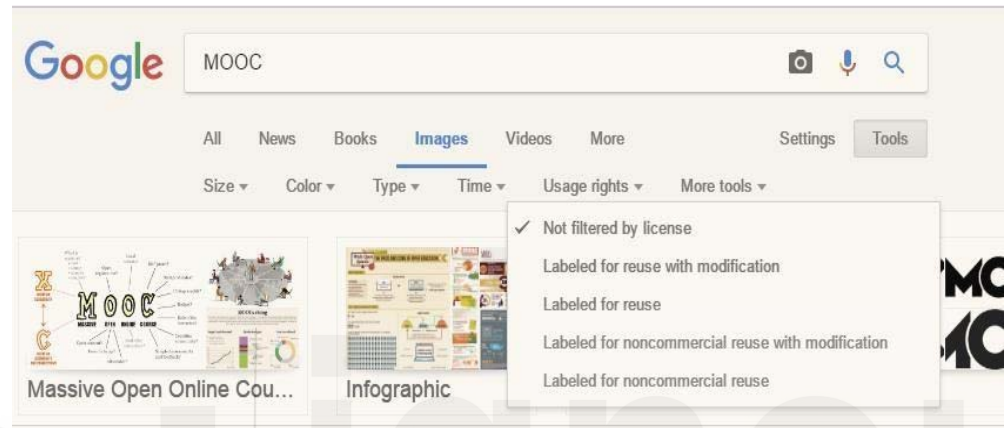


YouTube is a wide repository of visual and textual data. There are a few less explored functions in YouTube, like filters (to filter the videos as per date, type, duration, view count, rating, licenses, view quality, most liked videos etc.). There are various playlists (collections of videos) on a topic or company where multiple videos in a series can be watched). The built-in video analytics, too, can throw light on videos from Vimeo and Facebook and can be downloaded and imported into video analysis software. YouTube data (video comments) can be downloaded with the help of open-source software.

This type of data harvesting facility has given rise to new paradigms of audience research, where the public (user) is seen participating and giving

feedback on the media content. Such data can be imported into corpus linguistics software to analyse various lexical patterns and discourse analysis.

Visual communication research through still images has undergone a sea change with the advent of image collection algorithms employed by Google Images (www.images.google.com). The site doesn't have its own database of images. Rather, it retrieves photos from other websites and blogs in real time when searched. "A picture is worth a thousand words", though a cliché, can still make the job of visual researcher much easier to work:



The site allows you to apply various filters before choosing your suitable image. You can select images according to size (small, big, larger and pixel defined by you), colour (colour, black & white, transparent, etc.), image type (face, photo, clipart, drawing, animated, etc.), time (time of picture upload) and, most importantly, the usage rights.

The digital literacy paradigm expects the user's fair use of the referenced material. This contrasts with "digital skills", which measure the user's familiarity with the digital tools. Indiscriminate (improperly referenced, too) use of copied images, text, video, etc., may result in plagiarism, which is gross academic misconduct. A few websites (www.pixabay.com) offer free images and video downloads that are copyright-free, allowed for commercial use and without the requirement of attribution. www.pexels.com and www.imagesource.com are a few of them. The best choice is www.zipplist.co, which hosts links to the best free resources for photography, audio, icons, graphics, videos, textures, and colours. A search on Google may yield more useful and free websites. Flickr (www.flickr.com) is another great collection of photos on various topics, but a few usage restrictions may apply. The comments posted on the photos by the users pose yet another data mining opportunity.

Regarding hugely crowdsourced data on social media like Facebook, Twitter, YouTube, LinkedIn, etc., a big opportunity lies in these networks for brands' public sentiment, engagement, and reaching out to the next-gen audience. A separate chapter deals with social media analysis and data harvesting using sophisticated (yet free) tools.

Readers' comments on online newspapers' news present a new challenge and opportunity for a researcher, who is working on a totally new kind of reader feedback mechanism. A few websites like NDTV (www.ndtv.com/trends/)

most-commented) give ample opportunity to a researcher to gauge readers' reactions to the media's performance.

Podcasts (audio files in MP3 format) of speeches, conversations, dialogues, scripts, etc., are relatively less valuable for media researchers unless transcribed first and subjected to qualitative data analysis.

