

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.

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

			online research ethics (Ess, 2002).	Internet Mediated Research
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.

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 predicated 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/DigitalDevicesDistCopyJSep t2013.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 straightforwardly, 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 hypotheses 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.

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 Inflibnet.

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.communit.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 Inflibnet 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.

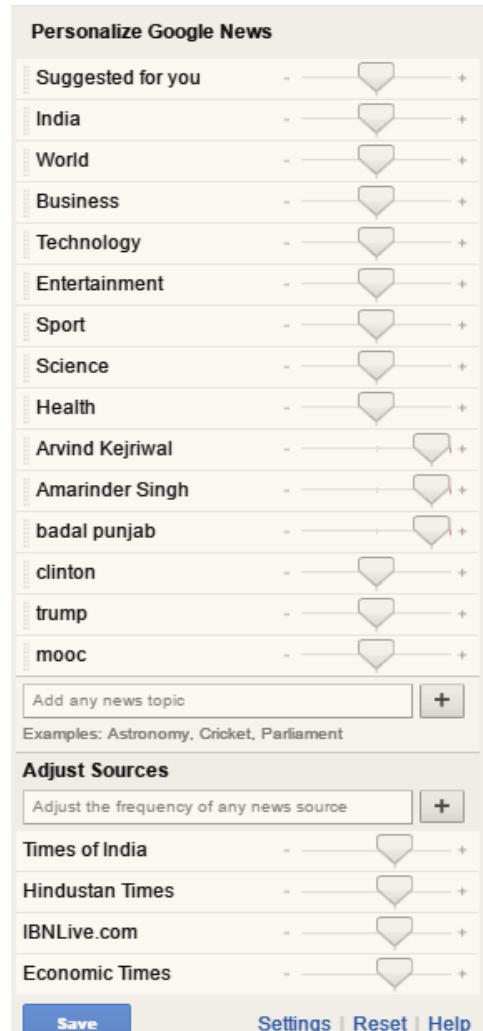
www.en.wikipedia.org/wiki/Wikipedia:List_of_online_newspaper_archives—It has a list of online newspaper archives of newspapers from different countries, and the links are hosted here for the third-party websites.

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](https://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 screenshot shows the Google News homepage at <https://news.google.com>. On the left, there's a sidebar with 'Top Stories' (Arsenal F.C., Rodrigo Duterte, Donald Trump, Manchester, French Open, Indianapolis 500, North Korea, Sebastian Vettel, Emmanuel Macron, FC Barcelona, Hisar, Haryana) and 'Suggested for you' (India, World, Business, Technology, Entertainment, Sport, Science). The main area features a search bar with placeholder text: 'Find news stories that have all these words:' followed by four input fields. Below these are sections for 'this exact phrase:', 'at least one of these words:', and 'none of these words:'. There are dropdown menus for 'occuring' (set to 'anywhere in the article') and 'Date added' (set to 'recent'). A note says 'Looking for scanned newspapers? Learn more'. Below these are date range inputs ('between M/d/y and M/d/y'), 'Source' (e.g. CNN, New York Times), and 'Location' (e.g. Florida, California, India). To the right, a news article about India's ties with Russia is displayed, along with related links to Narendra Modi and Mann Ki Baat, and images of Prime Minister Narendra Modi and other political figures. The interface is clean with a light blue and white color scheme.

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.

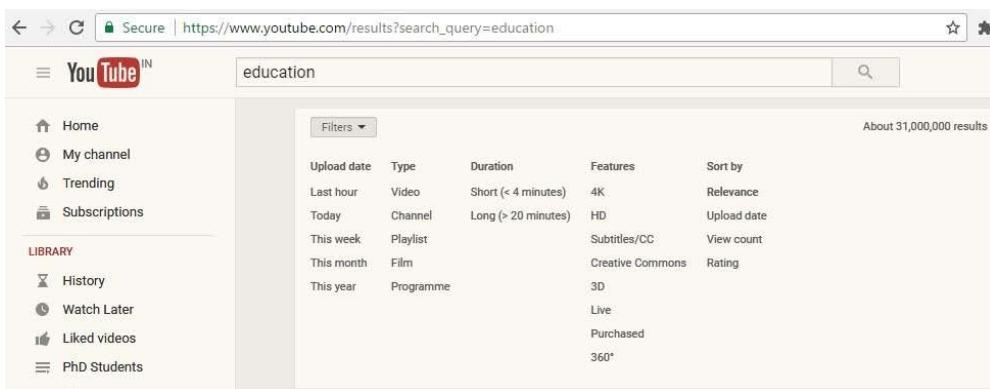
All newspapers have their data live data as archived editions open to data retrieval (also known as data scraping). A few of them have the option of "advanced search" on their sites. The New York Times has had digital copies of all newspapers since 1851 at www.query.nytimes.com, and it allows searching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

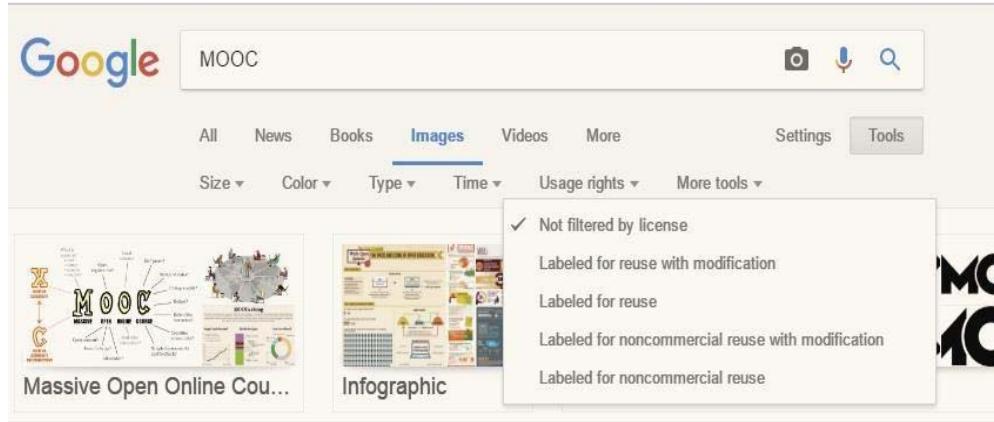


YouTube is a wide repository of visual and textual data. There are a few less explored functions in YouTube, like filters (to filter the videos as per date, type, duration, view count, rating, licenses, view quality, most liked videos etc.). There are various playlists (collections of videos) on a topic 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 Inflibnet 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 k th 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.

- A. **Define the target population:** The population can be based on demographics, research interests, or other variables relevant to the study.
- B. **Choose a sampling method:** The next step is choosing a sampling method that depends on the research topic, objectives, and population characteristics.
- C. **Calculate the sample size:** The number of participants required for the study is calculated.
- D. **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.
- E. **Analyse data:** After collecting and analysing data using statistical tools,
- F. **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.

Block

2

INTERNET RESEARCH METHODS - I

UNIT 6

Interview Method

UNIT 7

Experimental Research Method

UNIT 8

Visual Analysis

UNIT 9

Social Network Analysis

BLOCK 2 INTRODUCTION

After learning the basics of Internet research in the previous Block, in Block 2, you will learn the various data collection methods. Each method is unique in this Block; you will learn two mostly used qualitative data collection tools (Interview Techniques and Visual Analysis), one classic example of a quantitative tool—experimental Research—and a Unit on social media data collection tools.

Unit 6 Interview Techniques, explains the diverse world of interviews, exploring concepts, types, and ethical considerations. From informal to structured interviews, including telephonic and online surveys, we'll navigate through semi-structured and unstructured formats, emphasising interviewing skills and ethical guidelines.

Unit 7 Experimental Research Methods, explains the intricate nature of experimental methods. We'll explore classic research designs and the experimental process, including controlling extraneous variables and randomisation. Additionally, we'll discuss experimental design variations and field experiments and analyse the merits and demerits of this method.

Unit 8 Visual Analysis, where we explore the intricacies of perception and interpretation in images. Delve into topics such as Visuality, Iconology, Formalism, and Simulation. It uncovers the diverse methods of analysis, from Semiotics to Discourse Analysis, and investigates the scope of inquiry spanning historical, contextual, and personal dimensions.

Unit 9 is Social Media Analysis, where we learn about the intricate web of digital interactions. We explore Social Networks and Social Network Analysis, unravelling the essence of connectivity. We will learn about data management and understand its nuances from collection to governance. We discover Network Theories in Cultural Anthropology, unveiling the dynamics of societal structures. Join us in deciphering the digital realm.

These Units will give you a broader understanding of data collection from the online platforms, and you will be able to visualise your intended enquiries appropriately.

UNIT 6 INTERVIEW METHOD

Structure

- 6.0 Introduction
 - 6.1 Learning Outcomes
 - 6.2 Interview: Concept and Types
 - 6.3 Informal Interviews
 - 6.4 Structured Interviews
 - 6.4.1 Telephonic Interviews
 - 6.4.2 Online Surveys
 - 6.5 Semi-structured Interviews
 - 6.5.1 Planning Focus Group Discussions
 - 6.5.2 Process of Focus Group Discussions
 - 6.6 Unstructured (In-depth) Interviews
 - 6.7 Interviewing Skills
 - 6.8 Ethical Issues
 - 6.9 Let Us Sum Up
 - 6.10 Further Readings
 - 6.11 Check Your Progress: Possible Answers
-

6.0 INTRODUCTION

The interview is one of the most commonly used fundamental research techniques. The interview is a 'focused, purposive conversation' between the researcher and an informant - it involves asking questions, listening to individuals and recording their responses. You would have watched interviews with political leaders, celebrities, sportspeople, and even common people on television channels. Interviews are also conducted for jobs. The broad purpose of any interview is to obtain information. However, when we refer to 'interview as a research technique', its objective and focus are much more systematic and scientific in nature and approach, hence the need to understand the technique. In this Unit, we shall focus on interview techniques, which enable a researcher to obtain information that cannot be gained through other research methods.

6.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Discuss Interview as a tool of data collection;
- Describe different types of interviews - Informal, Structured, Semi-structured and Unstructured Interviews;
- Examine the strengths, limitations and applications of each type of Interview and

- Apply interview techniques in your research study.

6.2 INTERVIEW: CONCEPT AND TYPES

The term interview draws upon the French term '*entreuve*', which means 'to see one another or meet'. One of the important purposes of research is to get information about people. Many research methods and techniques are used for this purpose. These include methods such as observation. While the observation method involves the study of a phenomenon in natural settings and uncovers the ongoing and present activities of the respondent, it does not explore the past and the opinions, beliefs and motivations of a person. To get that information, it is imperative that the 'right questions' are asked; the interview technique allows us to do so. Interviews aim to explore why people behave in a particular way and what they think about something. It usually involves face-to-face interaction; however, technology allows interviews to be conducted online on social networking sites and telephones. Before proceeding further, let us first look at the different types of interviews.

Interviews have been classified as structured, unstructured, formal, informal, and so on. For the purpose of our discussion, we shall follow the classification given by Berger (2000), as all these types of interviews are extensively used in communication research:

- Informal Interviews
- Structured Interviews
- Semi-structured Interviews
- Unstructured (In-depth) Interviews

The focus of our discussion shall be on qualitative interviews—semi-structured and unstructured interviews—while the remaining two types shall be briefly touched upon to place them in a context.

6.3 INFORMAL INTERVIEWS

Informal interviews, also known as 'informal conversational interviews,' can take place anywhere. As the name suggests, these are not organised systematically or focused on specific questions. The informal nature of the method allows the information to build naturally. Informal interviews are commonly used in field research to gain the subject's confidence. Informal interviews are generally one-on-one and generate highly individualised responses, which may vary from one person to another. Informal interviews have been found helpful in developing a big picture by obtaining and clarifying information. The individualised nature has helped obtain information on sensitive issues and can add depth to data obtained by other methods, such as participant observation. The interaction can be held even after intervals. In such types of interviews, especially those conducted in rural areas, it is possible that the subject may not be fully aware of the information being probed.

The technique is mainly proper at an individual level. It is less applicable to groups as it may not be possible to clearly understand what is being said in a group situation. Like any other data collection tool, it is necessary to establish a good rapport with the respondent to open up; hence, it may take more time than other forms of interviews. It also requires specific skills and care in asking questions, such as avoiding leading questions. The information obtained is random, so it may take a long or more than one session to cover all areas.

Moreover, data from one respondent may not be comparable with other respondents as the nature of questions and the sequence is likely to vary from one person to another. Given the informal nature of the method, subjectivity may also creep in recording responses. According to Bertrand & Hughes (2005), if the primary concern of the research is to obtain individual descriptive responses, for instance, building upon observational data from fieldwork - these problems may not matter. However, if you wish to use your interviewees as a sample of a larger population, you will need to choose a more structured interview.

Activity-2

Conduct an informal interview on a topic of your choice for five to ten minutes. Identify the pros and cons of undertaking such an Interview and list the issues involved.

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 some strengths of Informal Interviews.

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

2. List some limitations of Informal Interviews.

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

6.4 STRUCTURED INTERVIEWS

Structured interviews are helpful in obtaining information about an issue's prevalence and distribution from many people. We discussed that questionnaires or interview schedules are designed in advance so that questions are carefully worded and delivered identically to all the respondents. The resulting data are quantitative and comparable, hence easier

to analyse. Such interviews can be completed quickly in the field and thus are relatively economical in terms of cost and time.

We also discussed that greater skill is required while designing questionnaires or interview schedules, which must be pre-tested. Highly structured questionnaires yield little insight into how people feel about the issues involved, and the interviewer may not be able to respond to a valuable issue/situation that does not appear on the schedule and emerges during an interview. In a larger study, individual differences among interviewers may affect data quality. At times, it may be difficult to find those selected in the sample, while some people may be reluctant to answer some questions due to lack of time, disinterest, language barrier or sensitivity. Structured Interviews are also conducted using telephones and the Internet.

6.4.1 Telephonic Interviews

Telephonic interviews are relatively cheaper and easier to conduct. In such interviews, closed-ended questions are generally included as these are easier to administer to obtain feedback on an issue under study. Telephonic interviews have been found more useful in journalistic research, exploring a few focused questions. However, these have severe limitations as a data collection tool for academic research. Yet another limitation of a telephonic interview is that in the absence of a comprehensive mobile phone database, there is the possibility that some segments of the population may be systematically missed out. About a decade ago, access to telephones was limited to the affluent sections, but the spread of mobile phones has facilitated access to a diverse section of the population. In earlier times, the telephone directory published by the telephone departments in different cities used to provide address details and the location of the person listed, which is not available any more. Thus, it is possible that the sample may not be representative, leading to a skew in findings. Moreover, those contacted over the phone may not be interested in providing feedback, so they discontinue the interview midway or give superficial answers, thus affecting the data quality.

6.4.2 Online Surveys

Interviews are also conducted online by emailing questionnaires and uploading website links. The links take the respondents to different web pages listing the closed-ended questions. The respondent is expected to tick the appropriate response, and when all the questions on a page are responded to, s/he has to proceed to the next pages till all questions have been responded. An example of an online survey from different stakeholders to obtain feedback on the New Education Policy on Open and Distance Learning and Online courses is given below. In this survey, three questionnaires were uploaded on the IGNOU website requesting feedback from practitioners/experts/ learners and employers with a brief introduction and links:

New Education Policy: Survey on Open and Distance Learning and Online Courses

The Ministry of HRD has entrusted the responsibility of developing a Draft Policy on Open and Distance Learning and Online Courses to IGNOU. In this connection, you are requested to provide us with your valuable feedback in the questionnaire relevant to you, available at the below link:

[Click here to participate in the Practitioners and Experts survey](#)

[Click here for the Learner's survey](#)

[Click here for the Employer survey](#)

A similar process is followed while using an email listserv in which the link to the questionnaire is emailed requesting feedback. It needs to be mentioned that in an online survey, the number of questions is generally kept small - not exceeding 25 to 30 as many a time, the respondent may lose interest midway, leading to a high rate of incomplete responses; thus, data loss. Further, using mobile phones and the Internet may reduce the cost of conducting surveys, but technology also introduces its own biases. For example, despite its increased outreach, access to technology is still limited in rural areas and certain sections of society, including women. In addition, linguistic and socio-cultural barriers may also affect data quality. Finally, response rates in online surveys are very low; as a result, data from online surveys can be taken only as indicative and not as definitive in terms of results.

Activity-2

Prepare an Interview Schedule comprising 15-20 questions to study the social media usage patterns of the youth in your area. You may draw upon the do's and do n'ts of designing an Interview schedule.

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 uses of telephonic surveys?

.....

2. What are the limitations of online surveys?

.....

6.5 SEMI-STRUCTURED INTERVIEWS

Semi-structured interviews, as the name suggests, are neither structured (such as a field survey) nor freewheeling (like unstructured interviews) but placed somewhere in between. The roots of semi-structured interviews are in Ethnographic Research, and Focus Group Discussion (FGD) comes under this category, which has the inherent strength of collecting qualitative data from several people in the quickest possible time. In recent years, FGDs have been used precisely to find out how people respond in a group and how their feelings and opinions can be shaped by the experience of discussing the subject with others (Bertrand & Hughes, 2005).

FGD is considered a user-friendly and 'non-threatening research method' that participants find stimulating and enjoyable. It has been found useful in assessing people's reactions to new products, services, messages, problems or ideas in a participatory mode. The sample size can be increased without increasing time and costs, which is relatively economical. It invites participants to monitor each other, providing checks and balances that do not operate for individual interviews or surveys (Bertrand & Hughes, 2005). FGDs also help people express themselves openly about sensitive issues and bridge social and cultural differences (Morgan, 1998). FGDs can provide a forum for participants to express themselves. For instance, feminist researchers use FGDs to "provide women with safe space to talk about their own lives and struggles" (Kamberelis & Dimitriadis in Brennen, 2013).

6.5.1 Planning Focus Group Discussions

Focus group discussions need to be carefully planned and executed. The researcher outlines a basic checklist of the areas/topics to be covered and who will attend the discussions. A group comprising 6-9 members is selected purposively to elicit data on opinions, beliefs, and experiences. FGDs are free-form discussions by a group of people, led by a moderator, designed to obtain information about some topic. To the greatest extent possible, the researcher tries to maintain the quality of the casual interviews found in unstructured interviews (Berger, 2000). Focus groups can be exploratory, and exercising less control over the groups can lead to wide-ranging discussions. The participants respond to the moderator and one another, and the conversations and reactions closely approximate normal conditions. It also gives freedom to the researcher to respond to any situation that emerges during the interview.

However, FGD should not be considered as a simple way to get a lot of interviews in a hurry - they are 'focused efforts at gathering data'. Merely gathering people together does not guarantee a meaningful discussion, and there must be an effort to gather research data through such focused discussions. Some scholars argue that, like any other research method, conducting FGD in combination with other qualitative or quantitative research methods is useful. Often, focus groups are used as an exploratory technique and followed up with surveys to generalise results for larger populations. Morgan (1998), however, argues that for many purposes, the strengths of focus groups will be entirely sufficient and can be used for

varied purposes such as problem identification, planning, implementation and assessment.