The data must be cleaned (Google et al. and other freeware) before being analysed. Various online tools are freely available to tweak the data as per your requirements. Google's Fusion Tables are a very nice tool that can create valuable new datasets by fusing more than one secondary dataset already available in the public domain.

This way, the secondary datasets from different countries (the first column being the same) can be joined, and a whole new primary dataset can be generated quickly while sitting on the desktop.

The data is by default in Excel or CSV format for the databases, but a few new formats have come up, e.g. Jason (for tweets & Facebook), txt (Text) format with UTF.

Eight formats for corpus-enabled textual analysis. HTML (Hypertext Markup Language) format is well-known for the webpages. Similarly, PDF, JPEG, CSV, DAT, FLV, MP4, doc, PPT, etc., are the widely known common data formats which need to be converted into the desired formats (mostly Excel for categorical data and numerical data and in text format for textual analysis). API (Application Programming Interface) is the latest form of data extraction protocol in which the data is requested from websites (e.g. Facebook, Twitter, YouTube, etc.) with API facilities for developers. The App requests the information from the user's behalf, and the API allows the continuous streaming of data in JASON format, which is converted into the desired format (excel or txt) by the App.

Open data, according to a study at Harvard University, "enhances government's efficiency, fights govt's corruption, improves accountability" and leads to more transparency, openness, and participation, creating new journalism models by separating the signal from the noise and incubation of multi-billion-dollar businesses based on data of public sector (Kundra, 2012) citizen engagement and information based effective governance & public policy. While not limiting the benefits to the government sector alone, open data benefits all sectors across the board – teaching and research, manufacturing, NGOs, community workers, media, civil society, analytics industry, etc.

4.6 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 importantly, set the contexts and boundaries within which a research study can be undertaken.

Primary data are collected in the field at a preliminary stage to provide accurate information on building a sampling frame or setting the field realities. They also form the heart of research because they are the 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 secondary data guides the study.

We further examined how we can collect data from Internet sources that are available for everyone to use. We also explored 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 when doing your research.

Through the discussion, we emphasised the importance of data and data sources in communication research because failure to use proper data sources will affect your research's quality, reliability, and accuracy. Undertaking research is time- and effort-consuming, but there are ways to simplify it. This Unit explained these simple ways of sourcing data.

4.7 FURTHER READINGS

1. Chawla, D. & Sondhi, N. (2011). *Research Methodology: Concepts and Cases*. New Delhi: Vikas Publishing
2. Hamilton, J. (2005). *Primary and Secondary Sources*. ABDO Publishing.
3. Singh, K. J (1978). *Gandhi and Mao as Communicators: A Comparative Study of Practice and Theory*. Unpublished doctoral dissertation. University of Pennsylvania 1978

4.8 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 another person or agency.
2. Availability, relevance, accuracy, and sufficiency are 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, including over 2,500 journals from 57 countries.
3. Two important Inflight features that are 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.

UNIT 5 ONLINE SAMPLING METHODS

Structure

- 5.0 Introduction
- 5.1 Learning Outcomes
- 5.2 Probability and Non-Probability Sampling
- 5.3 Probability sampling-based methods
- 5.4 Non-probability sampling-based methods
 - 5.4.1 Characteristics of an Effective Online Survey
- 5.5 Online sampling for interviews and Focus Group Discussions
- 5.6 Sampling for Social Networks
- 5.7 Mixed-mode surveys using online and traditional media
 - 5.7.1 Advantages of Online Data Collection
 - 5.7.2 Issues and challenges in Internet-based survey sampling
- 5.8 Ethics of Online Research
- 5.9 Let Us Sum Up
- 5.10 Keywords
- 5.11 Further Readings
- 5.12 Check Your Progress: Possible Answers

5.0 INTRODUCTION

Online data collection is an emerging area in social science research. With the proliferation of the Internet and social media tools, researchers are finding it easy to collect data through the Internet. Facebook, Instagram, X (formally Twitter) and other social media tools have millions of users worldwide who can participate in online research. Online sampling refers to identifying and selecting a subset of respondents from a population with the help of web tools for research.

Online Data collection methods have emerged as a standard media, communication, and market research tool. Also known as Internet or web-based research, researchers and statisticians use various online research techniques to gather data from respondents. Social media research is a relatively new addition to this range of online research techniques but is a significant development. This method presents more complexity, creating new opportunities for investigation.