6.5.2 Process of Focus Group Discussions

To conduct the FGD, you may start by introducing yourself and team members to establish rapport and explain the purpose of the session. You should share the need for audio/video recording or note-taking, mention the need for a checklist and explain its purpose. You must also obtain the group's informed consent for any such recording and ensure their participation is voluntary. Start the discussion with a focus question and invite responses; if you do not get any, address someone and keep the conversation going. Monitor the discussion and keep track of what needs to be covered. As a researcher, it is important for you to listen carefully - this will enable you to respond to new ideas and encourage those with something interesting to say. You should remain flexible if a useful debate is happening, but if the discussion goes off track, interrupt politely but firmly. You should refrain from giving your opinions or trying to reach a consensus, as that is not the purpose of the FGD. You must also make eye contact to involve everyone and stay in charge without being obtrusive. The discussion should keep moving and on track in light of research objectives. Time needs to be carefully monitored to elicit desired information.

Limitations: FGDs are time-consuming and expensive. Selecting a representative sample is difficult, as people may refuse for various reasons. They are not easy to conduct effectively without proper training. The discussion has to be kept moving, and it should not falter at any stage. The data obtained from FGDs is not comparable, and complex analytical and interpretative techniques are required.

Activity-3

Prepare a list of areas you wish to explore using FGD as a tool for data collection. Sequence them logically and frame them in the form of open-ended questions. As discussed above, administer these questions to a group of 6-8 participants and record the experience.

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 some limitations of FGD.

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

2. Describe the role of a moderator for the success of FGD.

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

6.6 UNSTRUCTURED (IN-DEPTH) INTERVIEWS

An unstructured interview is an 'open-ended conversational exploration of an individual's worldview and gives an insider's perspective'. In-depth interviews, also known as "intensive interviews", fall in this category. It is a key ethnographic data collection technique that is useful for a holistic inductive approach. The method is useful for collecting complex information on the respondents' opinions, beliefs, attitudes and personal experiences. In-depth Interviews collect qualitative data and are generally used to supplement the information collected through survey methods or participant observation. It does not follow a rigidly set structure - some basic questions are outlined, and the discussion is largely free-ranging. In-depth interviews elicit verbal responses, which are richer than written ones. These provide useful inputs/feedback for gaining a meaningful understanding of what might be a very different perspective.

The sample size of an in-depth interview is generally kept small as a select few purposively selected people are subjected to a detailed interview, which can be scheduled at a mutually convenient time and place. Compared to the survey method, in-depth interviews allow the researcher flexibility to ask follow-up questions and seek clarification on an issue. The topics of in-depth interviews are generally in the form of tentative questions that allow the researcher to cover all the topics and probe further.

In studies involving a large number of field investigators for data collection using in-depth interviews, you will have to orient and train the interviewers to create a common understanding and approach to the study. In such studies, it is useful to exchange notes with other interviewers at the end of the day/week to make necessary changes in approach based on their experience.

Limitations: In-depth interviews, like any other form of data collection, also have inherent limitations. These can be conducted on a limited number of respondents and can be time-consuming. The researcher has little control over the responses as respondents tend to speak more about issues concerning them than the subject under study. Less caution can generate enormous data, which may be later difficult to organise and analyse, hence requiring more effort and skills on the researcher's part. Interviewing several people every day could be a tiring experience and may cause fatigue. The tendency to get personally involved with the interviewees may lead to the risk of introducing bias in the results as a researcher may impose his/her own perspective.

It is possible that respondents may not be able to properly articulate what they actually think or believe at times. In such situations, the researcher will have to ask probing questions to seek clarity. Some respondents may give

'politically correct' statements for various reasons, and it will depend upon your skills as an interviewer to adequately examine their responses and arrive at findings.

The purposive sampling technique is generally used for in-depth interviews, which involves choosing the region and the respondents for a specific purpose. In this regard, refer to the points discussed in Unit 4, Block-I on Sampling Methods.

Activity-4

Do some background research before undertaking in-depth interviews with block-level officers on a development project of your choice. List some key areas for exploration and formulate questions about them.

Check Your Progress: 4

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

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

1. List some limitations of in-depth interviews.

.....

2. What steps will you take to plan an in-depth interview?

.....

6.7 INTERVIEWING SKILLS

From the above discussion, it is clear that despite being a common data collection tool, interviews are difficult to conduct and require skills and a certain amount of tact. The following points are common to all the interviews discussed above and need to be kept in view while conducting Interviews:

- You need to build rapport with respondents by making some introductory remarks about the purpose of the study and seeking their cooperation. Ensure you take informed consent that their participation is voluntary and they are willing to cooperate.
- For better results, the interview may be preceded by a more general information gathering and observation exercise to identify critical areas for pursuing in greater depth.

- You need to ask clear, neutral, and unambiguous questions. Loaded or leading questions should be avoided, which will elicit leading answers (We have discussed different types of questions - leading, ambiguous, double barrel, etc. in Unit 7 of Block 2).
- You need to stay focused, and the interview should keep moving in a direction consistent with the objectives of the study. If the discussion loses focus, the researcher should bring the discussion back to the main point.
- If you do not understand something said by the respondent during the Interview, ask supplementary questions and explore the issue further to seek clarity.
- As a researcher, it is extremely important for you to be a good listener and not give your own views on an issue. The Interview is *not* an exchange of information but *obtaining* information from the respondent.
- You need to maintain neutrality and objectivity throughout the process of data collection and remain non-judgmental. Any show of signs through facial expressions, body language, etc., will likely impact the responses.
- You should maintain a polite and formal interaction and be sensitive to the subject's entirely different perspective.
- You should refrain from offering unsolicited advice and false promises, which may put you in an awkward position at a later stage.
- Finally, you should check that all questions are covered, thank the respondents, and ask them if they have any questions.

Recording Medium

The recording medium needs to be carefully selected to ensure the accuracy of the information collected during the interview. These generally include handwritten notes or audio/video recordings. While taking notes, you must be extremely careful when listening and noting down the responses. It is often impossible to note down everything in the presence of the respondents. A good researcher goes back in time and tries to recall the entire sequence of events once the interview is over. In larger studies involving many interviewers, debriefing sessions are organised regularly for the exchange of information. If audio recording has to be conducted in rural areas, an adequate battery supply must be carried out as there may be a power failure. Video recording has been found useful for certain studies. However, the presence of a camera can make the subjects conscious, and they may put on an act or give politically correct answers. Please remember that informed consent for audio or video recording is mandatory.

Transcription

Once the Interview has been conducted and recorded, the entire conversation must be transcribed, which is time-consuming and tedious. Sometimes, it may take 5-6 hours to transcribe an hour-long interview. After transcription, the entire material has to be checked for accuracy and to make sense of the entire conversation. Then, the data has to be classified and coded under certain themes and categories using appropriate qualitative data analysis

techniques such as the constant comparative technique, the analytical induction strategy, etc. According to Berger, there are no absolute rules about coding, as a great deal depends on the nature of the material being coded. You will read more about qualitative data analysis in subsections 15.5 and 15.6 of Unit 15 of Block 4.

6.8 ETHICAL ISSUES

The purpose of research is knowledge production and generation. However, as researchers, you are not expected to adopt any means to obtain information; hence, some ethical norms will be in order while using Interviews as a research tool. Some of these ethical issues have been discussed in previous units, such as the participation of people in Interviews should be voluntary, and nobody should be compelled to participate. They need to be informed about the nature and objectives of the study and how the findings will be used. Further, they can opt out if they do not feel comfortable at any stage during the Interview. Thus, 'informed consent' forms the basis of selecting respondents, and if confidentiality has been promised, you must uphold it. You must treat all your research subjects respectfully and protect them from invasion of privacy and any physical, financial or emotional harm. Some subjects may not be capable of fully understanding the implications of your study and thus may be more vulnerable or sensitive for a particular subject. For example, when selecting children as subjects, you need to get the approval and cooperation of parents. Parents and legal guardians must be present when children are being interviewed.

The recruitment of respondents is yet another ethical area, especially in FGD. Including respondents known to a researcher who does not meet the criteria for selection will lead to a skew in findings. Payment for participation is generally made in market research to cover the travel expenses and the time used. However, academic research has ethical implications as it may be argued that the study's findings have been compromised. Like other research methods, ethics are involved in data analysis and the reporting of interviews. According to Wimmer and Dominick (2003), data should not be tampered with, and the responses should not be fabricated or changed. Further, science is a public activity, and researchers have an ethical obligation to share their findings and methods with other researchers who may wish to examine the research instruments, methodology, sample selection, and other relevant items and apply them in their own research work.

Activity-5

You have to collect data using interview methods from rural women on their reproductive health. Discuss some ethical issues involved.

Check Your Progress: 5

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

1. List some ethical issues that need to be considered while doing research.

.....

.....

.....

.....

.....

.....

.....

2. Why is publishing considered to be an obligation for the researcher?

.....

.....

.....

.....

.....

.....

6.9 LET US SUM UP

In this Unit, we discussed interviews as a research method under the broad classification of informal, structured, semi-structured, and unstructured interviews. The strengths and limitations of each type of Interview were thoroughly discussed and analysed. The application, interview process, and care to be taken were described in detail. We also looked at some guidelines for conducting interviews and emphasised the need to follow certain ethical norms largely pertaining to minimising harm, respecting autonomy, protecting privacy, and others. As a researcher, you must consciously and systematically use the interview method to generate new knowledge and insights through your study.

6.10 FURTHER READINGS

1. Berger, ArtherAsa (2000) Media and Communication Research Methods, Sage Publications, London, New Delhi
2. Bertrand Ina & Peter Hughes (2005). Media Research Methods: Audiences, Institutions, Texts, Palgrave MacMillan, New York.
3. Brennen Bonnie S. (2013). Qualitative Research Methods for Media Studies, Routledge, New York
4. Krueger Richard A. (1998). Analysing and Reporting Focus Group Results, Sage Publications, London
5. Morgan David L. (1998). Planning Focus Groups, Sage Publications, London
6. Priest S. H (2010). Doing Media Research: An Introduction, Sage Publications, New Delhi
7. Wimmer R.D. & Joseph R. Dominick (2003). Mass Media Research: An Introduction, 4th edition, California, Wadsworth Series in Mass Communication.

6.11 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Informal interviews are useful in:
 - Gaining confidence in the subject in the initial stages of field research
 - Obtaining information on sensitive issues
 - Adding depth to data obtained by other research methods.
2. Some limitations of informal interviews are:
 - The information obtained is not systematic
 - Data from one respondent may not be comparable with other respondents
 - It is not effective for interviewing a large section of the population.

Check Your Progress: 2

1. Telephonic interviews are:
 - Relatively cheaper and easier to conduct
 - Select closed-ended questions can be easily administered to respondents
 - Useful in journalistic research exploring a focused area.
2. Online surveys have the following limitations:
 - Access to technology is limited in rural areas and cannot cover all sections of society.
 - If a representative sample is not included, the findings of a study will be skewed.
 - Linguistic and socio-cultural barriers may affect the response rate and data quality.

Check Your Progress: 3

1. Some limitations of FGDs are:
 - Not easy to conduct without proper training
 - Difficult to select a representative sample who meets the specific criteria
 - The data obtained is not comparable and requires complex analytical techniques.
2. The moderator's role is crucial for the success of FGD. He or she has to facilitate the discussion and keep it on track in light of research objectives. He or she has to respond to new ideas and encourage those with something interesting to say. He or she has to involve everyone in the discussion and stay in charge without being obtrusive.

Check Your Progress: 4

1. Some limitations of in-depth interviews are:
 - It requires more effort and skills on the part of the researcher
 - The researcher may impose his/her own perspective and introduce bias in the results
 - Interviewing several people every day may cause fatigue
 - Over-enthusiasm can generate enormous amounts of data that may be difficult to analyse later.
2. While planning an in-depth interview, I shall take the following steps:
 - Outline some basic questions in light of the study objectives
 - Identify the respondents for the Interview
 - Outline the purpose of my study
 - Persuade them to participate in the study
 - Follow up with them, as it may take several attempts
 - Schedule the interviews at a convenient place and time.

Check Your Progress: 5

1. The following ethical issues need to be kept in view while doing research:
 - Informed consent and voluntary participation of the respondents
 - Providing information to them about the research objectives and process
 - Protecting them from invasion of privacy
 - Protecting them from any form of physical, financial or emotional harm.
2. Publishing scientific results is an ethical obligation on the researcher. As science is a public activity, researchers need to share their methods and findings with other researchers who may wish to examine the research process followed and apply it in their own work. The knowledge thus accrued is shared with other scholars, enabling society at large to benefit from the results.

UNIT 7 EXPERIMENTAL RESEARCH METHOD

Structure

- 7.0 Introduction
 - 7.1 Learning Outcomes
 - 7.2 Nature of Experimental Method
 - 7.3 Classic Experimental Research Design
 - 7.4 Process of Experimental Research
 - 7.4.1 Control of Extraneous Variables
 - 7.4.2 Randomisation
 - 7.5 Experimental Design
 - 7.5.1 Factorial studies
 - 7.5.2 Quasi experimental design
 - 7.6 Field Experiments
 - 7.7 Merits and Demerits of Experimental Method
 - 7.8 Let Us Sum Up
 - 7.9 Keywords
 - 7.10 Further Readings
 - 7.11 Check Your Progress: Possible Answers
-

7.0 INTRODUCTION

Communication research has established itself firmly and is a fertile field for young researchers. The area is rich, vast, and varied. The experimental method is a well-established media research method, but its use is restricted. There is a tendency among researchers to opt for the survey method. Experimental research is highly popular in psychology and social psychology. In the case of media research, the experimental method can produce an enormous amount of data. However, media researchers find several issues in conducting experiments. For instance, the method is largely quantitative and uses extensive statistical procedures for analysis. Issues relating to infrastructure, i.e. getting the appropriate testing conditions and/or getting matching subjects for experiments, pose problems for researchers. There can be laboratory experiments or field experiments. The purpose is to find out and establish a causal relationship between an independent variable and a dependent variable. This method is also amenable to statistical analysis depending on the data generated.

We have discussed earlier that the selection of a research method depends upon the topic and the researcher's methodological approach to it. The experimental technique supports an empirical approach to research.

7.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Discuss the nature and method of experimental research;
- Describe the process of classic experimental research design;
- Differentiate between experimental and quasi-experimental designs;
- Explain the nature and scope of field experiments; and
- Outline the advantages and disadvantages of the experimental method of research.

7.2 NATURE OF EXPERIMENTAL METHOD

As already stated, the experimental method is widely used in natural sciences and certain social science disciplines. Being quantitative, experiments attempt to determine the cause-and-effect relationship between factors. In other words, they help establish and find out cause and effect. Of course, there are differing views. Some scholars think finding a cause-and-effect relationship between two variables, independent and dependent, is not easy. Nonetheless, the experimental method is one of the effective methods in which the researcher can control the time order when presenting two variables. This enables him/her to ensure that the cause happens first and then the effect and s/he can also control other possible causes of the variable being studied.

The experimental method is scientific in nature and character. As science makes progress, its forms of explanations also keep changing. Since social sciences adopt scientific methods of empiricism, their explanations of social phenomena also keep changing. Explanations are categorised into deductive and probabilistic. This classification is based on the types of generalisations an explanation employs. Let us take the deductive explanation for evaluation. The pre-conditions of a deductive explanation are;

- 1) A universal generalisation.
- 2) A statement of conditions which generalises appear true.
- 3) An event to be explained (explicandum), and
- 4) The rules of logic.

The law of gravitation is a good example of a deductive explanation. Its applicability encompasses past, present, and future. In natural sciences, deductive explanations are made possible with manipulation and control through experimentation. Experiments help us demonstrate the effect of one phenomenon on another phenomenon. However, only a few phenomena of interest to social scientists are susceptible to manipulation and control. Yet, certain ethical questions relate to social scientists' manipulation of social phenomena. Besides, most human factors are not amenable to natural scientists' research methods.

Social scientists mostly prefer to have probabilistic explanations. A generalisation is supposed to provide evidence for the phenomena under investigation. No explanation can be deduced from the generalisation. In universal generalisation, we may say 'All A is B', whereas a probabilistic generalisation will say that 'n per cent of A and B'. Probabilistic explanations are weak compared to deductive explanations. Probabilistic explanations have a minimal predictive capacity. The deductive explanation is strong in its predictive power.

7.3 CLASSIC EXPERIMENTAL RESEARCH DESIGN

A classic experimental research design contains two groups for comparison. One is an experimental group, and the other is a control group. These two groups are equal in every respect. However, there is one difference. The experimental group is administered or exposed to an independent variable called the experimental stimulus, and the control group is not exposed to the experimental stimulus. Measurements on the dependent variable are called scores and are taken twice from each group. The first measurement is known as the **pre-test**. It is taken before administering experimental stimulus to the experimental group. The control group also undergoes measurement. The second set of measurements takes place after the experimental group is exposed to the experimental stimulus. This is known as the **post-test**. The control group, of course, is not exposed to the experimental stimulus. But both groups are measured. The difference in measurement between the post-test and pre-test will be compared in the experimental and control groups. In the case of the difference in the experimental group, we can safely assume that the independent variable causes changes in the dependent variable or that the independent variable affects the dependent variable and is causally related to the dependent variable.

Group	Pre-test	Post-test	Difference
Experimental	M1	M2	M2-M1 =Ra
Control	M3	M4	M4- M3 =Rb

Selden Menefee and Audrey Menefee, who visited the Tiptur area of Karnataka, conducted an experimental study using a mimeographed newspaper to measure the information level of the sample subjects chosen. When the measurements were compared between the experimental group and the control group, it was found that the information level of the experimental group had gone up due to the exposure to the mimeographed newspaper more than the control group that was not exposed to the newspaper, the experimental stimulus. This meant that exposure to the newspaper caused an increase in the information level of those exposed to it.

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. What is the primary purpose of the experimental method?

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

2. Identify the pre-conditions of a deductive explanation.

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

3. Briefly explain the classical experimental method.

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

7.4 PROCESS OF EXPERIMENTAL RESEARCH

The experimental method involves two operations: manipulation and observation. Researchers manipulate an independent variable in scientific jargon and observe the subjects' consequences or responses to the dependent variable. Each experiment may be different, but every researcher must consider the common steps to follow.

Below is a direct quotation from Wimmer and Dominick (2003), who have identified the steps as follows:

- 1) "Every researcher has to select the setting. Most experiments are conducted in a laboratory or any other environment that is under the investigator's control. Experiments are also conducted in natural settings, but the researcher has limited control over them.
- 2) Any research will have to select the experimental design. The hypotheses or research question will determine the appropriateness of the design. The types of variables available for manipulation or evaluation, the availability of respondents, and the availability of resources will impact the design selection.
- 3) Operationalise the variables: In experimental research, the independent variables are operationalised after their creation through manipulation. Scales or rules are constructed to categorise behaviour observations to operationalise the dependent variables.

- 4) Decide how to manipulate the independent variable. A set of clear and specific instructions, events or stimuli is developed to manipulate the independent variables and present them to the experiment subjects. The two types of manipulations are straightforward and staged. "In straightforward manipulation, written materials, verbal instructions, or other stimuli are presented to the subjects". The investigators create events and circumstances that help them manipulate the independent variable if it is staged manipulation. It can be simple or elaborate. The use of a confederate is also in vogue. A confederate is a person who acts like a subject, but in reality, he is also part of the manipulation.
- 5) Select and assign subjects to experimental conditions. The best way to guarantee external validity is to select the subjects randomly from the population and assign them to the experimental and control groups.
- 6) Conduct a pilot study: A pilot study involving a small number of subjects will reveal problems and enable the investigator to make suitable corrections and revisions. It is done to determine whether manipulating the independent variable has had the desired effect.
- 7) Administer the experiment: The main phase of research is data collection. Experimental manipulations are done on the subjects, either individuals or groups. Then, the effect of the independent variable is measured, and the subjects are debriefed. While debriefing, the investigator explains the objectives and consequences or implications of the study undertaken.
- 8) Analysis and interpretation of results: The subjects' scores on the dependent variables measured are tabulated and statistically analysed. Then the investigator interprets the results.

7.4.1 Control of Extraneous Variables

In media research, an investigator encounters unwanted extraneous variables. He or she must remove or control them to ensure the internal validity of the instrument/ method employed for research. These extraneous variables are also called **confounding variables**, which normally contaminate and complicate the findings. The environment, experimental manipulations, experimental design, and the assignment of subjects to both experimental and control groups control the confounding variables.

Generally, three types of techniques are used in assigning subjects to groups. They are:

- Randomising
- Matching
- Generalisation

Matching and randomisation are the two methods of control. Matching control variables are known to the researcher before the research process starts. Randomisation will enable him/her to remove the effect of unexpected or not-seen-before factors.

Matching is a method to equate the experimental group with the control group on external variables presumably related to research hypotheses. The two methods of matching the groups are precision matching and frequency distribution. Precision matching is also known as pairwise matching. Each subject in the experimental group is matched by another one in the control group with the same characteristics. For example, if one person is 30 years of age in the experimental group, another subject in the control group will also be of the same age. The problem with this method is the difficulty of matching many variables. Pairwise matching is a tough task. Precision matching has a very high number of losses of cases in the absence of compatible twins.

This is usually substituted by frequency distribution. Here, experimental and control groups are matched separately rather than in combination for each variable. It is not one-to-one matching, but these two groups are matched on identified central characteristics. For instance, the average age of one group should match the average age of another. This may be less precise, but it is easy to manage. As we know, matching can control a limited number of pre-identified extrinsic variables. A good number of confounding variables may affect the relationship between independent and dependent variables.

7.4.2 Randomisation

One cannot be sure of eliminating the confounding variables. Hence, there is scope for wrong interpretation of results. This can be avoided by opting for randomisation - a process by which subjects are assigned to the experimental and control groups. The assignment can be done by flipping the coin, using a table of random digits, using the lottery method, or any other suitable method. This will ensure an equal probability of being selected for the experimental or the control group. Randomisation will remove any systematic mistake because extrinsic variables may relate to independent or dependent variables.

Internal validity is important in communication research. Another equally significant aspect of research is the generalisability of the findings. The effect of one variable on another in a natural setting and on a bigger population is known as the external validity of research designs. It also refers to the ability to generalise the research findings. Two factors related to external validity that can affect the generalisation of research results are the representativeness of the sample and the reactive arrangements.

Random assignment of subjects to the experimental and control groups will ensure equality between the groups and the investigation's internal validity. However, it may not represent the larger population, only the sample selected. A high rate of refusal and enormous cost may affect opting for a highly representative sample. Next will be the reactive arrangements. The results of a study should be generalised not only to a large population but also to a real-life situation. This is difficult to accomplish. It may result in a reactive effect of testing on the subjects. For instance, a pre-test may affect subjects' responses to the experimental stimulus. Such an effect can be avoided by opting for a post-test only.

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. Explain the two operations involved in the experimental method.

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

2. List the common steps to follow in experimental research.

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

3. Briefly explain the three factors that can be used to control the confounding variables.

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

7.5 EXPERIMENTAL DESIGN

The experimental design refers to the 'total experimental plan or structure of the research'. It is selecting and planning the experiment related to a research problem. The simple procedures are pre-test, experiment, and post-test. Other factors like variables, sampling, control, and the construction of a measuring instrument should also be included in the process. Normally, research designs are classified as experimental and quasi-experimental. We have already discussed the basic procedure of a classical research design. Some of the popular experimental designs are recounted here:

a) Pre-test-post-test control group

The pretest-posttest control group design is one of the basic procedures commonly adopted by media researchers. This procedure helps control rival hypotheses because of the effects of maturation, testing, history, and other sources. Each group faces the same situations during the experiment. This is the other name for classical experimental design.

b) Post-test, the only control group

There is a possibility of subjects becoming sensitive to the experiment during the pre-test. In the post-test-only control group method, neither the experimental nor control group is exposed to the pre-test. Only the experimental group is exposed to the independent variable, which will be a post-test of the group. Then, the two groups are compared to the presence of statistical significance. This helps control the rival explanation as both groups are equally affected by factors like maturation, history, and others.

c) Solomon four-group design

We have already observed that a pre-test has its own merits and demerits. It enables us to assess the time sequence and is the basis of comparison. Yet, we may have reactive effects in the process. More than that, these are occasions where we may not be able to have the luxury of a pre-measurement period. The Solomon four-group design will have the same features as a classic experimental research design. The difference is that an additional set of control and experimental groups will not be made subject to pre-testing. When we have two experimental and control groups each, the comparisons will show whether the experimental stimulus (x) has independently affected the groups that were not made sensitive by a pre-test. If we find that the experimental stimulus had its effect without a pre-test, the findings can be generalised to the target population. The element of causality will increase in its strength.

7.5.1 Factorial studies

When a researcher analyses two or more independent variables simultaneously, it is called factorial design. In this, each independent variable is called a factor. The factorial study has advantages like saving time, money, and resources. It allows the investigator to know the interaction between two or more - dependent variables and the effects they cause on the dependent variable. If we say that it is a two-factor design, it means that two independent variables are manipulated. The same applies to a three-factor design where three independent variables are manipulated. Factors can have two or more levels. An example here will make understanding it easy. Imagine that there is an advertising campaign on radio and television. The subjects of these media are assigned two groups, one that is exposed to the radio and the other to television. These are known as two levels of investigation.

7.5.2 Quasi experimental design

The quasi-experimental design is not suitable for application in certain research situations. This is especially true when manipulation and random assignment of respondents are impossible. In social sciences, two types of relationships are identified. The first is the stimulus-response, and the second is the property- disposition relationship. For the first one, reward and satisfaction or advertisement and consumption levels can be given as examples. For property-disposition relations, attitudes, values, and orientations can be examples. Stimulus- response situations are best suited for experimental methods, while quasi-experimental designs can tackle property- disposition.

The four areas of difference between these two types of designs are identified.

1. The time gap between the effect of an independent variable and the response to it. In stimulus — response relations, the time interval is short. In contrast, property disposition relations are longer.

2. A stimulus is easy to isolate, and its effect can be well-described. On the other hand, properties like social class are unclear, and dispositions like prestige and education have their own effect on other elements.
3. The 'before and after' comparison is easy in experimental studies and difficult in quasi-experimental studies. For example, two groups with different properties and dispositions cannot be compared. An upper-class group differs from a lower-class group on many counts, like values and social practices.
4. The time sequence of events varies in experimental research, and the direction of causality is clear for pre- and post-comparisons. However, the time sequence is difficult to establish in a property-disposition relationship. As such, manipulation, comparison, and control cannot be easily applied to property-disposition relationships.

Experimental methods are strong on control but weak on representation. Quasi-experimental studies, particularly surveys, are strong on representation but weak in control. Experiments can control both extrinsic and intrinsic variables through randomisation. They also have control over the introduction of experimental stimuli. This helps them in determining the direction of causation. These advantages cannot be seen in quasi-experimental designs. On the other, laboratory experiments cannot replicate social life. Quasi-experiments will not have the random assignment of subjects to experimental groups. The major quasi-experimental designs are pretest-posttest, non-equivalent control group design, and interrupted time-series design. The one-shot case study, correlational designs, panel, and trend studies are also considered types of quasi-experimental studies by some scholars, though they can be longitudinal studies.

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. Explain the Solomon four-group methods.

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

2. Distinguish between stimulus-response and property-disposition relationships.

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

3. List the four areas of difference between experimental and quasi-experimental designs.
-
.....
.....
.....

7.6 FIELD EXPERIMENTS

As you are aware, laboratory experiments are artificial, so they do not have external validity. In contrast, field experiments have external validity. Subjects react in their natural setting with normal behaviour. For example, watching a TV commercial in a laboratory setting differs from viewing it in a natural setting. One of the characteristics or advantages of field experiments is that they are closer to reality and often non-reactive. Complex social processes and events can be studied. Field experiments can be less expensive as they do not need special equipment. Yet, if the size of the study increases, then the expenditure may also increase. Sometimes, field experiments are the only option left in certain social situations.

There are a few practical difficulties with the field experiments. For instance, certain research studies cannot be done in the field for ethical reasons. The impact of television violence on children cannot be measured through field experiments. Field experiments now and then face external obstacles that may be unexpected. It is time-consuming when the researcher wants to establish contacts, seek cooperation, and obtain necessary permission. Often, this procedure takes weeks and months. The field experiment's most difficult part is controlling intervening variables. The extraneous variables will affect the accuracy of experiments and, in turn, the entire research work.

Types of Field Experiments

Field experiments have been categorised into two: experiments in which the investigator manipulates the independent variable and the other in which manipulation of the independent variable happens naturally. Wimmer and Dominick provide an example for the first type. Imagine an investigator who wants to know the effects of being unable to read a newspaper. The option is to choose two comparable samples and not allow one sample to read a newspaper for some time. The second sample will continue to read the newspaper. Later, a comparison will be made to determine the effect on other areas of life, such as interpersonal communication. The other types of field experiments deal with passive manipulation of the experimental stimulus.

In some cases of field experiments, the investigators may not be able to assign subjects randomly to experimental and control groups. Such field experiments are known as quasi-experiments. One major challenge for a researcher is to guard against alternative explanations.

7.7 MERITS AND DEMERITS OF EXPERIMENTAL METHODS

Like every research method in social sciences, the experimental method has advantages and disadvantages. The practical advantages of experiments are:

- a) **Evidence of causality:** Experiments enable us to determine the cause-and-effect relationship between an independent and dependent variable. They also support controlling the time order and other possible causes of research efforts.
- b) **Control:** As already stated, control is an advantage of the experiments. An investigator will control the environment and isolate the testing from other influences and effects. This means that they can structure the experimental setting according to their choice. However, the artificial nature of the laboratory setting is a disadvantage. The selection and control of subjects are other advantages of experimental research.
- c) **Cost:** Experiments can be less expensive compared to other research methods.
- d) **Replication:** The experiments can be replicated. Since the conditions under which they are conducted are clearly described, it becomes easy for others to repeat them. Modifications can also be made.

Disadvantages:

The experimental method has its disadvantages:

- a) **Artificiality:** As already stated, the artificiality of the research environment is a major problem in experiments, as behavioural alterations take place when media researchers employ the laboratory setting. Because of the artificial environment in the laboratory, results that emanate from these do not resemble behaviour in a natural atmosphere or real-life situation. In real life, subjects are exposed to competing stimuli from other sources. The artificiality may affect the external validity of the research.
- b) **Researcher bias:** The experimenter can influence the experiment so that he or she can get the result according to the pre-decided hypotheses. To avoid this type of bias on the researcher's part, a technique called double-blind is used. In this, neither the researchers nor the subjects will know whether a particular subject is given to the experimental or control group.
- c) **Limited scope:** Only select research questions can be experimented upon. When issues and topics in mass communication concern a large number of people, experiments on a massive scale are not possible.

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 the characteristics of field experiments?

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

2. Identify the categories of field experiments.

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

3. Enumerate the advantages and disadvantages of the experimental technique.

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

7.8 LET US SUM UP

The experimental method is a technique used in communication research. However, its adoption is restricted in media research, unlike in psychology and social psychology. Natural sciences make the best use of experimental research. Experiments attempt to find out the relationship between factors of cause and effect. These enable us to ensure that the cause is scientific. Explanations in social sciences are classified as deductive and probabilistic.

Deductive explanations are strong as their applicability encompasses the past, present, and future. In natural sciences, deductive explanations are made possible with manipulation and control. But in social sciences, only a few phenomena are amenable to manipulation and control.

A classic research design contains two equal groups: experimental and control. The experimental group is exposed to experimental stimulus, and the control group is not. The measurements taken during the pre-test and post-test are compared to know whether the experiment involves manipulation and control, which are difficult to achieve in social sciences. The experimental technique normally has eight steps. Control of confounding or extraneous variables can be done by matching, randomisation, and generalisation. Internal and external validity will have to be achieved in experimental research. There are two designs when we deal with experimental methods: Experimental design and quasi- experimental designs, pretest-posttest control group, post-test only group, and Solomon four-group designs belong to experimental design.

When a researcher analyses two or more independent variables simultaneously, it is called a factorial study or design. Here, each

independent variable is called a factor. Pretest-posttest non-equivalent control group designs and interrupted time series designs belong to the category of quasi experiments.

Experimental designs are unsuitable for application in certain research situations, especially when manipulation and random assignment of respondents are impossible. Experimental designs are best suited to stimulus-response relationships but not to property-disposition relationships. Quasi-experimental designs, including one-shot case studies, correlational designs, and panel and trend studies, can tackle the property-disposition relationships. Field experiments are conducted in natural settings. They are closer to reality and often non-reactive. They do not need any special equipment. The costs can be lower compared to other methods. The advantages of experimental methods are evidence of causality, control, cost-benefit, and replication. The disadvantages are artificiality, researcher's bias, and limited scope.

7.9 KEYWORDS

Confounding variables : Variables which are extraneous to the experiment. They contaminate the findings.

Deductive explanation : An explanation that has the element of generalisation regarding past, present and future.

Empiricism : Verification with strong evidence.

Explicandum : An event or phenomenon to be explained.

Manipulation : Deliberately produce a change in the phenomenon selected.

Pilot study : Involves a small number of subjects on whom the study is conducted.

Subjects : People who are selected to participate in the experiment.

Verstehen : A tradition of understanding human behaviour.

7.10 FURTHER READINGS

1. Broota, K.D (1992), Experimental Design in Behavioural Research, New Delhi, Wiley Eastern Limited.
2. Nachmias, David and Chava Nachmias (1976), Research Methods in the Social Sciences, New York, St. Martin's Press.
3. Priest, Susanna Hornig (2010), Doing Media Research: An Introduction, New Delhi: SAGE Publications India Pvt. Ltd.
4. Wimmer D. Roger and Joseph R. Dominick (2003), Mass Media Research, Belmont (USA) Wadsworth/Thomson Learning.

7.11 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. The main purpose of the experimental method is to determine the relationship between factors of cause and effect. Experiments are more suitable for natural sciences than social sciences. In other words, experimental research attempts to determine the cause-and-effect relationship between an independent variable and a dependent variable.
2. The four pre-conditions of a deductive explanation are:
 - A universal generalisation
 - A statement of conditions which generalises appear true
 - An event to be explained (explicandum)

The rules of logic.

3. A classical experimental research design has two equal groups for comparison: experimental and control. The experimental group is exposed to the experimental stimulus, but the control group is not, for post-test. The difference between the post-test and pre-test on the measurements taken will prove that the independent variable affects the dependent variable.

Check Your Progress: 2

1. The two operations involved in experimental research are manipulation and observation. In scientific jargon, researchers manipulate an independent variable and then observe the subjects' response to the dependent variable.
2. The eight steps involved in experimental research are selecting the setting, selecting the experimental design, operationalising variables, manipulating the independent variable, selecting, and assigning subjects to experimental conditions, conducting a pilot study, administering the experiment, and analysing and interpreting extraneous variables.
3. The confounding variables can be controlled by matching, randomisation, and generalisation; matching can be either precision matching or frequency distribution. Randomisation can be done by assigning the subjects to experimental and control groups using different methods. It is called generalisation if the findings can be applied to a large population.

Check Your Progress: 3

1. The Solomon four-group design will have the same features as the classic experimental research design, except that it will have an additional control and experimental set of control and experimental groups not subjected to pre-testing. The findings can be generalised if we

- find that the experimental stimulus had its effect without a pre-test.
2. The stimulus-response relationship is based on a reward-satisfaction proposition. The property-disposition relationship is based on attitudes, values, and orientations. The first one is amenable to experiments, while the property-disposition relations are taken in the case of quasi-experimental designs.
 3. The four areas of difference between experimental and quasi-experimental designs are the time gap between the effect of the independent variable and the response to it; a stimulus can be easily isolated, but properties are not; before and after comparison; and the variation in time sequence.

Check Your Progress: 4

1. Field experiments are carried out in natural settings, which gives them external validity. Subjects will react with their normal behaviour. Field experiments are non-reactive, less expensive, and do not need special equipment or facilities.
2. There are two types of field experiments: Experiments in which the researcher manipulates the independent variable and the other in which manipulation of the independent variable happens naturally.
3. The advantages of the experimental method are evidence of causality, control, less cost, and replication. The demerits are artificiality, the researcher's bias, and limited scope.

UNIT 8 VISUAL ANALYSIS

Structure

- 8.0 Introduction
 - 8.1 Learning Outcomes
 - 8.2 Visual and Visuality
 - 8.3 Iconology
 - 8.4 Formalism
 - 8.5 Simulation
 - 8.6 Scope of Inquiry
 - 8.6.1 Contextual Inquiry
 - 8.6.2 Historical Inquiry
 - 8.6.3 Personal Language
 - 8.6.4 Reflexivity
 - 8.6.5 Communities and Culture
 - 8.6.6 Philosophy and Visuality
 - 8.6.7 Hybridisation
 - 8.7 Sites and Modalities
 - 8.8 Methods of Visual Analysis
 - 8.8.1 Semiotics
 - 8.8.2 Content Analysis
 - 8.8.3 Discourse Analysis
 - 8.9 Let Us Sum Up
 - 8.10 Keywords
 - 8.11 Further Readings
 - 8.12 Check Your Progress: Possible Answers
-

8.0 INTRODUCTION

Today, more than ever, we live in a world of images. We are bombarded by images everywhere. The centrality of images in our lives cannot be ignored anymore. Our perception of ourselves and the world around us is largely influenced by images. In fact, images have become pervasive, with their application in every field or discipline, be it medicine or anthropology, defence, or communication.

In the entire history of mankind, people have never consumed or produced images on the scale we are witnessing now. The images are almost saturated. "The spectacle is the leading production of present-day society," claimed Guy Debord.

The possibilities of image craft have been extended boundlessly with digital technologies, particularly information and communication technologies, that aid in creating, altering, and transmitting images. Though image-making or

image-reception is as old as the hills and has been common across human cultures and civilisations, digital devices and the Internet have revolutionised the very impulse of visual imagination, redefined the ways of living, and made the ubiquity of images inevitable. Indeed, there can be no denial of the pre-eminence of technology. As Marshall McLuhan professed, every technology alters man's psychic and social complex, all the more so with digital technology, which has a profound impact on our sensory environment.

In a way, the arrival of the Internet, has led us to the age of image. From emojis, selfies, memes, PPTs, and VR games to 3D views, a whole gamut of new images has come to occupy our lives. We are confronted with an image-based culture that is challenging the print culture in many ways. The expanse of images has given birth to terms like ocularcentrism and visual pollution to mention a few.

On the one hand, as many scholars have argued, there is the democratisation of image production, transmission, and consumption, with a majority population involved. On the other hand, there is the overarching concern over distraction or loss of control over it. Then, there is this discourse on the relationship between the image world and the real world - whether the real world constructs the image universe or vice versa. There are arguments that 'life has receded into representation' (Guy Debord).

Further, there is a debate on the authenticity of the image itself. With software like Photoshop enabling immense editing possibilities and after-effects, the dichotomous relationship between the original and the manipulated image is being re-examined in a different plane of postmodernity. In the context of the mechanical reproduction of art, accentuated by printing technologies, Walter Benjamin argued about the loss of authenticity, aura, and cult value.