The sampling technique should be appropriate for that research purpose so that statistically relevant claims can be made. Online web surveys might be different from each other, i.e. they should be treated as equal. They can differ based on the following:

1. Sampling technique
2. Sample size

3. Types of questions
4. Mode of implementation (Email, Forms, Social Media)

We can use both probability sampling and non-probability sampling to conduct online surveys. The data can be collected from the web through tools like Google Forms, Microsoft Forms, and emails. As in offline simple random sampling, we also need the sample frame.

This chapter describes specific online sampling techniques that can be used in qualitative and quantitative research. The researcher must identify whether the survey data need to be generalized. If the findings are to be generalized, then the researcher must use a random sampling procedure through online methods. Non-probability procedures are helpful in qualitative studies since these studies require low sample sizes. However, these results can't be generalized over a larger population. For online probability sampling, the researcher needs a sampling frame.

5.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Know the main types of online sampling approaches in survey research;
- Understand how online probability sampling and non-probability sampling differ;
- Understand the sampling techniques used for interviews and focus group discussions; and
- Understand the issues and challenges involved in online sampling.

5.2 PROBABILITY AND NON-PROBABILITY SAMPLING

If the researcher decides to use quantitative and qualitative techniques, then the researcher needs a sampling frame. Random sampling can be done through the use of a sampling frame. Random sampling can also be done without the sampling frame, but statistically, it won't stand firm because the procedure will lack rigour. So, online sampling can be used for a web survey. This online sampling can either a) probability-based sampling or b) Non-probability-based sampling. In probability sampling, every population element has an equal chance of appearing in the sample. It uses a random procedure for selecting the elements from the population and generally uses a sampling frame. The sampling frame is a list of all elements that qualify to be the population for a study. Non-probability sampling, on the other hand, does not use a random procedure for selecting samples. Here, the sampling frame is not required, so it does not take a sample from the whole population.

For example, if the population happens to be a university, then the sampling frame will have a list of all students studying in the university. In the case of a company or an organization, the sampling frame happens to be a list of all the employees. The researcher should try to include all the elements or respondents in the sampling frame. Having said this, the sampling frame is never complete since some students may not be available, may be ill, or

maybe travelling. Inclusion criteria also determine the number of elements in the sampling frame. Once the sampling frame is in place, the researcher can execute any probability sampling technique that suits the purpose of the research study. Some of the methods which the researcher may use are:

- a) Simple random sampling.
- b) Stratified random sampling
- c) Systematic sampling
- d) Cluster Sampling.

The respondents are selected based on convenience or availability using a non-probability sampling technique. Different online non-probability sampling techniques might be used, such as offline techniques. Some of them are;

- a) Availability sampling
- b) Purposive sampling
- c) Quota sampling
- d) Snowball sampling.

Steps in online sampling: The following steps are followed when selecting a sample from a population.

- a. Population Identification: The first step in online sampling is to identify a population of Internet users who can participate based on the research design.
- b. Identify the sample size for the research study.
- c. Use the right sampling technique: Various probability and non-probability methods can be used to select the sample. Some of them are simple random sampling, stratified sampling, cluster sampling, quota sampling, convenience sampling, purposive sampling, and snowball sampling. The technique to be used depends on the research topic, research design, and the characteristics of the respondents.

5.3 PROBABILITY-BASED SAMPLING METHODS

In probability-based sampling surveys, the data can be collected using the following methods:

- a) Through a sampling frame
- b) Random sampling without a sampling frame
- c) Intercept surveys
- d) Pre-recruited panels

The researcher may use online sampling for Probability-based interviews and focus group discussions.

- a) Through a sampling frame
- b) Random sampling without a sampling frame

Let's discuss these methods and understand how they can be used.

1. Survey -- Through a sampling frame

An online or web survey is one of the most commonly used techniques for gathering data. This method sends a set of closed-ended questions to the selected respondents for the study. Individuals can be approached through emails, survey links, email groups or links embedded in the websites. The sample, sample size and response rates can vary in all these different modes of data collection. Online surveys are cost-effective and save time. Since the surveys are conducted through Internet tools, the respondents can fill out the forms at their convenience. These forms are designed so the respondents can easily indicate their replies, which are saved by the software.

Like a normal face-to-face survey, the online mode also seeks information from respondents on a specific topic or phenomenon. Web surveys are conducted through the Internet, and the respondents receive a questionnaire that has been checked for validity and reliability. The way the questionnaire in a web survey is administered is the researcher's choice. He makes this choice after considering quite a few factors. The web survey can be administered through:

- a) Emails
- b) Web forms like Google Forms and Microsoft Forms
- c) Tools like Qualtrics and Survey Monkey.

These tools can be designed to obtain structured and unstructured data. After all the respondents have submitted the forms, the data can be downloaded in CSV or Microsoft Excel format, making the data analysis much easier.

For these surveys based on probability sampling, the researcher needs a sampling frame to start with. The sampling frame has to be corrected and updated. The researcher conducts probability sampling and identifies respondents based on the sample size from the sampling frame. The researcher also needs the contact information, preferably the respondents' email. Let's say the researcher happens to study university students and employees of an organization, or it's a study on the school students. The researcher may approach the university, school, or the company to seek contact details about all the employees. It is important that the researcher keeps this data private and uses it strictly for research purposes only. Once the sampling frame has been finalized, the researcher can apply a probability sampling technique and select the respondents for the study. The selected respondents can be contacted, and the questionnaire can be shared. This list-based sampling method is random, and the results can be generalized to a larger population.

Stratified sampling can also be used in list-based probability sampling. The sampling frame should also have other information to enable the researcher to put respondents in different strata. The total number of people in the population should also be known for systematic sampling. The researcher must ensure the sampling frame has all the details required for cluster sampling. The researchers may choose to follow up the emails with a telephone call to increase coverage and response rate. Systematic and cluster

sampling can also be implemented for web surveys.

Steps involved in an Online Survey:

- a) Outline the research objectives clearly.
- b) Identification of the research design.
- c) Clearly Defined Target Sample
- d) Precisely crafted survey questions.
- e) E. Execution of the sampling process
- f) F. Collection and Analysis of data.
- g) G. Reporting of results.

Check your Progress: 1

Note: 1) Use the space provided below for your answers.
2) Compare your answers with those given at the end of this Unit.

1. What are online probability and non-probability sampling?

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

2. What is sampling error?

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

3. How is list-based probability sampling different from non-list-based probability sampling?

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

2. Random sampling without a sampling frame

This method does not use the sampling frame to select a sample. It's difficult to do online surveys without a sampling frame. Here, the researcher can randomly send emails or links to web forms to Internet users. This may have a high non-response rate since the email IDs might be incorrect. Moreover, the mail recipients may not be eligible for the study.

Another method can be to locate and find email groups on the web. After discovering more about those organizations, the researcher can send his questionnaires to the email groups. The researcher must ensure that the group members qualify for the study. In that case, an accompanying email or some background information about the research project would be good.

Also, the questionnaire should be designed so that people who don't qualify can be eliminated after attempting the first few questions. Data consistency is questionable here as respondents might get biased, and their responses may give incorrect results.

3) Intercept surveys

In this method, every key visitor to a website is included in the sample. So, it can be seen as a form of systematic sampling. The responses are not recorded once the number of respondents per sample size is achieved. One major problem is that every kth visitor may not qualify to be a part of the sample. This person may not qualify in terms of demographics, subject area, or research purpose. This makes intercept survey sampling may be a very time-consuming process. This method is mainly used for consumer satisfaction surveys or by marketing companies who want to better understand their customers. The results from this systematic sampling can be generalized to the larger populations. Here, the population includes visitors who visit this website, regardless of their physical location. To record and specify the location of the visitors, the survey may allow devices with specific IP addresses only based on the region.

Other issues with intercept surveys are the non-response and its bias. It is difficult to identify the reasons for non-response as no information is available about the persons who declined to participate in the intercept survey. There can also be a bias between heavy and light Internet users since heavy Internet users may ignore pop-ups. Bias can also come in since Internet users prefer different web browsers, and some may use pop-up blockers. The code in the browser used for stopping pop-ups may introduce bias in intercept surveys.