With digital technologies enabling instant replication, sharing, and altering images, the discourse has moved to simulation and hyperreality, where there is a copy but not the original. Stuart Hall's ideas on reception provide another dimension - how the audience can decode these images - dominant reading, negotiated reading and oppositional reading.

When it comes to analysing images/visuals, there are six basic perspectives—personal, cultural, historical, technical, critical, and ethical. An image or a set of images can be analysed with one or a combination of these perspectives, which leads to different kinds of inquiries. Generally, the focus is on formal elements and/or content. Later in this chapter, we will investigate the types of inquiry and basic visual analysis methods.

8.1 LEARNING OUTCOMES

After completing the Unit, you should be able to:

- Appreciate the significance of visuals in society;
- Grasp basic debates about visual culture;
- Outline the types of inquiries in studying visuals;

- Identify the sites and modalities; and
- Understand various methods in visual analysis.

8.2 VISUAL AND VISUALITY

By virtue of our vision, all of us can see images/visuals. We see photos, memes, movies, ads, paintings, etc. Images can convey information and ideas. They can also be used for aesthetic pleasure and entertainment. The purposes of images have varied with time, place, and person. They can serve various purposes, including spiritual, cultural, political, aesthetic, and commercial.

So, what is an image?

In his book *Ways of Seeing*, John Berger defines an image as "sight which has been recreated or reproduced. It is an appearance, or a set of appearances, which has been detached from the place and time in which it first made its appearance and preserved for a few moments or a few centuries."

Images can be broadly classified into one-dimensional (1D), two-dimensional (2D) and three-dimensional (3D) images. A one-dimensional image has only one plane, i.e., length and no width. For example, a straight line. In the case of 2D images, there are two planes - length and breadth, like in a drawing, and 3D, there are three planes - length, breadth, and height, as in sculpture. A two-dimensional image is flat. A three-dimensional image has volume. Also, images have visual elements or properties like line, shape, volume, colour, size, texture, tone, and pattern depending on the medium and form.

On the other hand, visuality refers to how vision is constructed in various ways: 'how we see, how we are able, allowed, or made to see, and how we see this seeing and the unseeing therein,' as Gillian Rose quotes. Every image leads us to see something in a certain way. This is precisely where the scope for analysis and interpretation lies. Visual analysis deals with how the visuals communicate, the meaning they generate, the effect they produce and the functions they serve. It can be applied to visual materials like art, photography, cinema, posters, sculpture, etc. Visual analysis can describe and analyse the visual material, situate its context, and interpret the content.

8.3 ICONOLOGY

The term iconology is derived from the Greek words *ikon* and *logia*, which mean image, reason, thought and discourse, respectively. Iconology evolved to study individual artworks or bodies of artworks with a specific focus on the subject matter or meaning of works. Iconology is a branch of art history that deals with the social, cultural, and historical background of the themes and subjects in visual arts.

The leading proponent of iconology, Aby Warburg, stated that iconology's idea is to trace motifs to different cultures and visual forms and explore artworks through 'synthetic intuition', i.e., the basic drives of the human psyche. Iconology tries to unearth the psychology and the imprint of culture on creating artworks. Both material and mental images are considered, and

artwork is investigated as a product of time and place.

It probes the following questions:

1. Who created the artwork/s? In the case of patronised work/s, who commissioned the work and for what purpose?
2. Why were the particular artist chosen? Why at a specific location and time?
3. Why was it represented in a certain manner?
4. What influences did the artist bring to his work/s—religious, social, historical, ideological, and personal?

German Professor Erwin Panofsky developed a method of analysing artworks known as iconography, which emphasises the description of the representation. Basically, iconography studies a particular range or system of images used by an individual artist or group of artists to convey particular meanings. For instance, the iconography of Amrita Shergil and Frida Kahlo as autobiographical works reflect the personalised experience of the artists as women.

Panofsky identified three layers/strata of meaning in artworks – representational, iconographical, and iconological. He devised the following three-point system to apply to the works of art:

I. Primary/Natural level:

The first level, the pre-iconographic level, is about the natural subject matter. Here, we see what we get, i.e., we can identify only the basic subject matter of a painting. Here, we recognise the objects, gestures, activities, people, places, and things. At this level, we do not need any inside cultural, conventional or art historical knowledge to understand the artwork. Our practical daily life experience comes into play, and we identify the motifs used. We can understand only as much as someone from any other culture would understand. This level is subdivided into the factual and the expressional sections.

II. Secondary/ Conventional level:

In the second level, we look at a deeper level of meaning. The preliminary work of iconology begins here. We must know the cultural and artistic conventions in use to understand an artwork. Here, we go beyond recognising objects to identifying ideas, themes, and concepts through artistic motif/s. We have progressed to the first step of iconology in interpreting the artwork/s. This step is also referred to as iconographical symbolism.

III. Tertiary level:

The third level is the highest goal of iconology. We explore the intrinsic meaning or content. At this level, Panofsky says, we understand the underlying basic attitude of a nation, a period, a class, a religious or philosophical persuasion. Here, we look at the ideological meanings. Panofsky considers art a symptom of something else and expresses itself in countless other symptoms; therefore, the compositional and

iconographic features are interpreted as particularised evidence of something else. Symbolical values that artists consciously intended or even unaware of form the basis of iconology. This level is also referred to as iconological symbolism.

For further understanding, see Table.1 (Panofsky's System)

The iconographical study is most suitable for studying the artworks of the past. It entails observation, documentary research, intertextual comparison, and classification. Both textual and contextual research are conducted. As Theo Van Leeuven, in *The Handbook of Visual Analysis*, notes, iconographers will

1. Find out as much as possible about the circumstances under which the objects of their studies were created.
2. Collect and verify all the available factual information.
3. Read books on theology and mythology to identify the subject matter and
4. Observe the interplay between the influences of literary sources and the effect of self-dependent.

Table 1. Panofsky's System

	Object of Interpretation	Act of Interpretation	Equipment for Interpretation	Controlling Principle of Interpretation
1.	Primary or natural subject matter - (A) factual, (B) expressional, constituting the world of artistic motifs	Pre-iconographical description (and pseudo-formal analysis)	Practical expression (familiarity with objects and events)	History of Style (Insight into how, under varying historical conditions, objects and events were expressed by forms.)
2.	Secondary or conventional subject matter, constituting the world of images, stories, and allegories	Iconographical analysis	Knowledge of literary sources (familiar with specific themes and concepts)	History of Types (Insight into how, under varying historical conditions, specific themes or concepts were expressed by objects and events.)

3.	Intrinsic meaning or content, constituting the world of 'symbolic' values	Iconological interpretation	Synthetic intuition (familiarity with the essential tendencies of the human mind), conditioned by personal psychology and 'Weltanschauung.'	History of cultural symptoms or symbols in general (Insight into how, under varying historical conditions, essential tendencies of the human mind were expressed by specific themes and concepts.)
----	---	-----------------------------	---	--

8.4 FORMALISM

As its name suggests, a formal analysis analyses the form of the visual, i.e., the visual attributes, material attributes and composition. Figure, shape and delineation are of high significance. Formalism, an antithesis of iconography, is a methodology that prioritises "form" over "content." It focuses on the figure and plays in the form.

Form constitutes all the elements that are used to express content. Both form and content are interrelated and, in a way, co-dependent and inseparable. However, Roger Fry, an artist and art critic who emerged as one of the most influential and leading formalist critics, stated that art has little or no meaningful connection with the artist or the cultural and historical context from which it emanated. He encapsulated art as a product of people's imagination and emphasised the emotional effects of artwork.

In "An Essay in Aesthetics," Fry argues that the response felt from examining art comes from its form, meaning that the use of line, mass, colour, and overall design invokes an emotional response. His specific contribution was to perceive the elements that give an artist his significance. Fry's approach mainly relies on dissecting each element in the art and understanding its contribution to its overall impression. There is ample scope for a subjective view in interpreting the artwork/s.

Basically, a formal analysis aims to study formal elements like line, colour, contrast, shape, and mass and how they are composed in various ways to create balance, proportion, and order as well as rhythm and pattern. It compares style and form and explores visual, spatial, material, and compositional aspects of artwork. Artwork as pure form, emancipated from the artist's intention and context, was the crux of formalist thought.

However, later critics and scholars have contested the rigid distinctions of

iconography and formalism and stressed the combination of the two approaches to studying artworks. This mixed method has comprehensively accounted for the differences in style in various historic periods, cultures, artistic schools, and movements.

8.5 SIMULATION

Have you ever wondered why the images of products or celebrities we see are more impressive than the product or the celebrity? This is the power of image today. The images are nothing but copies. Yet, the copies seem more real than the original or more real than the real. The problem of representation doesn't arise here as in modernity, where the artist was troubled with representation, as there was this quest and curiosity to represent reality or truth.

Now, in postmodernity, the signs are detached from reality. Copies fundamentally characterise them without the original or, at least, copies whose original doesn't exist anymore. Therefore, there is no distinction between the real and the unreal, and everything is a matter of simulation/simulacrum, and the significance of the original recedes.

French philosopher Jean Baudrillard theorised about this phenomenon of simulacra and hyperreality. Baudrillard argued that this is not a question of imitation, duplication, or parody but that of substituting the real for the real. The realm of the hyperreal is more real than real. Examples are malls, virtual reality games, theme parks, Disneyland, TV sports, Instagram, film cities (Ramoji Film City), and other such simulations. Further, Baudrillard encapsulates hyperreality as "the meticulous reduplication of the real, preferably through another reproductive medium, such as photography." The hyperreal world's images, models and codes control our thoughts and behaviour.

As explained in *The Stanford Encyclopaedia*, in this postmodern world, individuals flee from the "desert of the real" for the ecstasies of hyperreality and the new realm of computer, media, and technological experience. *As simulations increase, they refer only to themselves: a carnival of mirrors reflecting images projected from other mirrors onto the omnipresent television and computer screen and the screen of consciousness. This refers to the image's previous storehouse of images produced by simulatory mirrors.*

Baudrillard, in his seminal essay, "The Precession of Simulacra," describes four successive phases of the image:

- it is the reflection of a profound reality;
- it masks and denatures a profound reality;
- it masks the absence of a profound reality;
- it has no relation to any reality whatsoever: it is its pure simulacrum

Baudrillard summarises that the masses seek spectacle and not meaning.

8.6 SCOPE OF INQUIRY

Depending on the framework or objectives of the study, we can engage in different kinds of inquiries into visual arts/phenomena. The inquiries are derived from the six perspectives of visuals - personal, cultural, historical, technical, critical, and ethical. Here, we outline some of the inquiries:

8.6.1 Contextual Inquiry

Contextual inquiry aims at interpreting subject matter, describing function/s and identifying the cultural/historical significance of the visuals. Primarily, we try to explore what is shown and why. We try to understand the work by examining the historical context in which it is situated. We try to compare it with texts and other imagery from the time and go beyond the aesthetic purpose. The context can be ascertained through history and identifying types of work - portrait, tablet, tomb, worship place, and palace. The motivation of the artists and patrons can help them understand the purpose and appearance. The inquiry can be based on formal analysis (understanding the shared conventions of art in its historical iconography, context, function, style, art as a cultural artefact, visual experience, and critical thinking (with ideological frames of Feminism/Marxism/Pluralism...)). This contextualisation helps to situate the work within a larger framework/background to identify its signification, significance, and impact at the time of its production.

8.6.2 Historical Inquiry

In historical inquiry, we are firstly concerned with the who, what, when, where and why of an image. In the case of patronised works, why and how it was commissioned-produced are also taken into consideration.

Secondly, we try to locate and analyse historical resources to establish the historical evidence required to construct an interpretation of the work. We look at major events and developments (social, political, religious, etc) during the period and what caused them, how they influenced the artist's decision, how they shaped the artist's work and the meaning produced. It can probe the medium and techniques used by understanding their historical usage as certain mediums and techniques are integrally connected to specific periods, times, and purposes. It could be part of a cultural tradition.

8.6.3 Personal Language

This pertains to an artist's style, i.e., the visual language used or evolved by her/him - the choices made in composition, technique, medium, idea, theme, etc. Artists may work in a single medium or use multiple mediums within a specific genre, defy generic conventions, or hybridise on consistent themes or specific settings - historical or contemporary. The influences on the artist, the inspiration, upbringing, life and experience, beliefs and values, and technical skills could be valuable here. Some artists can evolve their stylistics. For instance, Dutch painter Rembrandt's use of lighting is referred to as Rembrandt lighting, Stanley Kubrick's use of one-point perspective, and Orson Welles' deep focus photography in films.

[Image of Rembrandt's painting with Rembrandt Lighting, a still of one-point perspective in Kubrick's *The Shining*, a still of Orson Welles *Citizen Kane*]

8.6.4 Reflexivity

Images often involve a high degree of reflexivity, with the artist reflecting on her/his positionality. Their perception of art or self itself is embodied in the work/s. Therefore, the inquiry is centred around the artist's interrogation of self and his/her preoccupation with art. Self-portraits, mirror reflections, illusion, and metapainting are a few practices in reflexivity. In photography, reflections, shadows, framing, irregular angles, posture, expression, and distortion are employed. Often, the audience is made aware of the work as an artificial construction. The inquiry focuses on the visual projection of self, the embedded narrative/s and the artist's self-awareness. Examples of self-reflexive practices - Bengali filmmaker Ritwik Ghatak's film *Jukti Takko Aur Gappo* (1975) on his status as a partition refugee in Bengal, Iranian Filmmaker Abbas Kiarostami's *Koker* Trilogy in which self-referencing is extensively used. Artists Frida Kahlo, Van Gogh and photographer Vivian Maier's self-portraits are good examples. Selfie as an art form can be studied under reflexivity.

8.6.5 Communities and Culture

Particular communities or cultures have traditions of producing certain kinds of art - ritual-based illustrations, performance, and spectacle. For example, totems, idols, and tribal/indigenous art. Worli art is a genealogical register of the Gond Tribes in Central India; the *Theyyam* dance in Kerala with intricate, colourful face painting for invoking ancestral spirit during worship, and mandala floor drawings of *Nagamandala* ritual of coastal Karnataka for worshipping snakes to mention a few community-based art forms.

Another dimension that can be investigated is the relationship of an artist's work/s to particular cultures and communities. Robert Flaherty's documentary film *Nanook of the North* (1922) on the Nanook tribe, Maya Deren's *Ritual in the Transfigured Time* (1946) on Voodoo worship, and Vinod Raja's *Bird Trapper or Beggar* (2017) on nomadic Hakki Pikki community can be explored in this light.

8.6.6 Philosophy and Visuality

In this inquiry, we attend to the philosophical ideas behind the image. We are keen on mapping the search for the meaning of life that the artist is engaging in a work or series of works. It can be philosophies that led to art movements like impressionism, expressionism, surrealism, dadaism, etc. It can be the philosophical leanings of the artist, like feminism, Marxism, postmodernism, etc. It can also be religious philosophies like Buddhism, Judaism, Confucianism, etc. Each of these *isms*/philosophies has certain stylistic and formal elements. The inquiry can be conducted through the application of the frameworks of the philosophy as well as analysis of the formal and stylistic components of the work. Discourses on representation of race, caste, gender, class, etc. can also be discussed here.

8.6.7 Hybridisation

Hybridisation is blending two or more mediums/materials/genres to produce a visual work. Hybridisation is of various kinds - intermixing new/unusual materials with traditional mediums, crossing over aesthetic elements, fusion of genres, etc., to produce new effects. It can also be a hybrid of objects like paintings, drawings, and photographs, as in collages, between images and techniques through VFX. It can blur the distinctions between media, materials, art, and science. Hybrid products break away from rigid conventions and push the limitations and boundaries of art practice/s. Examples of hybrid works are body art, memes, parodies, remixes, and dub smash.

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. What is visuality?

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

2. List three levels of iconography.

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

3. Define formalism.

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

4. Elaborate on simulation and hyperreality.

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

5. Differentiate denotation and connotation.

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

8.7 SITES AND MODALITIES

In the book *Visual Methodologies*, Gillian Rose summarises, “interpretations of visual images broadly concur that there are three sites at which the meanings of an image are made: the site(s) of the production of an image, the site(s) of the image itself, and the site(s) where various audiences see it.” He adds that each site(s) has three different aspects called modalities.

The three modalities Rose suggests are:

1. Technological, where visual technology is defined as ‘any form of apparatus designed either to be looked at or to enhance natural vision’, such as oil paintings, photography, television, and the Internet.
2. Compositional, where compositionality refers to the specific material qualities of an image or visual object, i.e., the formal strategies like content, colour, and spatial organisation.
3. Social means a wide range of economic, social, and political relations, institutions and practices that surround an image through which it is seen and used.
4. All these three modalities have a bearing on all three sites.

8.8 METHODS OF VISUAL ANALYSIS

8.8.1 Semiotics

We can all identify common signs around us—traffic signs, toilet signage, flags, emblems, etc. We have signs in schools, railway stations, malls, and almost every public space. All communication happens through signs only. Then, how do we understand sign/s?

The sign is defined as something that stands for something other than itself. The meaning of the sign lies outside of it. Now, we wonder how the sign is connected to the meaning it represents. A sign's meaning can be understood through resemblance, as in a photograph, through causal relation, as in the case of smoke (the fire is the cause, smoke is the effect), and through learning, as in the case of language. Any entities like words, images, and objects are signs. Gestures, facial expressions, and body language are signs. Signs have material forms, i.e., can be seen, heard, touched, smelt, or tasted, but act as a link or vehicle to something immaterial.

Swiss linguist Ferdinand de Saussure, while studying the structure of language, invented the term semiology to refer to the science of signs. Semiology was derived from the Greek root *semeion*, which means sign. This branch of science deals with what signs consist of and the laws that govern them. According to Saussure, semiology is the "life of signs."

It looks at how the relationship between the sign and the ‘something else’ results. The society teaches us this relationship.

Saussure developed a two-part sign model - A sign consists of a signifier and a signified.

The signifier is the form the sign takes (which carries or produces meaning), and the signified is the concept or object it represents (the meaning itself). The relationship between the signifier and signified is called signification. Saussure identified the signifier as the sound image and the signified as the concept.

E.g.: T- R-E-E

Signifying practices refer to how rather than what meanings are produced. Saussure noted:

1. The relationship between signifier and signified is arbitrary.
2. The relationship is not intrinsic but relational; it is not necessary or inevitable.
3. The relationship is differential, i.e., a system of differences and oppositions exists.

All of Saussure's discussions of signs were related to structural linguistics.

Later, American scholar Charles Sanders Peirce proposed a three-part model in which the sign is divided into the sign or representation, the object to which it refers, and the interpretant to the meaning it conveys.

A sign can have multiple signifieds/interpretants. This polysemy can lead to confusion or miscommunication as the intended meaning may not be received.

Pierce's major contribution was in identifying images as signs. He divided signs into types based on the relationship between the signifier and the signified.

Three of these have been widely accepted - icon, index, and symbol.

1. An icon shares a relationship of similarity/resemblance. A photograph, a statue, or a portrait are icons as they resemble the person or place, they depict. This resemblance can be identified through seeing.
2. An index has a relationship of connection, so it points to what it stands for either through causal relationship or contiguity. This can be figured out. Fire and smoke share a causal relationship; therefore, smoke is an indexical sign.
3. A symbol has a relationship of arbitrariness. A symbol's meaning can be understood through learning its conventions. The National Flag is a symbol. The country the flag represents is identified by learning about it. Alphabets, words, and language are all symbols.