4) Pre-recruited panel

In this method, groups of individuals who have already agreed to participate in the research constitute the sample. These groups are selected in advance by using probability sampling techniques. These individuals are selected initially by contacting them by phone, email, or letter. This is a time-consuming process since the researcher must also ensure that the panel members qualify to be included in the research study. Pre-recruited panel surveys are generally used in longitudinal studies. Longitudinal studies justify the effort made to form and recruit panel surveys. Higher participation in the longitudinal studies makes the entire process cost-effective.

Quite a few companies maintain panels drawn using probability sampling. Sub-samples can be drawn from these panels using probability sampling for research purposes. Some of these companies ensure that equipment and Internet access are available to all participants to maintain a panel that is a statistically valid cross-section of the population. In longitudinal panel surveys, there is a chance that over some time, participants may respond differently to surveys than first-time participants. This is called 'panel conditioning'.

5.4 NON-PROBABILITY SAMPLING-BASED METHODS

In non-probability sampling-based surveys, the data can be collected using the following modes;

1. Entertainment polls
2. Surveys using 'harvested' email lists
3. Unrestricted self-selected surveys
4. Opt-in panels
5. Snowball sampling

1) Entertainment polls

Entertainment polls are 'surveys' conducted to know audience reaction to the TV shows & films they watch for entertainment. It captures the likes and dislikes of audiences through the online tools. This method collects data through websites which post open surveys. On the other hand, various TV shows conduct polls where the viewers are asked to vote for their favourite characters and their reasons for liking the TV show. These polls are designed to get an idea about the viewers' choices. TV producers plan contests around their shows and offer prize money to increase participation. Viewers who watch certain shows qualify to participate in the study. The genre of the show and the age category form the inclusion criteria for this method.

2) Surveys using 'harvested' email lists

In this method, email lists are harvested from the posts on the web. These posts can be on social media, emails, and email groups or may be available in user-generated content on websites and blogs. Third-party vendors run crawlers and extract the email from these posts. They then sell these emails to researchers or other marketing companies who want to target specific users based on their demographics, etc. The samples derived from harvested email lists are non-probability samples because they are based on a convenience sample of email addresses. Non-probability sampling techniques for harvested email lists differ from list-based probability samples, where the email addresses are drawn from the sampling frame given by some organizations.

3) Unrestricted self-selected surveys

Anyone can participate in this non-probability survey method. These surveys are posted on websites to facilitate participation in the surveys. These surveys can be promoted through online advertisements, pop-ups, and advertisements in traditional media. These surveys are not rigid about design. These surveys can be designed to be region-specific since they look at the products or services being offered there only. For example, a researcher can conduct a web-based survey to better understand the risks of consuming junk food among residents of Mumbai by posting a link to a survey on their " Food and Health" website. The inclusion criteria permit people of all ages and backgrounds to be participants.

This method uses convenience sampling. As we know, the survey link increases the reach and participation but is not seen as equivalent to online probability sampling. Alvarez et al. (2002) argue that samples drawn through non-probability sampling through the Internet can be used in experimental design where control and treatment groups can be randomly formed from the sample drawn.

4) Opt-in panels

In opt-in panels, respondents are selected based on their availability. The participants choose to participate in this method. The results cannot be generalized here since it is not probability sampling. Selection bias is an issue; for example, those working in tele-calling centres will likely be younger and computer literate. Limiting the respondents to a specific region is difficult in opt-in panels. The participants choose to join the panels and participate in the research.

5) Snowball Sampling

Snowball sampling is a non-probability sampling method where the initial participants give referrals to other participants. So, the initial set of participants helps identify more participants who can be included in the study. In the case of online sampling, the researcher can share the survey link, interview questions or the schedule with the first participant. Then, he can be requested to identify more participants. This method is generally appropriate for studies where the topic of research is rather sensitive, like alcoholism, studies on gays & lesbians, drug, and substance use, etc. The initial set of participants can share phone numbers or emails of other people who can be contacted through email or messaging apps. Through this procedure, the sample size grows bigger. The researcher generally interviews the participants selected through snowball sampling. The researcher may plan online interviews using Zoom, Google Meet, etc. The sample size in snowball sampling is generally kept low offline and online. In the case of snowball sampling, the first respondent may belong to a specific network or may have formed a closed network. For example, a network of hackers can be hidden away from the public. The researcher for this study may contact one of the members and then be told about other members he may contact online and gather data through interviews or telephonic conversations.