Visuals can be iconical, indexical or symbolical.

American scholar Charles Morris developed the behavioural theory of signs. He claimed that signs have three types of relations: 1. to objects, 2. to persons, and 3. to other symbols. Correspondingly, his semiotics has three branches: semantics, syntactics, and pragmatics. Semantics inquires into the relationship of signs to what they stand for, Syntactics (or syntax) examines the formal or structural relations between signs, and pragmatics investigates the relationship of signs to interpreters.

Visual Semiotics

When studying visuals in semiotics, we have to consider two aspects—denotation and connotation, referred to as two-order signification. French scholar Roland Barthes developed the theory of two-order signification, where any sign has multiple signifieds / layered meanings. Apart from the general meanings, there can be cultural meanings attached to a sign.

A sign has two sets of meanings –

- a) First-order meaning is literal, apparent, and manifest. This is the common-sense meaning. This is denotation.
- b) The second order means latent and hidden. This is the ideological meaning. This is connotation.

Denotation addresses representation, and connotation addresses the ‘hidden meanings’ of images. Barthesian semiotics and Panofsky’s three-point system are concerned with the meaning of the visual/artwork. The denotation is akin to the primary level in that we only identify the people, places, and objects as they were in the image. The connotation is the amalgamation of the secondary and tertiary levels, locating cultural and ideological meanings. The only difference is the consideration of the context in the iconographic method.

8.8.2 Content Analysis

Content analysis is a research method used to identify patterns in a set of texts, which can be written, oral, or visual. It is mostly used as a quantitative method, but at times it involves coding and numerical values. Here, when we use the term text, we mean any object whose meaning and significance you want to interpret in depth. These objects are not necessarily written and spoken. They can be paintings, films, visual artefacts, etc.

Content analysis can be conducted on a set of images, generally many images. The data is systematically collected from the set of visual texts chosen as the sample. Content analysis can handle large quantities of data. Typically, they try to work based on a formulated, more specifically, a comparative hypothesis. For example, Lipstick advertisements are red.

The method allows the description of one or more defined areas of representation, periods, or types of images. It can be applied to formal elements like shape, size, position, colour, etc. It can be applied to representation in movies, television, etc., or any object whose meaning and significance you want to interpret in depth. Each text, whether verbal or visual, is called text.

The units of analysis (of manifest content) must be clear, reliable, observable, and defined logically to allow quantification. Then, under each unit of analysis, specific and relevant categories have to be devised. It is necessary to ensure that each of the categories is built from within the set of the unit, i.e., it should be the dimensions or qualities of the particular unit of analysis alone. Each of the units has to be mutually exclusive and exhaustive. For example, in a study about photography, shots can be a unit, angle can be a

unit, orientation can be a unit, etc. Under each of these units, we can have categories like close-up, mid-shot, and long shot for shots and high, low, and irregular for angle, and portrait and landscape for orientation. The coding can be done based on the categories.

Leeuven writes the typical research questions which may be addressed using content analysis, including:

1. Questions of priority/salience of media content: How visible (how frequently, how large, and in what order in a programme) are different kinds of images, stories, and events represented? ...
2. Questions of ‘bias’: comparative questions about the duration, frequency, priority, or salience of representations of, say, political personalities, issues, policies, or of ‘positive’ versus ‘negative’ features of representation.
3. Historical changes in modes of representation of, for example, gender, occupation, class, or ethically codified images in particular types of publications or television genres.

Steps in Content Analysis:

1. Select the texts to analyse - Choose the medium, the genre, the artist, the period to be focused on, and any inclusion/exclusion criteria. Collect the determined sample.
2. Define the units of analysis and categories.
3. Set the coding rules.
4. Apply the rule and do the coding.
5. Analyse the results.
6. Conclude.

8.8.3 Discourse Analysis

Michel Foucault’s monograph *Archaeology of Knowledge* (1972) inspired an entire method/body of research called discourse analysis. Discourse analysis emphasises language in use, institutions, and social practices, particularly drawing insights from Foucault’s theories on power, knowledge, and discourse—the production and maintenance of knowledge and unequal power relations.

Originally applied to written and spoken language, discourse analysis is now applied to any sign language, hence its application in visual arts and media. Discourse analysis is not concerned with the structural aspects of a text. Rather, it explores the relations of power and knowledge construction in terms of works of art and the institutions that produce them.

In *Visual Methodologies*, Gillian Rose defines discourse as groups of statements that structure how we think, and act based on that thinking. In other words, discourse is a particular knowledge about the world that shapes how the world is understood and how things are done. Therefore, Rose notes that ‘art’ does not consist of certain kinds of visual images but of knowledge,

institutions, subjects, and practices that define certain images as art and others as not art.

As such, the method is complex and vague, as no clear steps are specified. Researchers evolve their own system and apply it. Broadly, Foucault's concepts of énonciation (enunciation), the practice of uttering something in a context, and enoncé (statement), the basic unit of discourse, can be applied to the study.

Enunciation, the practice of uttering, is a communication practice. Hence, various forms of communication, including visual communication, are brought under the ambit of discourse analysis. Ways of communication, strategies, signs, technologies, and visualisation can be studied under discourse analysis. The world or society that the visual arts create and the persuasion they engage in can be explored as a discourse. Visuality, as a discourse that renders certain things visible and invisibilises others, can be examined.

Further, the institutions that produce these images and their power in constructing the world can be investigated. In all of these, the images themselves are the argument. The researcher's task is deciphering the constructed visibilities and how they constitute power/knowledge.

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. Explain Morris' branches of semiotics.

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

2. Specify the steps involved in content analysis.

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

3. Elucidate the basic steps of discourse analysis.

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

4. Identify the types of inquiry in visual analysis.

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

.....

.....

.....

.....

8.9 LET US SUM UP

In this Unit, we have discussed the intricate realm of visual analysis, a meticulous process that involves dissecting visual material to understand its context and interpret its content. We have explored Panofsky's three-point system, which categorises artwork subject matter into primary, secondary, and tertiary elements, offering a structured approach to study. Also, this Unit explained formal analysis, such as line, colour, and shape, discerning how they merge to establish balance, proportion, and order within compositions. In the complexities of postmodernity, we discussed the blurring boundaries between the real and the unreal, where simulation and simulacra reign supreme. Understanding the multifaceted meanings embedded within images, we explored three critical sites: production, the image itself, and its audience, each imbued with technological, compositional, and social modalities.

Furthermore, in this Unit, we discussed the concepts of denotation and connotation, indispensable tools of semiotics, and distinguishing between literal and manifest meanings versus hidden and latent interpretations. Finally, we discussed the intricacies of content analysis, a methodological approach to identifying patterns across various textual forms, be it written, oral, or visual, and delve into the concept of discourses, the underlying structures shaping our perceptions and actions.

8.10 KEYWORDS

Iconology : The study of the subject matter of artwork.

Formalism : The study of formal elements of an artwork

Motif : An element in a composition that can be used repeatedly for decorative, structural, or iconographic purposes. It can have symbolic meanings.

Style : is a coherence of qualities in periods or people—consistency in things like medium, function and subject matter. Style varies in time and place, so much like medium and technique.

Dominant Reading : Closer to the preferred/intended meaning of encoder/producer.

Negotiated Reading : Partial acceptance, partial rejection of preferred meaning.

Oppositional Reading : Rejection of the preferred meanings

Semantics	: A branch of semiotics that studies the meaning/s of signs.
Syntactics	: A branch of semiotics that studies the syntax/structure of signs.
Pragmatics	: A branch of semiotics that studies the relation of signs to interpreters.

8.11 FURTHER READINGS

1. Chandler, Daniel (2022). Semiotics; The Basics. London: Routledge
2. Berger, Arthur Asa (2016). Media Analysis Techniques (4th Edition). New Delhi: Sage
3. Leeuven, Theo Van and Jewitt, Carl (2000). The Handbook of Visual Analysis. London: Sage
4. Nayar, Pramod K. (2008). Introduction to Cultural Studies. New Delhi: Viva Books
5. Rose, Gillian (2003). Visual Methodologies: An Introduction to Researching the Visual Materials (4th Edition). London: Sage

8.12 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Visuality refers to the study and interpretation of visual culture, encompassing the analysis of images, symbols, and visual representations across various media forms. It examines how visual elements communicate meaning, influence perception, and shape cultural identity, offering insights into the visual aspects of human expression and communication.
2. Iconography operates on three levels: primary, secondary, and tertiary. Primary iconography involves identifying and interpreting the primary subject matter or central themes depicted in an image. Secondary iconography delves deeper into the symbolism and allegorical meanings associated with specific visual elements. Tertiary iconography explores broader cultural, historical, and contextual interpretations.
3. Formalism is an analytical approach in art and cultural studies that focuses on the formal qualities of artworks, such as composition, colour, line, and texture, rather than their socio-political or historical context. It emphasizes the intrinsic aesthetic properties of art objects and their impact on viewers' perceptions and experiences.
4. Simulation refers to the imitation or replication of real-world phenomena through artificial means, often blurring the boundaries between reality and representation. Hyperreality, coined by Jean Baudrillard, describes a condition where simulations become more real or meaningful than reality itself, leading to a state of heightened artificiality and loss of distinction

between the two.

5. Denotation refers to the literal or explicit meaning of a sign or symbol, representing its surface-level definition or objective representation. Connotation, on the other hand, involves the subjective, cultural, or contextual associations and interpretations associated with a sign, going beyond its literal meaning to convey layers of symbolic significance and cultural connotations.

Check Your Progress: 2

1. Morris' semiotic theory encompasses three branches: syntactics, semantics, and pragmatics. Syntactics studies the formal relations between signs, focusing on their arrangement and structure. Semantics examines the meaning of signs and their relationship to the objects they represent. Pragmatics investigates how signs function in communication within specific contexts.
2. Content analysis involves systematically analysing textual, visual, or audio content to identify patterns and themes. The process typically includes selecting the content to be analysed, defining coding categories, coding the content according to these categories, and analysing the coded data to draw conclusions and insights. Finally, researchers interpret the findings within the context of their research objectives.
3. Discourse analysis involves examining language use in communication to uncover underlying meanings and social constructs. The process includes selecting the discourse to be analysed, transcribing, or recording the discourse, identifying linguistic features and patterns, interpreting these features within their social and cultural context, and drawing conclusions about power dynamics, identity construction, and discourse structures.
4. Visual analysis encompasses various types of inquiry, including iconography, iconology, and visual culture analysis. Iconography focuses on identifying and interpreting visual symbols and motifs within artworks. Iconology delves deeper, exploring the underlying meanings and cultural contexts of these symbols. Visual culture analysis examines broader societal meanings and interpretations of visual artifacts.
5. Visual research employs three modalities: production, documentation, and analysis. Production involves creating visual artifacts, such as photographs, videos, or artworks. Documentation entails capturing visual data through methods like photography or video recording. Analysis entails interpreting visual data to extract meaningful insights, often using techniques such as semiotics, discourse analysis, or content analysis.

UNIT 9 SOCIAL MEDIA ANALYSIS

Structure

- 9.0 Introduction
 - 9.1 Learning Outcomes
 - 9.2 Social Network and Social Network Analysis
 - 9.2.1 Social Network:
 - 9.2.2 Network Analysis:
 - 9.2.3 Social Media Analysis:
 - 9.3 Concepts of Data and Data Management
 - 9.3.1 Data
 - 9.3.2 Data Collection
 - 9.3.3 Data Management
 - 9.3.4 Data Storage
 - 9.3.5 Data Privacy and Security
 - 9.3.6 Data Cleaning
 - 9.3.7 Data Governance
 - 9.3.8 Data Analysis and Insights
 - 9.3.9 Data Reporting
 - 9.4 Cohesion, Centralisation, Decentralisation
 - 9.5 Social Capital, Structural Holes, Equivalence
 - 9.6 Block Models, Ego Models and Building Network Models
 - 9.7 Social Network Theoretical Concepts – Social Capital (structural holes versus Simmelian ties), diffusion, small world, reciprocity, social support
 - 9.8 Network Theories in Cultural Anthropology
 - 9.9 Let Us Sum Up
 - 9.10 Keywords
 - 9.11 Further Readings
 - 9.12 Check Your Progress: Possible Answers
-

9.0 INTRODUCTION

This Unit is concerned with acquainting the learners with some important concepts related to social networks and their analysis. The nature and scope of social media and the social capital that works in the networks will be discussed to deliver the ideas of Social Media Analysis. It would also lead to understanding social cohesion, diffusion of information, and social capital, as well as managing and analysing the data, which is of great value in the modern world.

Again, to get a proper overview of the subject matter, we can understand the definitions, thereby grasping that technology, psychology and marketing are intertwined. Social networks have become the key to understanding human

psychology in technological intervention. This psychology and behaviour are the core elements of social capital. The shifting interest of sociologists and anthropologists in the different nodes and connecting social media links can be an effective area for understanding human behaviour. Hence, social network analysis is crucial for study and central to modern communication.

9.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Identify the concepts and objectives of studying social network analysis;
 - Discuss the subject matter, nature, and scope of data management;
 - Discuss the key elements and factors that network;
 - Understand the concepts of social cohesion, centralisation;
 - Explain the workings of network models; and
 - Describe the theoretical concepts of social networks.
-

9.2 SOCIAL NETWORK AND SOCIAL NETWORK ANALYSIS

Social Media is today the place where all the action is taking place. Thus, social media is an evolution of word of mouth that has been upped with technology's intervention and the complex web's presence. The major components of social media span are 1) Social Networking, 2) Microblogs, 3) Blogs, 4) RSS Feeds, 5) Widgets, 6) Linking and posting, 7) Content Rating, 8) Bookmarking sites, 9) Audio podcasting, 10) Video podcasting. Social media and Social networks are two sides of the same coin. Social networks, also called social networking services, have become the most crucial element of the new world, dominated by new media, social media, and the web of networks. Today, a world highly connected by the Internet has social media as the core communication element. In this case, a deeper understanding of social networks is crucial and fundamental for any behavioural-related study.

9.2.1 Social Network

The element of connecting through social media is what social networks are, and it has changed the entire phizog of human connectivity. The Internet was taken by storm with the advent of blogging, and at the same time, wikis brought easy information home to the world. However, social media created a platform for users to develop their profiles containing basic information, inviting others to register and liaise with profiles of the people they knew or were keen on developing links with. The system created platforms for people with shared interests to connect easily and access information through other profiles. From information base like Wikipedia to connecting for professional linkups like LinkedIn, to sharing political and non-political opinions through Facebook and X (formerly Twitter) and creating a base for photographs through Flickr or Instagram or even uploading videographic content through YouTube, social networks developed new vistas. They created a complex network that needed analysis, even from a capitalist perspective. A social

network is based on a Sociological perspective, similarities in liking, and overlapping interests, among others.

9.2.2 Network analysis

Network analysis deals with simplifying social structures, formulating patterns and developing models that determine common properties of the networks and acceptable methods that can be used to analyse networks in general. The concepts and methods of network analysis can be coordinated through graphs and models.

These fall under the statistical and probabilistic methods and, to a lesser extent, algebraic techniques. These broad concepts, such as social roles or social groups, are moulded into formal models of networks that can become the base of academic or behavioural studies. Network analysis also has a considerable amount of relevance outside the theoretical perspective. The relationship between nodes is a major unit of analysis in social network studies (i.e., the links among individuals, groups, or other objects). Social network analysis requires data on the ties among entities (i.e., the links among the nodes), which could be friendship, co-membership, communication, and shared behaviour.

9.2.3 Social Media Analysis

Social media analysis is an emerging area that is largely the by-product of connectivity and interactive tools and sites. They are the core of studying issues about lifestyles and activities in the digital realms capable of; there is so much misinformation, fragmentation, and confusion within the marketplace today that it is tough to know how to choose the right platforms and how to set up the right processes to achieve our goals. Social Media Analytics is for anyone who seeks the right technologies to implement social listening and measurement programs. As a larger part of our lives has become ruled by social media, we need to know what these platforms are capable of and how to use them effectively to achieve our goals.

Social Network Analysis is thus a different approach to social phenomena, requiring new concepts and methods for data collection and analysis.

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. Elaborate with details about social media analysis.

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

9.3 CONCEPTS OF DATA AND DATA MANAGEMENT

9.3.1 Data

Data is the modern-day currency. Data is information, and all the information available on the web of social networks is the currency of information gold that can be used to understand human behaviour. This could be both sociologically relevant and a useful base for consumer-based companies.

9.3.2 Data Collection

There are different kinds of data: one extracted by companies or social forums from the users of the profile holders on social media and one that is user-generated content. Much user-generated content includes posts, comments, images, and videos. Personal details, demographics, and preferences provided by users when creating profiles are crucial data readily available online. For a data collector, working with spreadsheet programs could be a tedious task, and filtering and manipulating data are becoming core requirements for some jobs.

9.3.3 Data Management

Managing the information available online is one of the primary determinants of data analysis. Data management in social media refers to the strategies, processes, and practices crucial for managing and systematising the vast amount of data generated through social media platforms.

A major determinant of effective data management is maintaining the privacy and security of user information, checking data accuracy, and extracting meaningful insights for business or research purposes.

9.3.4 Data Storage

With data in abundance, there is a significant need for the hour to work for efficient storage and retrieval mechanisms that look at both structured and unstructured data. There are different means to store data.

Cloud Storage: Many social media platforms leverage cloud infrastructure for scalable and flexible storage solutions.

9.3.5 Data Privacy and Security

With great data comes great responsibility for working on its security. Privacy policies are crucial and significant, with threats targeted at vulnerable sectors. Most policies work on implementing encryption, access controls, and other security measures from unauthorised access or breaches.

9.3.6 Data Cleaning

A systematic review and cleaning of datasets to eliminate errors, duplicacies, or outdated information is crucial for smooth working. This requires validating that the data collected is accurate and reliable.

9.3.7 Data Governance

Adhering to data protection regulations and industry standards through establishing and enforcing internal data management and usage policies.

9.3.8 Data Analysis and Insights

Perhaps social media analysis emanates from analysing the existing data with the help of analytics tools. Employing tools and techniques to analyse social media data and derive meaningful insights are the core elements of Social media analysis. In the age of Artificial Intelligence algorithms to predict trends, sentiment, and user behaviour based on social media data. This is the most important part of the entire data-management process.

9.3.9 Data Reporting

Data reporting can mean a lot of things. Most significantly, how the correlations are significant and the applicable trends. Prompt data analysis requires marketing and technical skills to know what the numbers mean, and which are important. The marketing analytics process leads to suggestions for enhancement, planning new programs, and developing new budgets that are more sustainable to the market demands. Understanding the social media equation is crucial in a market-driven world where stringent marketing programs must allow marketing managers to update their case to the market's demands, leading to a decent sales pitch.

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 is Data management? Discuss how Data is analysed.

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

1. Discuss how the audiences can be reached with the help of Data analysis.

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

9.4 COHESION, CENTRALISATION, CORE-PERIPHERY

Let us understand three core concepts:

Social Cohesion: Social cohesion can be defined as the element of

connectedness and commonality among groups in society. It concentrates primarily on two elements: the sense of belonging to a community and the relationships among members within the community itself. Social cohesion consolidates plurality of citizenship by minimising dissimilarity, socio-economic disparities, and fractures in society, bringing in the necessity of people's needs seeking personal development and a sense of belonging.