5.5 ONLINE SAMPLING FOR INTERVIEWS AND FOCUS GROUP DISCUSSIONS

Interviews and focus groups provide rich data and are used to gather data that the researcher cannot get through surveys. The data is more detailed and subjective and covers different aspects of the research study. The sample size used in qualitative approaches is relatively small compared to the quantitative approach. The sampling technique for interview and focus group discussion can be probability or non-probability. A researcher can also use mixed-mode sampling, where probability and non-probability sampling are combined.

As discussed, the researcher can use a sampling frame to identify participants for an interview or can do without a sampling frame and adopt any technique

under the non-probability method. Researchers can use various web tools like text-only chat, voice chat, or video chat to gather data. The researcher may also use video conferencing tools like MS Teams, zoom, and Google Meet for interviews and FGDs and, with the respondents' permission, can record the interviews to be transcribed later. Researchers must be careful when using technology and adhere to research ethics. In an online focus group discussion, the researcher uses specialized software to conduct a discussion where 8-9 people participate. The participants should be computer literate and have Internet access and a device. As in an interview, the FGD can be recorded with the participants' permission. The participants can be chosen purposefully and can be contacted through an email or phone call beforehand. This process involves non-probability sampling without a sampling frame.

Online research interviews and FGDs have numerous advantages:

- Online Interviews can be conducted from any place without geographical constraints. The cost of conducting online interviews is relatively low.
- The researcher can also record facial expressions and body language using video conferencing. This makes the data much richer and gives a thick description.
- Respondents feel more comfortable discussing sensitive issues through the Internet than face-to-face conversations.

5.6 SAMPLING FOR SOCIAL NETWORKS

The researcher can follow these steps to sample social media platforms like Facebook and X (formally Twitter). It's quite difficult to use list-based online probability sampling in these cases. Intercept surveys, unrestricted surveys, and availability sampling are preferred modes of data collection on social networks. The users' data is stored on servers and kept confidential; only database administrators have access to the records and schema of the database.

The following steps are used in the sampling process of social networks.

- Define the target population:** The population can be based on demographics, research interests, or other variables relevant to the study.
- Choose a sampling method:** The next step is choosing a sampling method that depends on the research topic, objectives, and population characteristics.
- Calculate the sample size:** The number of participants required for the study is calculated.
- Take permission to access the data from social networks:** The researcher obtains consent from the users and requests the social network API to obtain data.
- Analyse data:** After collecting and analysing data using statistical tools,
- Report the findings:** Finally, the researcher presents the findings. He may use visualizations and statistical analysis tools, such as SPSS, R, and Tableau.

5.7 MIXED-MODE SAMPLING FOR DATA COLLECTION

In surveys, interviews, and focus group discussions, the researcher can also use a mixed-mode method for sampling. The researcher can use probability and non-probability sampling online to identify and contact the sample. The operational cost can be higher, but the benefit outweighs the cost since the response rate will go up in mixed-mode surveys. Vehovar and Manfreda (2017) argue that the reasons for combining data from online and traditional surveys are as follows: It has been seen that respondents are more comfortable with offline survey mode than online mode. The effective strategy is to ask respondents to use the online survey link, and then non-respondents can be requested to fill out the survey forms offline. This maintains the population coverage and response rates of a mail survey.

Advantages of Online Data Collection

Online data collection in social science has several benefits, some of them are:

- Wider and faster reach: Gathering a huge volume of data in real-time through the Internet is easier. It also provides faster and wider access to the sample population than traditional techniques.
- Cost-effectiveness: Online data collection technologies save time and are quite cheap. Data collection through the Internet involves no transportation, paper, or printing, and it saves the cost of the data entry operator.
- Data in different formats and user-friendly approach: Data collection through the Internet has higher response rates. Moreover, the Internet facilitates data collection in various formats, both structured and unstructured. The data can be converted to user-friendly formats for quicker access and calculation.

Issues and challenges in Internet-based survey sampling

A frequent problem with large-scale, online surveys is the lack of a sampling frame. It is difficult to find a single list of email addresses for specific populations (government organizations, corporations, etc.), making list-based online sampling difficult. It is also impossible to implement a strategy without using a sampling frame for probability sampling, as the email addresses for an organization or a university for an entire population cannot be created.