Centralisation: A centralised network architecture is developed around a single server that handles all the major processing and controls the logistics, data, and major decisions through an established central authority or company processes. Some moderately powerful workstations or smaller tech groups connect to the main server and submit their requests rather than working independently on the available data. This can include applications, data storage, and utilities. Centralised networks do have their limitations, too. For example, a single point of failure can be a big risk factor for organisations. The centralisation of social media can be compromised in terms of censorship and integrity of content created by users. Examples of centralised social media are Facebook, Instagram, and LinkedIn.

Decentralisation: Decentralised social networks operate on independently run servers rather than on a centralised server without the role of business ownership, and here, no single authority holds the data. Rather, it is decentralised over connection nodes among the networks. This, in turn, creates room. Despite the advantages of decentralised social media networks, there are apparent possibilities of attacks orchestrated by content creator communities that could lead to a large extent of loss of information. Mastodon is an example of such a decentralised network, free and open-source software for running self-hosted social networking services. It uses free, open-source program code.

Another good example is Steem, which provides a social blockchain. Steem is a growing social blockchain community that provides room for revenue streams that are possible for users. It works in gratification mode by rewarding for their content.

9.5 SOCIAL CAPITAL, STRUCTURAL HOLES, EQUIVALENCE

Social capital refers to the resources and advantages individuals or groups can access through social networks and relationships. It encompasses trust, cooperation, and shared values within a community or social group. Social capital is valuable because it facilitates collaboration, information exchange, and collective action.

There are two main types of social capital:

Bonding Social Capital: The connections and relationships within a close-knit group of people sharing similar characteristics, such as family, friends, or members of a particular community, are essentially bonding. This creates a deeper sense of relationships while making available support within the group.

Bridging Social Capital: Social groups or individuals with various backgrounds exist on social media. Bridging social capital refers to where connections are built, fostering interaction and cooperation across different communities, leading to a more all-encompassing and diverse network. Social capital can foster different changes in society. For instance, it can help aid:

Economic Opportunities: A strong and diverse social network can give individuals access to job opportunities, information about potential careers, and business connections.

Health and Well-being: Many supportive social networks work as a support system providing emotional and mental assistance. Social capital refers to this element that works for prospered mental and physical health.

Community Development: A community is essential for human sustenance. It fosters collaboration, reliance, and collective values, crucial in community development. Social capital here leads to greater civic engagement and collective problem-solving.

Education: In a highly networked world, education cannot be carried on without any support; social capital can enhance educational outcomes with easy access to resources, information, and accommodating networks that lead to academic upliftment.

Social capital has been widely studied in sociology, economics, and political science to understand its implications for social dynamics and human behaviour.

Check Your Progress: 3

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

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

1. What is social capital, and what are the different elements of social capital?

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

2. What is Centralisation and decentralisation of technology? Define with examples.

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

9.6 BLOCK MODELS, EGO MODELS, AND BUILDING NETWORK MODELS

The network that works today is simplistically divided into different models. These models are:

Block Models

In graph theory or network analysis, the process of simplifying a complex network, a "block model", could refer to simplifying by grouping nodes into blocks based on certain characteristics or properties. This can make it easier to analyse large networks.

Ego Networks

In social network analysis, an "ego network" refers to the connections closely connected to an individual node (ego). It includes the ego (central node) and its immediate connections (alters) and provides an in-depth analysis of a social network's local structure.

Building Network Models

This term could be related to constructing models for various types of networks, such as social networks, biological networks, or communication networks. Building network models often involves defining nodes, edges, and the rules or properties that govern their interactions.

Tools that can help one estimate the social capital:

Tools used to determine social capital

Google Webmaster Tools- shows how often impressions of site pages and keywords are displayed.

Brandwatch, Radian6, and Sysomos sort online messages about a brand or a target subject on social media channels such as X (formerly Twitter), Facebook, and MySpace.

Four Square -targeting audience by location is performed.

Web Watch -monitors the integrity and availability of your site from outside by simulating the end-user actions

Check Your Progress: 4

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

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

1. What do you mean by ego networks?

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

2. What do you mean by diffusion and small world in social media?
-
-
-

9.7 NETWORK THEORIES IN CULTURAL ANTHROPOLOGY

Social anthropologist John A. Barnes 1954 first coined the term social network, which has been used to study several areas broadly specified under anthropological studies. Cognitive anthropologists are also keenly interested in correlating social network perspectives with sociological behaviour. To begin with, we can examine the effects of certain social network positions on individual perceptions. Anthropological research or in-depth study could consider family connections among members, cordial friendship ties among individuals, alliances that hold together different tribes, and socio-economic exchange among communities. This, of course, considers material gifts as well. Considering the huge shifting of relations and networks to the social media level, many anthropologists have found that the prescribed network approach to exploratory social communication is a rich and effective fieldwork for ethnographic fieldwork. Anthropologists now determine human behaviour and cultural preservation through the classical method of producing and consuming media in modern society. Indeed, social networks also play important roles in theories of social control (Hirschi, 1969), social disorganisation and collective efficacy.

9.8 SOCIAL MEDIA DATA ANALYSIS

In this section, you will learn the seemingly abstract idea of "social media analysis" in a very easy and user-friendly manner. This will help you understand the concept of social media research in communication studies. Social media, unarguably, is referred to as "people's media". We are all witnesses to the users' instant reactions to some national issues on Facebook, WhatsApp, X (formerly Twitter), YouTube, Instagram, Snapchat, SMS, etc. A few mainstream media have started devoting special space or airtime to showcase a select few posts or comments from social media. Mass media traditionally was cut off from public feedback, but social media proved to be a game changer. Hence, systematic, scientific, and machine-enabled research is the need of the hour. For the benefit of understanding, this section has been divided into the analysis of Facebook, YouTube, and X (formerly Twitter) data. A paragraph on "social network analysis", a slightly different branch of social media research, has also been included to showcase the different colours of the DMRMs.

As per Toronto Metropolitan University's Social Media Research Tool Kit, the following tools can be used to extract publicly available data from different social media account.

Name	Description	Platforms	Pricing
Gephi	Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.	N/A (Must import datasets)	Free
Hootsuite	Social Media Marketing & Management Dashboard	various	Pay
NodeXL	NodeXL is a network analysis and visualization software package for Microsoft Excel.	Flickr, X, YouTube, Wikipedia	Freemium
Brandwatch / Crimson Hexagon	Get deep insights into consumer opinion on any topic from across the social web.	various	Pay
CrowdTangle	A tool from Facebook to help follow, analyse, and report on what's happening across social media.	Facebook, Instagram	Apply
Netlytic	Netlytic is a cloud-based text and social networks analyser that can automatically summarize and discover social networks from online conversations on social media sites.	YouTube, RSS Feeds, Google Sheet, text/csv file	Freemium
Hoaxy	Visualizes the spread of information using the Twitter Search API (from the last 7 days)	X (via X's paid API)	Freemium
Talkwalker	A social media analytics and monitoring tool that keeps track of all brand-related online, print, and TV/radio content with real-time social listening and AI-driven consumer intelligence	various	Pay

Digmind	Social media analytics for businesses, and the Influencer Network, a dynamic and interactive portal that lets you identify and analyse your brand's top influencers and strongest communities on Twitter.	various	Pay
Facepager	Facepager was designed for fetching public available data from Twitter and other JSON-based API.	X, YouTube, Websites	Free
Octoparse	A no-coding web scraping tool; offers web scraping templates that can extract post and user data	various	Freemium
snscreape	Social media scraping tool	various	Free
Audiense	Identify relevant audiences, discover amazing actionable insights and inform your strategies to grow your business.	Facebook, Instagram, X	Freemium
Youtube-dl	This is a command-line program to download videos from YouTube.com and a few more sites.	YouTube, supported video sites	Free
DiscoverText	DiscoverText is a cloud-based, collaborative text analytics solution. With dozens of powerful text mining features, the DiscoverText software _x0003_solution provides tools to quickly and accurately make better decisions.	Twitter, SurveyMonkey, Text files	Freemium
Webometric Analyst	Webometric Analyst is a Windows-based program for altmetrics, citation analysis, social web analysis and webometrics, including link analysis.	X, YouTube, Tumblr, Goodreads, Flickr, Mendeley, Altmetric,	Free

		Google Books, WorldCat.	
TGStat	This website allows users to search for Telegram channels and examine general Channel analytics.	Telegram	Free
Twitter Advanced Search	Twitter Advanced search using key words, phrases and hashtags.	Twitter	Free
YouTube Data Tools	This is a collection of simple tools for extracting data from the YouTube platform via the YouTube API v3.	YouTube	Free
Media Cloud	An open source and open data platform for storing, retrieving, visualizing, and analysing online news that collects data through RSS feeds. Users can download the collected data as a CSV.	news coverage	Free
Tweet Archivist	Tweet Archivist collects data directly from Twitter based on the user's search terms. The user can also use this tool analyse the data to show information such as top users, words, URLs, hashtags, and more. Finally, users can download the dataset as either an Excel or PDF file.	Twitter (via file upload)	Pay
Social Feed Manager	Social Feed Manager is open source software for libraries, archives, cultural heritage institutions and research organizations. It empowers those communities' researchers, faculty, students, and archivists to define and create collections of data from social media platforms. Social Feed Manager will harvest	Tumblr, Flickr, Sina Weibo (latest update: 2021)	Free

	from Twitter, Tumblr, Flickr, and Sina Weibo and is extensible for other platforms. In addition to collecting data from those platforms' APIs, it will collect linked web pages and media.		
Sentiment Viz	Reddit visualisation app (sentiment, topics, heatmap, timeline, etc)	Reddit	Free
Instaloader	Captures images/videos and related metadata from Instagram	Instagram	Free
OutWit Hub	OutWit Hub explores the depths of the Web for you, automatically collecting and organizing data and media from online sources.	Internet	Freemium
4plebs	This is a community-run archive of 4chan.org (since April of 2013)	4chan	Free
InVID verification plugin	The provided tools allow you to quickly get contextual information on Facebook and YouTube videos, to perform reverse image search on Google, Baidu or Yandex search engines, to fragment videos from various platforms (Facebook, Instagram, YouTube, Twitter, Daily Motion) into keyframes, to enhance and explore keyframes and images through a magnifying lens, to query Twitter more efficiently through time intervals and many other filters, to read video and image metadata, to check the video copyrights, and to apply forensic filters on still images.	various	Free

Social Searcher	Social media monitoring (users, hashtags, terms)	various	Freemium	Social Media Analysis
InfraNodus (formerly Textexture)	InfraNodus represents textual data as a network. It uses a combination of text mining, network analysis, data visualization, NLP, and artificial intelligence to provide insights about any discourse and enhance a research process.	X, Amazon Reviews, Google Search Results	Pay	
PhantomBuster	Mainly for marketers, this is a tool that can extract social media account details for leads and build campaigns	LinkedIn, Instagram	Freemium	
Telegram Analytics (tgstat)	Search Telegram messages	Telegram	Free	
Vtracker	Tool for monitoring social media	various	Pay	
4CAT (Capture and Analysis Toolkit)	A Python-based tool that captures and processes data across different platforms; data can be downloaded for further analysis	Telegram, Reddit, Tumblr, X, 4chan, 8kun, BitChute; via Zeeschuimer: Instagram, TikTok, 9gag, Imgur, LinkedIn, Parler, Douyin and X	Free	
Visibrain Focus	Visibrain is a media monitoring tool for PR and communications professionals, used for reputation management, PR crisis prevention, and detecting influencers and trends. Visibrain's unique technology brings together online press, blogs and social media, capturing all online activity around a brand, in real time.	various	Pay	

Pulsar Social	Pulsar helps you find the story in the data by bringing all your digital audiences into one place and leveraging the power of AI and data visualization	various	Pay
Communalytic	Communalytic is a Research Tool For Studying Online Discourse. Communalytic can collect, analyse and visualize data from various social media platforms and can automatically identify toxic and anti-social interactions, map shared interests and detect signs of possible coordination among seemingly disparate actors in online discourse.	Reddit, Telegram, YouTube, X (via X paid API), CrowdTangle (Facebook & Instagram), CSV and JSON	Freemium
ArchiveSocial	ArchiveSocial automatically archives original social media posts as well as comments made by other users, allowing archive viewers to expand posts, comments, and pictures. The archive mirrors the original post, and has been used by agencies to comply with FOI requests and data archiving software-related research.	various	Pay
Twitter Audit	A tool used to check how many of your Twitter followers are real.	X	Freemium
SteamID	The website can look up steam profiles, historic profiles or game information	Steam	Freemium
Affinio	Affinio is the marketing intelligence platform that is changing the way you relate to your consumers.	various	Pay

Dash Hudson	A social media management tool that offers social listening, audience insights and trends reporting	various	Pay
Disboard	Live updates of users active in a given Discord community; reviews communities based on Discord user preferences	Discord	Free
Meltwater Social (Sysomos)	Offering a comprehensive set of solutions that support the full spectrum of your advanced social analytics and management needs.	various	Pay
PSNprofiles	The website can search publicly viewable players, daily gameplay activity, country and other stats	PlayStation Network	Free
Who posted what?	Whopostedwhat.com is a non-public Facebook keyword search for people who work in the public interest.	Facebook	Free
KarmaDecay	Reverse image search of Reddit (beta)	Reddit	Free
Facebook Video Downloader	Downloads Facebook videos	Facebook	Free
Story Saver	Download Instagram Stories, Highlights and Videos Online Easily with one simple click.	Instagram	Free
Open Measures (formerly, Social Media Analysis Toolkit or SMAT)	Analyse & visualize harmful online trends such as hate, mis-, and disinformation online	Gab, Parler, 4chan, 8kun, Telegram, Gettr, Win Communities, Poal, Rumble, etc	Free
Facebook Video Downloader	Download Facebook Videos Online	Facebook	Free

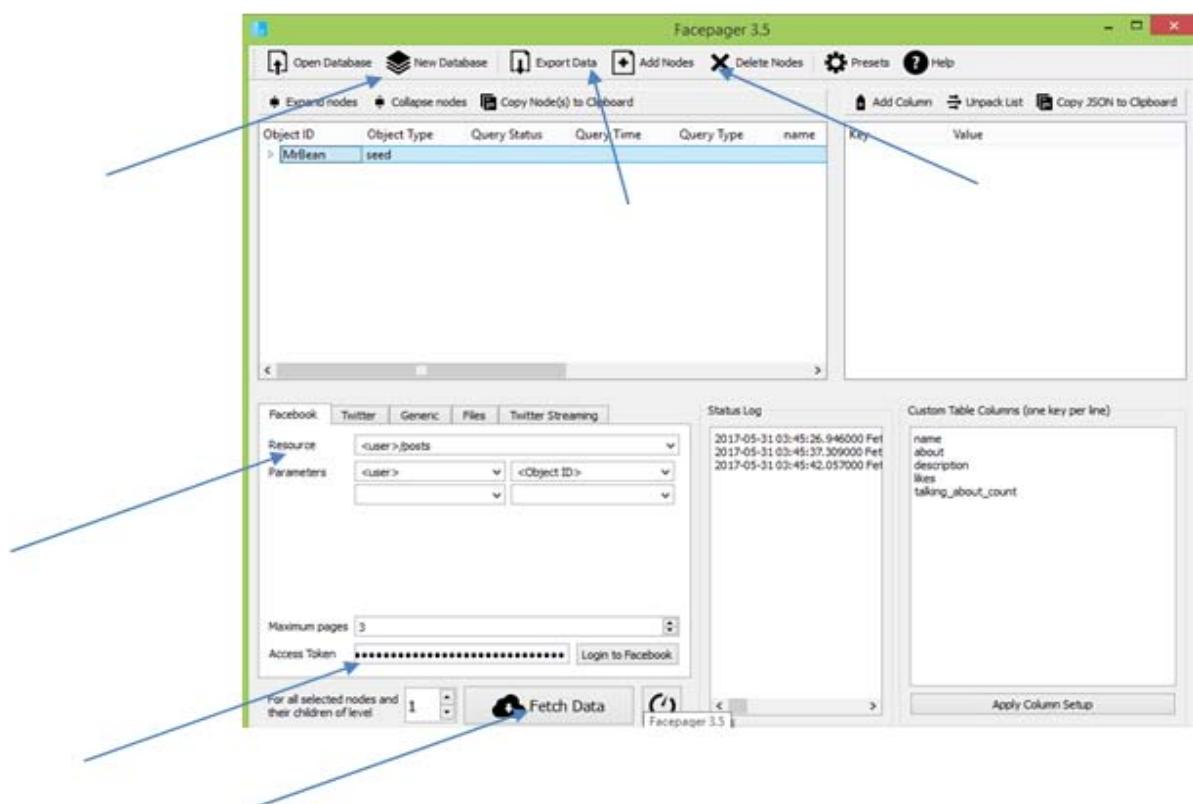
Skopenow	Collecting and analysing publicly available information from data sources including social media, the dark web, associated vehicles, court records, and contact data; Link Analysis feature generates a network and other visuals	various	Pay
Telemetr.io	The website provides Telegram channel analytics	Telegram	Freemium
Youtube Geofind	YouTube Geofind; three different search functions, location, topic, and channel.	YouTube	Free
Twitter Video Downloader	Download twitter videos & GIF from tweets	X	Free
tlgrm.eu/channels	This website allows users to search for Telegram channels	Telegram	Free
Redective	Searches for profile and engagement information of a Reddit user	Reddit	Free
TelegramDB	Telegram channel search engine	Telegram	Free
Geogramint	Locates Telegram users and groups based on geographic location (only users opted into the nearby feature)	Telegram	Free
Telepathy	The tool is described as the "Swiss army knife of Telegram tools," allowing analysts and researchers to archive Telegram chats (including replies, media content, comments and reactions), gather member lists, lookup users by given location, analyse top posters in a chat, map forwarded messages, and more.	Telegram	Free

metatargetr	The goal of metatargetr is to parse targeting information from the Meta Ad Targeting dataset as well as retrieve data from the Audience tab in the Meta Ad Library.	Facebook	Free
TikTokApi 6.2.0	Unofficial TikTok API in python that can retrieve video, hashtag, user data	TikTok	Free

Source: <https://socialmedialab.ca/apps/social-media-research-toolkit-2/>

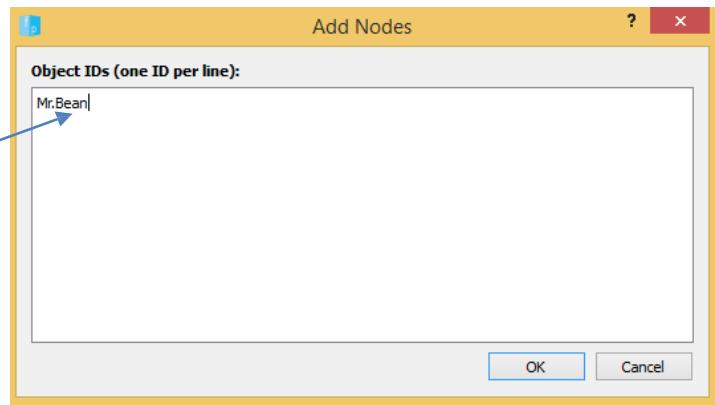
9.8.1 Facebook Data Analysis

If it were the nation, Facebook would have been the biggest one on earth. It comprises personalised Facebook pages, posts, likes, comments, photos, and videos. The researchers can harvest this rich data for insights. The biggest issues in social media data harvesting pertain to digital skills, i.e. technical knowledge, price issues, limitations of the free wares, multidimensional data treatment, etc. We would scrape the data from the posts here by taking Mr Bean's Facebook ID (among the world's top commented ones) at <https://www.facebook.com/MrBean>. A freeware from www.facepager.software.informer.com/3.5/ needs to be downloaded.