Random email addresses might be constructed using the names of individuals and organizations. Many of these constructed addresses may not work. In terms of coverage, it is widely recognized that online surveys using only samples of Internet users do not generalize to the general public. 52% of the Indian population had Internet access in the year 2022; thus, mobile phones are far from complete and vary state by state. So, any sample drawn from the Internet user population will have a coverage error.

Moreover, for online sampling and data collection, the target population must be computer literate and have Internet access. A university or an organization may maintain a list of email addresses of all students studied so far, but that does not mean that every student has access to a computer and the Internet.

5.8 ETHICS OF ONLINE RESEARCH

- a) **Prior permission:** As in traditional research, researchers should obtain consent for online research. Participants should be informed about the research and how their data will be used, as well as the risks and benefits of the study.
- b) **Privacy of respondents:** Researchers must respect participants' privacy, and their personal data should not be disclosed to any other person or organization. The researcher should take all measures to ensure the safety of this electronic data by encryption and preventing unauthorized access.
- c) **Physical or psychological harm:** During the research process, it should be seen that no psychological or physical harm is done to the respondents.
- d) **Data Accuracy:** The research process should ensure the accuracy and consistency of data, and the methods and instruments for data collection should ensure the validity and reliability of data.
- e) **Respect for cultural and social diversity:** Respondents should be explained the data collection process and how that data will be used. The research team should be sensitive to the respondents' cultural norms, ethnicity, and customs. Moreover, the team must be gender sensitive and demonstrate respect for social diversity.

Compliance and acknowledgement: Researchers should use citations and properly acknowledge the sources used to avoid plagiarism.

Check Your Progress: 2

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

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

1. Mention the advantages of interviews conducted through the Internet.

.....

.....

.....

.....

2. What are the challenges of doing online sampling for research?

.....

.....

.....

.....

3. What are intercept surveys?

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

5.9 LET US SUM UP

This Unit looked at sampling techniques using the Internet. Researchers can now contact respondents through emails, web links, and video chats and gather data. The selection process of respondents through a statistical process brings rigour to the study, and results can be generalized to a larger population. Not only this, but the respondents can also even be asked to fill out forms and submit them through the Internet. So, the entire process of identifying the respondents, contacting them, and then requesting them to participate and finally, the data collection can be done with the help of the Internet. The existence of email groups, social groups, social networks, and databases kept by different organizations has come to the rescue of the researcher as he often struggles to get an accurate and credible sampling frame for probability sampling. Sampling for web surveys, online interviews and online focus group discussions can be done using the Internet. Other data collection tools, such as observation and case studies, can also be used online. Here, the researcher must carefully outline the inclusion and exclusion criteria for his research. There are pitfalls in doing online sampling, and issues of coverage bias, non-response, sampling error, low sample size, and other types of bias appear. The researcher must consider all these points and plan his study accordingly to have significant findings.

5.10 KEYWORDS

Sampling error: Sampling error is the difference between a sample statistic and the population parameter estimated by the sample statistic. It can be understood as the difference between the true population value and the observed value from the sample.

Sampling bias: Bias in sampling happens when the sample drawn from the population does not reflect the characteristics of the population. This indicates that the sampling method doesn't adequately represent the population; rather, the sample either over-represents or under-represents certain groups or characteristics present in the population.

Purposive sampling: In this method, the researcher selects a sample that suits the purpose of the study. The researcher has a fairly good idea of the population and draws a sample of the appropriate size to meet the study's objectives.

Cluster Sampling: It is a random sampling procedure in which one cluster is randomly selected from a collection of clusters. Initially, elements are kept in different clusters. The clusters must be individually inclusive and collectively exhaustive. Each component of the sample should be in one and only one cluster.

5.11 FURTHER READING

1. Bhat, A. (2023, October 12). *Online research: Definition, Methods, Types and Execution*. QuestionPro. <https://www.questionpro.com/blog/execute-online-research/>
2. Clements, J. (2023, November 13). Online Research Interviews - Advantages and top tools. *MOS Transcription Company*. <https://www.legaltranscriptionservice.com/blog/online-research-interviews-advantages-and-top-tools/>
3. Kumar, R. (2010). *Research Methodology: A Step-by-Step Guide for Beginners*. United Kingdom: SAGE Publications
4. Menon, A. (2021). Online surveys. *SurveySparrow*. <https://surveysparrow.com/online-surveys/>
5. Sue, V. M., Ritter, L. A. (2011). *Conducting Online Surveys*. United States: SAGE Publications.
6. Vehovar, V., & Manfreda, K. L. (2017). Overview: online surveys. *The SAGE handbook of online research methods*, 143-161.