After installing the face page, click the "new database" as shown in the arrow. Then click Add Nodes. Copy the ID from the Facebook page, e.g.

www.facebook.com/Mr.Bean. Here, **Mr. Bean** is the ID. Insert it in the node as shown below. Multiple IDs can be inserted at once, one ID per line.



Now go to the source and select the <user>/posts. Afterwards, you can get the access token from the Facebook page by logging in to your account. Access tokens are security identifiers from your side, and social networking sites are quite careful about who uses their data. There is a limit to the posts or data that can be retrieved in 24 hours, but even this much data is sufficient for a good study. Assuming that the access token is obtained and entered as shown. Then click "fetch data", and the posts will be retrieved. When the node is expanded downwards, the data is visible, as shown. Once sufficient data is collected, select it by Shift + click and then click "export data". The data would be exported to CSV format and then converted into Excel format. The data can be converted into a text file in UTF-8 format and analysed in corpus analysis software like AntConc, as described in Part 3 of the DMRMs. The face pager is very easy to use as it has various data harvesting options.

For Example - comments, photo likes, comments on posts, video comments, and video likes, and it has almost all the functions that researchers can think of. However, since no single tool can serve every purpose of the study, Face pagers can't gather data beyond Facebook IDs. Companies, governments, and individuals have individual Facebook pages like personal domains to reach the world with a single click. Another free tool must be tried to collect data in the same manner as before. There is a disclaimer here that no standard software selection for data retrieval exists. Many other paid software can nicely do the same job, but here, open source and free tools are encouraged for academic purposes. Also, new APIs (application programming interface), access tokens, etc., keep on coming, and the already usable free software may not work in the changed scenario. In this case, a researcher is required to have good knowledge of digital skills and digital literacy. Hence, for collecting data from Facebook Pages, a non-programming exercise needs to be performed, and it would bring each Facebook post's number of likes, shares and comments, the number of Facebook reactions (Like, Love, Wow, Haha, Sad, Angry), comments and Facebook IDs of the users who commented, shared, or liked (privacy issues apply here). You can click the link: <https://nocodewebscraping.com/how-to-extract-data-from-facebook-page-competitor-analysis/>. This link would open a web page where, by following a few steps, you can harvest the data in the following manner.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
message	link	created_time	type	id	shares/count	reactions/	reactions/	likes/sum	likes/sum	comment/	comment/	comment/	comment/	name
1 MBEYA CITY YOUR	https://w/2016-08-18T19:video	204153042	15	379	NONE	378	TRUE	FALSE	ranked	23	TRUE			
3 TANGA!!!! TANGA!!!!	vhttps://w/2016-08-18T17:video	204153042	42	1076	NONE	1069	TRUE	FALSE	ranked	75	TRUE			
4 Mda mchache leo nli	https://w/2016-08-17T19:photo	204153042	328	17016	NONE	16845	TRUE	FALSE	ranked	829	TRUE	Timeline Pho		
5 KENYA!!!! KENYA!!!!	yhttps://w/2016-08-17T12:photo	204153042	153	4532	NONE	4495	TRUE	FALSE	ranked	208	TRUE	Timeline Pho		
6 My KENYAN people!!	https://w/2016-08-17T08:video	204153042	161	3535	NONE	3464	TRUE	FALSE	ranked	182	TRUE			
7 A Big Birthday	https://w/2016-08-16T11:photo	204153042	55	3019	NONE	3005	TRUE	FALSE	ranked	150	TRUE	Timeline Pho		
8 Watora Mari Behind	https://w/2016-08-15T16:video	204153042	197	2768	NONE	2734	TRUE	FALSE	ranked	172	TRUE			
9 KENYA!!! Your favonit	https://w/2016-08-15T15:photo	204153042	322	8079	NONE	8004	TRUE	FALSE	ranked	324	TRUE	Timeline Pho		
10 MOMBASA!!!! @RICHMAVOKO	2016-08-14T14:status	204153042	81	4697	NONE	4643	TRUE	FALSE	ranked	435	TRUE			
11 https://youtu.be/U8	https://w/2016-08-14T12:video	204153042	534	5184	NONE	5073	TRUE	FALSE	ranked	351	TRUE			
12 Jana Usiku Msanii @ja	https://w/2016-08-13T14:video	204153042	152	3613	NONE	3565	TRUE	FALSE	ranked	309	TRUE			
13 https://youtu.be/U8C	https://yc2016-08-13T01:video	204153042	346	6198	NONE	6090	TRUE	FALSE	ranked	433	TRUE	Jah Prayzah F		
14 Add me on Snapchat	https://w/2016-08-12T04:video	204153042	98	3469	NONE	3437	TRUE	FALSE	ranked	263	TRUE			
15 THE VIDEO IS OUT NO!	https://w/2016-08-11T17:video	204153042	650	5961	NONE	5878	TRUE	FALSE	ranked	296	TRUE			

The methods mentioned above have a few limitations. First, there is a limit to the quantum of data extraction as per the API limit. The common restrictions associated with freebies apply here also. Second, the above freeware can collect data, as it is, from a Facebook ID or Page, which means you can't collect data as per your own choice of words or query, e.g. harvesting all the Facebook posts or comments which contain "demonetisation" and India is not possible. To overcome this problem, Google Power Search commands (**site:facebook.com demonetisation OR demonetization AND India**) are a powerful method that brings 1,53000 results.

The number of posts can be randomised, and content can be saved manually or by using the Google Chrome extension (link clump). This would harvest the data per the posts' headline and the corresponding URL, as shown in the Table. The latter can be opened separately in browsers, and the content can be copied and pasted.

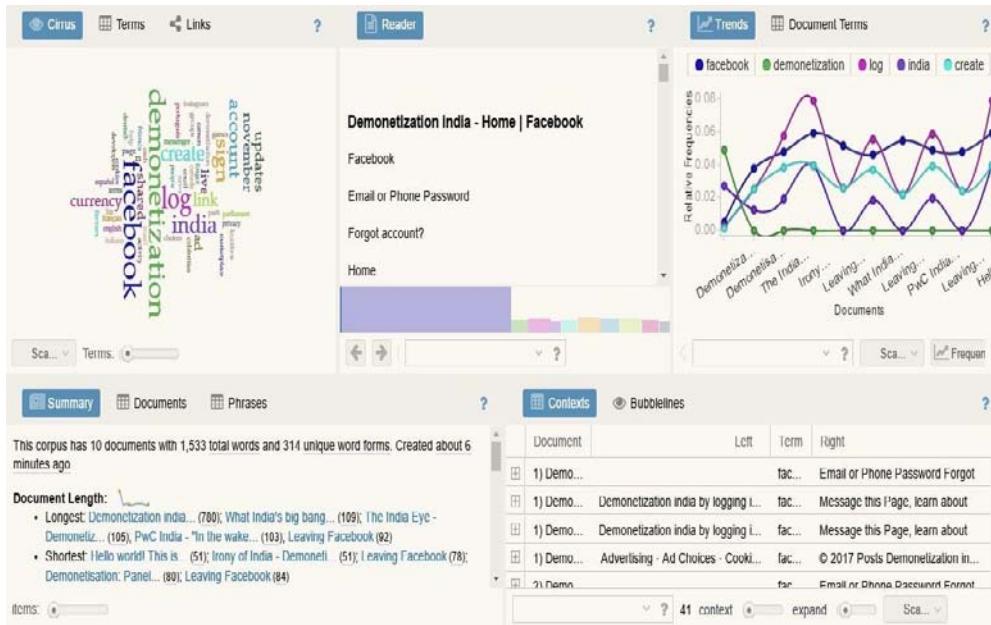
Hello world! This is a documentary on	www.facebook.com/indiaindetails/videos/22740703436291
Our new video is about Demonetization...	www.facebook.com/IndiaBakchod/videos/1287137644663
Demonetisation – Shook the country...	www.facebook.com/natgeotv.india/videos/1015565551643
Indians for Actualization of Democracy Public Group	www.facebook.com/groups/demon2016/
What India's Big	www.facebook.com/wsjinida/posts/977992965638557

Bang demonetisation... -	
Demonetisation: Panel Discussion convened by Oxford India Centre	www.facebook.com/events/211008546012741/
MarkandeyKatju - Demonetization India is not a developed...	www.facebook.com/justicekatju/posts/1477368848970323
Indian Economy Article:- Demonetisation...	www.facebook.com/pratiyogitarpaneng/posts/10154847
Demonetisation India - Home	www.facebook.com/Demonetisation-India-7460140322182

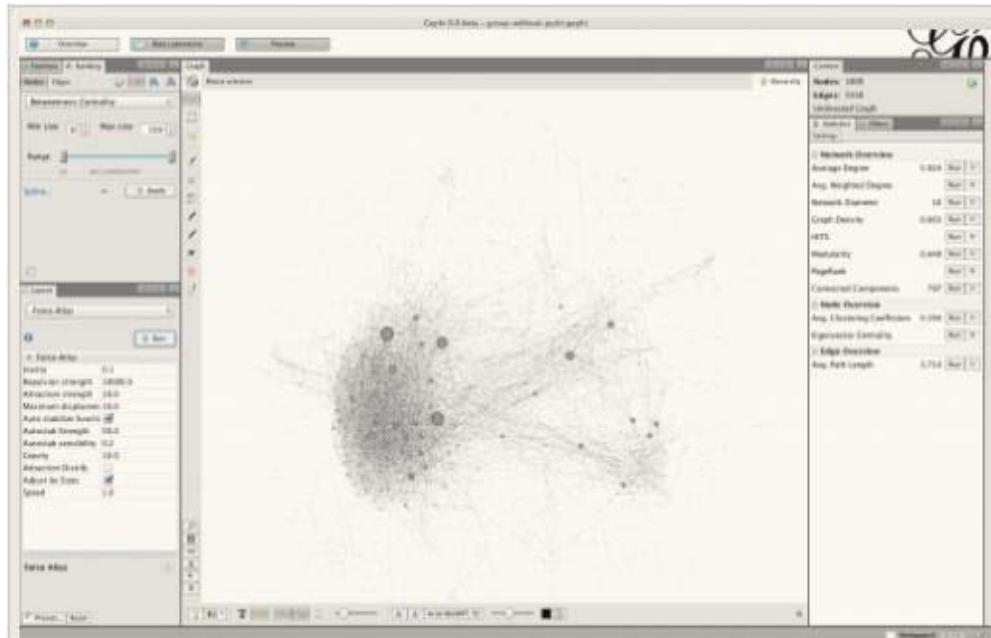
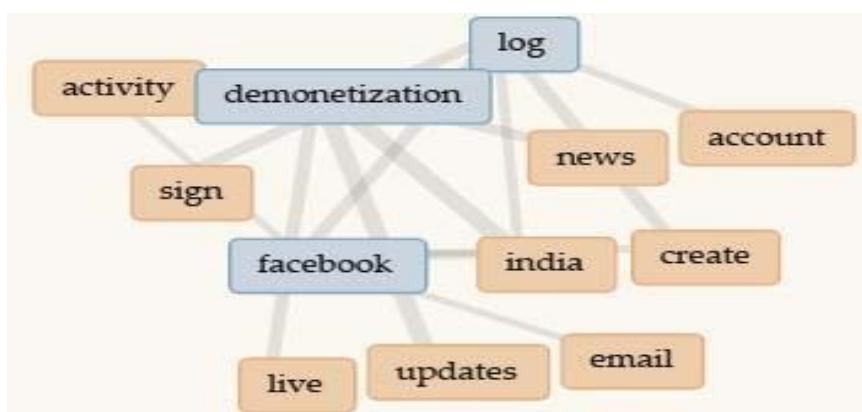
Another automated method is to open the content of these URLs all at once by pasting in www.voyant.com



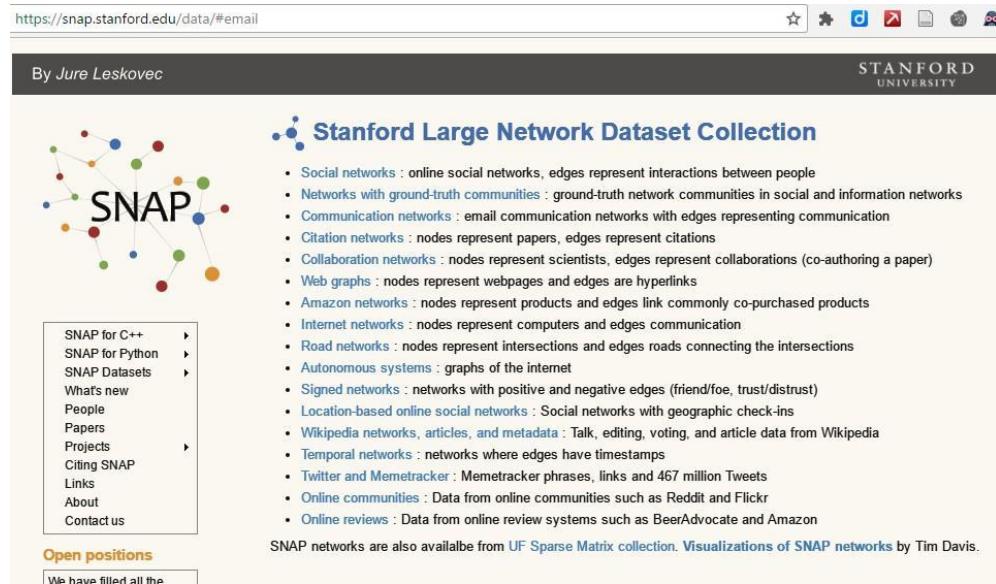
The corpus of Facebook posts or comments being searched this way presents multiple data analysis and interpretation styles, both visual and text. Moreover, Facebook data in Hindi, Punjabi and other regional languages can be collected and analysed easily. The site even generates an interaction of selected terms (e.g. demonetisation or BJP) with other words in the bubble form. The links between the objects show a dynamic interaction between the actors in the social media data. The other functions and interpretations of the data are the same as those of the "Digital Data Analysis" chapter of DMRMs Part 3 (Corpus method of analysis).



Another deep, insightful way of looking at Facebook's social media data(social network analysis) is downloading your own Facebook data as a PDF file (using the Netvizz App) and opening it in Gephi (www.gephi.org/users/download/). Once open, the network may look like the one in the next image.



The circles depict the nodes (people), and the arrows between them show the "interaction" between the nodes. The direction of arrows indicates the direction of communication from A to B nodes or vice versa. There may not be single arrows between two or more nodes, meaning there is no communication between these groups of nodes (people). Gephi is an open-source and free software that can map the interaction between people on Facebook. However, its limitation is that it can only analyse your network, and issues related to privacy, permissions and law emerge while analysing other people's or organisations' networks. However, to promote social network analysis, Stanford University has created a few open network datasets available at <https://snap.stanford.edu/index.html>. "Social network analysis is the mapping and measuring relationships and flows between people, groups, organisations, computers or other information/knowledge processing entities." (Valdis Krebs, 2002). Social Network Analysis (SNA) is a method for visualising our people and connection power, leading us to identify how we can best interact to share knowledge" (ibid). Gephi can also analyse the non-Facebook network dataset, as shown below. It has many more features, like measuring the strength of connection, identifying the community in the network, and ways to spread the message quickly in a network. You can choose a good video tutorial available on www.youtube.com for more detailed functions of Gephi.



With the new technological developments, ready-made network data sets of Facebook interaction, X (formerly Twitter), worldwide communication patterns of e-mails, and the like may soon be available in the public domain. These data sets can enrich our insight into social media analysis through social network analysis. A Google Power search may reveal a few of them that are already available for free download and distribution.

9.8.2 YouTube Data Analysis

A very simple analysis of harvesting the YouTube Comments, video views, likes, video description, duration, etc., can be done through <https://ytdt.digitalmethods.net/>

YTDT Video Info and Comments

This module starts from a video id and retrieves basic info for the video in question and provides a number of analyses of the comment section. Comments are retrieved via the `commentThreads/list` API endpoint.

The number of comments the script is able to retrieve can vary wildly. In some cases, only a relatively small percentage is made available, while in others well over 100.000 comments have been successfully retrieved. This seems to be mainly related to the age of the video in question.

The module creates the following outputs:

- a tabular file containing basic info and statistics about the video;
- a tabular file containing all retrievable comments, both top level and replies;
- a tabular file containing comment authors and their comment count;
- a network file (gdf format) that maps interactions between users in the comment section;

The first three elements can be shown directly in the browser by enabling HTML output.

video id: (video ids can be found in URLs, e.g. <https://www.youtube.com/watch?v=aXnaHh40xnM>)

HTML output: (adds HTML result tables in addition to the file exports)

This website also allows you to export the comments network as a PDF file for analysis in Gephi.

Note: -

1. X (formerly Twitter) allows only 7-10 days old tweets by freeware, and you may have to pay for old tweets.
2. Data extraction from Facebook is more challenging than YouTube and X (formerly Twitter).
3. Every type of social media analysis must be done using external text analysis software (corpus) or specialised software like Mozdeah, which is only for X (formerly Twitter).
4. A random search on Google would show many websites that can analyse X (formerly Twitter) data for trending hashtags, X (formerly Twitter) statistics, the sentiment of any person or product on X (formerly Twitter), etc., but most of them (free) do it with a very small sample of tweets, and paid software provides exhaustive and deep analysis.
5. For collecting older tweets, Google power searching commands can be applied, just like the one applied in the case of Facebook above. This way, the old tweets can be retrieved free of cost, but they are slightly difficult to harvest because of their format.

Ethical Guidelines in Social Media Research – A researcher must deal with this very carefully since network data is people's data. "One of the biggest areas of concern with social media data is the extent to which such data should be considered public or private data" (Townsend & Wallace, n.d.). The process of evaluating the research ethics cannot be ignored simply because the data are seemingly public" (Boyd and Crawford 2012, p672). Whether online postings are public or private is determined to some extent by the online setting itself and whether there is a reasonable expectation of privacy on behalf of the social media user (British et al., 2013). Though data is shared publicly on networks for general viewing, it is always advisable to mask the identity of the user's IDs for the loss of reputation, motive

attribution, and public shame. Hence, social media analysis tends to compromise the anonymity of the users. The same is well guarded in the traditional method of research. Exposing the identity may result in embarrassment and even prosecution. Data belonging to the vulnerable sections of society, like children, women, and minorities, warrants extra care. When reporting the results or referring to any individual post, the researcher should think about either paraphrasing the same or removing the user IDs or both. Alongside this, the researcher must consider the legal issues and the terms and conditions of data extraction. The data extraction must follow the institution's ethical guidelines, the funding agency's policies, and the data policies of the land. According to the Association of Internet Researchers (2012), no set of Internet research guidelines can be static because technologies and the way that technologies are used are constantly changing. Hence, the researcher, the custodian of the public's data, is the trustee, too.

9.9 LET US SUM UP

In this Unit, we have discussed the key features of social media analysis and its characteristics. We have also discussed how data is collected and managed, its significance, and the challenges of data management. The Unit has defined concepts of social cohesion, the different centralised and decentralised elements of social networks, and the different terminologies associated with social media network analysis. We also briefly discussed audience perception and network theories connected to social anthropology.