5.12 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Online probability sampling involves selecting participants from a known population using random sampling methods, ensuring that every member of the population has a chance of being included in the sample. This approach typically requires a sampling frame, such as a list of email addresses or registered users and employs random selection techniques like simple random sampling or stratified sampling to ensure representativeness. In contrast, online non-probability sampling involves selecting participants based on convenience or availability, without ensuring every member of the population has an equal chance of being selected. Common non-probability sampling methods include convenience sampling, snowball sampling, and quota sampling. While probability sampling offers greater generalizability and reduces sampling bias, non-probability sampling is often more practical and cost-effective for online research.
2. Sampling error refers to the discrepancy between the characteristics of a sample and the characteristics of the population from which it is drawn. It arises due to the inherent variability in samples and the fact that samples are not perfect representations of their respective populations. Sampling error can occur due to factors such as sampling method, sample size, and random chance. It is important to acknowledge and quantify sampling error when interpreting research findings to assess the reliability and validity of the results.
3. List-based probability sampling involves selecting participants from a predefined list or sampling frame, such as a directory of email addresses,

registered users, or customer databases. This approach ensures that every member of the population has a known and equal chance of being selected, thereby enhancing representativeness, and reducing sampling bias. In contrast, non-list-based probability sampling does not rely on a predetermined list but instead utilizes random sampling techniques to select participants from the population of interest. This may involve using random number generators or other random selection methods to identify potential participants from a broader population pool. While both approaches adhere to probability sampling principles, list-based sampling offers greater control and precision in participant selection, particularly when a comprehensive sampling frame is available, whereas non-list-based sampling may be more suitable when a sampling frame is not readily accessible or when random selection is preferred.

Check Your Progress: 2

1. Internet interviews offer unparalleled accessibility and convenience, breaking down geographical barriers and allowing researchers to reach participants from diverse locations without the constraints of physical presence. This not only broadens the pool of potential respondents but also facilitates flexible scheduling, accommodating participants' preferences and availability. Moreover, Internet interviews are cost-effective, eliminating expenses associated with travel, accommodation, and venue rentals, thus making them a more budget-friendly option, particularly for researchers with limited resources. The anonymity provided by online platforms encourages participants to share candid and honest responses, enhancing the quality of data collected. Additionally, the incorporation of multimedia elements enriches the data and improves accuracy, while real-time or asynchronous interview options enhance efficiency, enabling rapid data collection. Furthermore, Internet interviews contribute to sustainability efforts by reducing carbon emissions and environmental footprints associated with traditional research methods.
2. Online sampling for research poses several challenges that researchers must navigate. Sampling bias is a prevalent concern, as certain demographics are more likely to participate in Internet-based research, leading to skewed results that may not accurately represent the target population. Furthermore, the digital divide exacerbates inequalities, excluding individuals from socio-economic backgrounds or geographic regions with limited Internet access or digital literacy. Privacy and data security issues also arise, necessitating robust measures to protect participants' personal information and ensure compliance with data protection regulations. Moreover, variability in response rates and technical difficulties such as Internet connectivity issues or device compatibility problems can compromise the representativeness and reliability of the data collected. Additionally, verifying the authenticity of respondents in online surveys presents challenges not encountered in face-to-face interviews, raising concerns about the integrity of the data.

3. Intercept surveys, conducted in public places by approaching individuals directly, offer unique advantages and challenges. These surveys provide spontaneous engagement, capturing immediate reactions and opinions from respondents in real-time. With higher response rates compared to other survey methods, intercept surveys yield valuable insights into public sentiment and behaviour. Moreover, conducting surveys in the context where decisions occur enhances understanding and relevance. However, intercept surveys also present logistical challenges, including identifying suitable locations, obtaining permissions, and managing interviewer training and supervision. Despite these challenges, intercept surveys remain a valuable tool for gathering on-the-spot data and gaining insights into public opinions and behaviours.





ignou
THE PEOPLE'S
UNIVERSITY