9.10 KEYWORDS

Structural Holes: In social media, the gaps or spaces between different clusters or groups of users who share a common platform are called structural holes. Platforms can facilitate connections across these structural holes, enhancing diverse interactions and information flow.

Simmelian Ties: Strong relationships between individuals or within specific online communities that develop close-knit connections signify simmelian ties.

Diffusion: A viral content or trend that quickly progresses through the social media user base is called diffusion. It usually refers to the spread of information, trends, or novel ideas through the network.

Small World: The "small world" concept, defined within the domains of social media, suggests that individuals are interrelated through surprisingly diminutive chains of acquaintances. Concepts like friend recommendations or mutual contacts positioned in the web world draw attention to this concept of social networks' small-world nature.

Reciprocity: A sense of mutual link-up prevails in social media, where if one user follows another, there's a huge possibility of a reciprocal action. Briefly, in social media, reciprocity correlates to common connections, where users, whenever possible, often follow back those who follow them.

Social Support: Social media provides a platform for individuals to offer

and receive emotional, informational, or instrumental support. This support can come from close ties or broader network connections, contributing to users' well-being.

9.11 FURTHER READINGS

1. Chamria, R. (2022, June 24). Decentralised social media: Will it replace popular centralised social media sites? Blockchain Deployment and Management Platform | Zeeve. <https://www.zeeve.io/blog/decentralized-social-media/>
2. Freeman, L. C. (2014). The development of social network analysis – with an emphasis on recent events. The SAGE Handbook of Social Network Analysis, pp. 26–39. <https://doi.org/10.4135/9781446294413.n3>
3. Ganis, M., & Kohirkar, A. (2015). Social media analytics: Techniques and insights for extracting business value from social media. IBM Press.
4. Hunsinger, J., Klastrup, L., & Allen, M. (2010). International handbook of Internet research. Springer Science & Business Media.
5. McGloin, J. M., & Kirk, D. S. (2009). Social network analysis. Handbook of Quantitative Criminology, pp. 209–224. https://doi.org/10.1007/978-0-387-77650-7_11
6. McGloin, J. M., & David S. Kirk. (2010). "An Overview of Social Network Analysis." Journal of Criminal Justice Education 21(2): 169–181. <http://dx.doi.org/10.1080/10511251003693694>
7. Murray, L. (1999). Final report on social cohesion. Ottawa: The Standing Senate Committee on Social Affairs, Science, and Technology.
8. R. Jon McGee & Richard L (2016). Theory in Social and Cultural Anthropology: An Encyclopedia, L. Theory in Social and Cultural Anthropology, Sage Publications.
9. Rombach, M. P., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2012). Core-periphery structure in networks. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2002684>
10. Sajithra K, S. K. (2013). Social media – History and components. IOSR Journal of Business and Management, 7(1), 69–74. <https://doi.org/10.9790/487x-0716974>
11. Sponder, Marshall. (2014). Social Media Analytics: Effective Tools for Building, Interpreting, and Using Metrics.
12. Wasserman, S., & Galaskiewicz, J. (1994). Advances in social network analysis: Research in the social and behavioural sciences. SAGE Publications.
13. Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge University Press.

9.12 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Social media analysis is an up-and-coming area that is largely the by-product of connectivity and interactive tools and sites. They are the core of studying issues about lifestyles and activities in the digital realms capable of; there is so much misinformation, fragmentation, and confusion within the marketplace today that it is very difficult to know how to choose the right platforms and how to set up the right processes to achieve our goals. Social Media Analytics is for anyone who seeks the right technologies to implement social listening and measurement programs.

Check Your Progress: 2

1. Data management in social media refers to the strategies, processes, and practices crucial for managing and systematising the vast amount of data generated through social media platforms. The core elements of data analysis are employing tools and techniques to analyse social media data and derive meaningful insights. In the age of Artificial Intelligence, using algorithms to predict trends, sentiment, and user behaviour based on social media data can help reach the audience and generate an idea for the consumer-driven world. This is the most important part of the entire data-management process. Prompt data analysis requires marketing and technical skills to know what the numbers mean, and which are important. The marketing analytics process leads to suggestions for enhancement, planning new programs, and creating new budgets more sustainable to the market demands. Understanding the social media equation is crucial in a market-driven world where stringent marketing programs must allow marketing managers to update their cases to the market's demands, leading to a decent sales pitch.

Centralisation: A centralised network architecture is developed around a single server that handles all the major processing. Less powerful workstations connect to the main server and submit their requests to it rather than performing them directly. This can include applications, data storage, and utilities. Centralised networks do have their limitations, too. For example, a single point of failure can be a big risk factor for organisations. The centralisation of social media can be compromised in terms of censorship and integrity of content created by users. Examples of centralised social media are Facebook, Instagram, and LinkedIn. **Decentralised** refers to the social networks that operate on independently run servers rather than on a centralised server without the role of business ownership, and here, no single authority holds the data. Rather, it is decentralised over connection nodes among the networks. This, in turn, creates room. Despite the advantages of decentralised social media networks, there are apparent possibilities of attacks orchestrated by content creator communities that could lead to a large extent of loss of information. Mastodon is an example of such a

decentralised network. It uses free, open-source program code. Another good example is Steem, which provides a social blockchain.

Check Your Progress: 3

1. Social capital refers to the resources and advantages individuals or groups can access through social networks and relationships. It encompasses trust, cooperation, and shared values within a community or social group. The two main elements of social capital are:

Bonding Social Capital: The connections and relationships within a close-knit group of people sharing similar characteristics, such as family, friends, or members of a particular community, are essentially bonding. This creates a deeper sense of relationships while making available support within the group.

Bridging Social Capital: Social groups or individuals with various backgrounds exist on social media. Bridging social capital refers to where connections are built, fostering interaction and cooperation across different communities, leading to a more all-encompassing and diverse network.

Check Your Progress: 4

1. In social network analysis, an "ego network" is the connections closely connected to an individual node (ego). It includes the ego (central node) and its immediate connections (alters), providing an in-depth analysis of the local structure of a social network.
2. Diffusion refers to the spread of viral content or trends that quickly progresses through the user base in social media. It usually refers to the spread of information, trends, or novel ideas through the network. The "small world" concept, which is defined within the domains of social media, suggests that individuals are interrelated through surprisingly diminutive chains of acquaintances. Concepts like friend recommendations or mutual contacts positioned in the web world draw attention to this concept of the world nature of social networks.

Block

3

INTERNET RESEARCH METHODS-II

UNIT 10

Web Archiving

UNIT 11

Textual Analysis

UNIT 12

Online Ethnography

UNIT 13

Narrative Analysis

BLOCK 3 INTRODUCTION

Block 3 continues the earlier one on various data collection tools. Studying Internet histories is one of the unique opportunities for data collection of past data from online spaces. Similarly, textual, and narrative analysis will enable research on Web 2.0-based interactions. As Online Communities are yet another unique feature of cyberspace, online ethnography will help you unravel the nuances of interest groups.

Unit 10 Web Archiving - Understanding Internet Histories and Beyond" explains the intriguing world of preserving digital footprints. We will explore the origins of web archiving, its pivotal role in Internet history research, ethical considerations, and its significance for media professionals.

Unit 11 Textual Analysis explores perspectives, types, and presentation methods. We'll learn to understand diverse viewpoints on analysing texts, examine various types of texts and how to collect them and learn effective strategies for presenting our textual analysis.

Unit 12 Online Ethnography, explains the digital landscape, exploring the necessity of online ethnography. We'll examine qualitative and quantitative methods, weighing their advantages and disadvantages. Moreover, we'll discuss the ethical considerations and responsibilities inherent in this fascinating field.

Unit 13 Narrative Analysis - We'll explore its definition and functions, delve into its social dimensions, discuss methods for collecting narratives, and learn techniques for analysis. Through an example, we'll illustrate how narrative analysis unveils layers of meaning within personal and collective stories.

These Units would give you a broad perspective on exploring the various dimensions of cyberspace.

UNIT 10 WEB ARCHIVING

Structure

- 10.0 Introduction
- 10.1 Learning Outcomes
- 10.2 Understanding Internet Histories as a Research Methodology
 - 10.2.1 Introduction to Internet Histories
 - 10.2.2 Areas of Interest in Internet Histories
 - 10.2.3 Research Trends in Internet Histories
- 10.3 Web Archiving as a Tool for Internet Histories
 - 10.3.1 What is web archiving?
 - 10.3.2 Basic Concepts in Web Archiving
 - 10.3.3 Types of web archiving
 - 10.3.4 Why Archive Websites
- 10.4 Historical Context of Web Archiving
 - 10.4.1 Early efforts and challenges
 - 10.4.2 Milestones in the development of web archiving
- 10.5 Basics of Web Archiving Tools
 - 10.5.1 Introduction to user-friendly tools for archiving (e.g., browser extensions)
 - 10.5.2 How professionals can integrate archiving into their workflow
 - 10.5.3 Web archiving features, strengths, and limitations.
- 10.6 Records and Information Management
 - 10.6.1 Websites as records
 - 10.6.2 Selecting and collecting
 - 10.6.3 Hints and tips on saving and archiving
- 10.7 Ethical Considerations and Legal Framework
 - 10.7.1 Legal and Ethical Issues in Web Archiving
 - 10.7.2 Copyright, fair use, and permissions in the digital age
 - 10.7.3 Balancing the public's right to know with privacy concerns
- 10.8 Web Archiving for Media Professionals
 - 10.8.1 Why is it important for media professionals?
 - 10.8.2 Introduction to preserving digital content for journalistic and media purposes.
 - 10.8.3 The need for archiving in the context of digital media
- 10.9 Future of Web Archiving in Media
 - 10.9.1 Emerging technologies and trends in archiving
 - 10.9.2 Opportunities for media professionals in the evolving digital landscape
- 10.10 Let Us Sum Up
- 10.11 Keywords
- 10.12 Further Readings
- 10.13 Check Your Progress: Possible Answers

10.0 INTRODUCTION

Web archiving refers to collecting and storing web pages and websites to ensure their availability for future access. The principal objective of web archiving is to preserve the content and structure of websites over time, allowing researchers, historians, and the public to access and study the evolution of the new medium. Web archiving involves using dedicated tools and technologies to capture and save the content of web pages, including text, images, videos, and other multimedia elements.

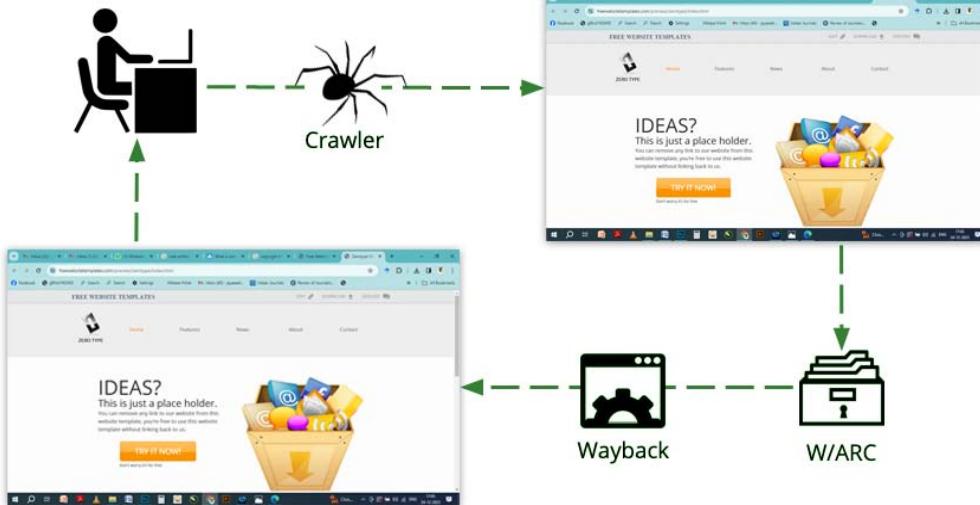
Imagine web archiving as a way to capture and save a snapshot of websites so that people in the future can explore them. It is analogous to taking pictures of web pages, meticulously preserving how they appeared and functioned during a specific period.

The dynamic nature of websites heightens the intricacy of web archiving. They undergo constant updates, relocations, or redesigns, requiring a continuous and ongoing archiving effort. The backbone of web archiving is the technique known as "crawling." This involves using specialised tools to systematically navigate the links on web pages, generating copies of their content. Typically beginning from a chosen web page, which often is the homepage, the process aims to hoard a diverse collection of web pages. However, challenges arise in capturing content concealed behind logins or interactive elements like videos. They add a layer of Complexity to the archiving process. Regular crawling of pages is significant in skilfully capturing the evolving iterations they naturally undergo. The objective extends beyond preserving visual aesthetics; it encompasses the absolute conservation of functionality, providing a holistic representation of the Internet's evolution.

Archived pages find a home in specific file formats such as WARC (Web ARChive) or WACZ (Web Archive Collection Zipped), ensuring their enduring existence. Equally significant is the accompanying metadata, a trove of information guiding archivists and future users. This metadata searches the inner workings of a website and bears a timestamp, which is important for authenticating the archived resources.

Imagine you have a tool that lets you "play back" these archived files—a digital time machine for the Internet. This tool facilitates seamless interaction with archived websites, mirroring their original functionality on the live web. It transcends the mere preservation of visual elements, encapsulating the website's underlying code, images, videos, and design aspects. This meticulous process is not just about saving pictures; it is about preserving the entire immersive experience, allowing future generations to delve into and comprehend the evolution of the Internet over time.

Web archiving is a careful endeavour to capture the essence of the Internet, employing crawling techniques to navigate the dynamic web landscape. Archived pages, stored in specific formats with accompanying metadata, serve as a time capsule and playback tools enable users to explore and interact with the preserved digital heritage of the web.



Keeping Website Life in Web Archiving

10.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Develop a comprehensive understanding of Internet histories as a research methodology.
- Gain in-depth knowledge of the historical context and basics of web archiving tools.
- Explore principles of records and information management in web archiving.
- Appreciate ethical considerations and legal frameworks associated with web archiving.
- Recognise the application of web archiving in media professions.
- Envision future trends and developments in web archiving within the media landscape.

10.2 UNDERSTANDING INTERNET HISTORIES AS A RESEARCH METHODOLOGY

Studying Internet Histories is like exploring a big online storybook. Instead of just seeing the Internet as a tool, researchers look at it as something influenced by people and cultures. They dig into digital archives, like online libraries, to see what individuals, groups, and organisations have done online. They use knowledge from history, communication, media studies, and technology to get the whole picture. The goal is to find stories in these digital archives and determine how society and culture have influenced the development of the Internet. It's like decoding how the Internet has changed over time.

10.2.1 Introduction to Internet Histories

Exploring the Internet's history is like diving into a big sea of information using powerful digital tools like "web archives." Instead of just keeping records, the Internet is seen as a cultural force shaping how we remember our

online lives from the 1990s. "Internet Histories" uses knowledge from different areas to understand how the Internet has made our daily lives. It is like looking in a mirror to see how the Internet significantly impacts our future, based on past events, and carefully saved in the Web's vast archives.

10.2.2 Areas of Interest in Internet Histories

Research in Internet Histories encompasses diverse areas, reflecting the multifaceted evolution of the Internet and its deep impact on various aspects of society. Here are key areas of interest within Internet histories:

Digital Research Methodologies:

The Internet, becoming a big information source, changed how we study history. Now, we use computers and digital tools for research, writing, and analysis. This shift is called digital humanities.

Impact on Historical Narratives:

Web archives are changing history by including more voices. Now, we hear from people often left out in the past, making our understanding of the past more complete and fairer.

Societal Implications:

Delve into the societal implications of preserving web archives, including how these archives contribute to the advocacy for justice, equality, and informed policy decisions. Explore how access to historical digital records influences a deeper societal understanding.

Interdisciplinary Perspectives:

Explore the interdisciplinary nature of Internet Histories, considering insights from history, communication studies, media studies, and technology studies. Investigate how this approach provides a nuanced understanding of the Internet as a cultural phenomenon.

Forces Shaping the Digital Present and Future:

Examine how Internet Histories serve as a mirror reflecting the profound impact of the Internet on various aspects of our lives, including work, interaction, and societal structures from the past to the future.

10.2.3 Research Trends in Internet Histories

Research Trends in Internet Histories encompass a dynamic array of methodological approaches aimed at unravelling the complex evolution of the digital realm. Delving into this domain likely involves discussions on various research methodologies, offering insights into how scholars navigate the challenges of studying Internet histories.

Archival research takes centre stage, enabling the exploration of historical websites, forums, and digital artefacts that shape our online past.

Qualitative methods, such as ethnographic studies and content analysis, provide a lens into Internet histories' cultural and social dimensions.

Quantitative web history methods involve the systematic analysis of large-scale datasets to uncover patterns, trends, and statistical relationships within the historical development of the web.

Longitudinal studies play a crucial role, providing a temporal dimension that tracks changes over time and addressing the challenges inherent in researching a rapidly evolving digital landscape.

User-centred approaches recognise the agency of individuals in shaping Internet narratives, incorporating user experience research and participatory methods.

Network Analysis is a powerful research method within the social sciences that can be applied to study web history. It involves examining and visualising relationships and interactions among various entities within a network.

Check Your Progress: 1

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

2) Compare your answers with those given at this Unit's end.

1. Explain the significance of web archiving in preserving online content.

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

10.3 WEB ARCHIVING A SA TOOL OF INTERNET HISTORIES

Web archiving is vital for Internet Histories, preserving online content for researchers and the public. Archives are a rich resource for history, sociology, media studies, and computer science. They're not just for academics. They also assist in legal cases, business continuity, and cultural preservation. Archived websites offer evidence in legal matters, aid disaster recovery, and protect digital artefacts showcasing societal trends, culture, and history. This is essential for understanding the transformations in our online world.

10.3.1 What is web archiving?

Like preserving paper documents, web archiving involves gathering and safeguarding websites from the vast World Wide Web. 'Web archivists' manage these archives, ensuring access for various groups, including government, businesses, researchers, and the public. Automated tools like web crawlers are used to collect and save information for archival purposes. Archived websites, accessible online, provide a snapshot of their live counterparts, allowing navigation and viewing at specific points in time. Some of us use simple tools for internal archiving, while national libraries and diverse groups focus on preserving culturally significant web content. We must remember that commercial solutions cater to businesses' specific

archiving needs, collectively forming a digital library that documents the Internet's evolution for exploration.



10.3.2 Basic Concepts in Web Archiving

The Internet changes a lot, and sometimes, we can lose important stuff because of updates, technical issues, or purposeful removal. Web archiving stops this by preserving content for ongoing access. It is not just about saving simple files; it also captures the live web and tracks how websites work. Copies are saved in standard formats (WARC or WACZ) with metadata for proof. Archivists use this to verify and load files into playback tools, letting users experience websites as if they were live, not just pictures. This keeps web history safe and deals with challenges.

10.3.3 Types of web archiving

Websites can be archived in three primary ways. (i) **Client-side web archiving** involves copying website content to a user's computer, making it the most common and readily achievable method. (ii) **Transaction-based web archiving** captures all website traffic, providing a complete record but requiring association with the website owner. (iii) **Server-side web archiving** entails installing a backup system directly on the website. This is the most accurate but costly approach. Unlike website backups, which are solely intended for restoring lost or damaged sites, web archiving enables users to access and steer through archived websites as if they were still active.

10.3.4 Why archive websites?

Websites are most important for organisations because they have crucial data. However, the Internet changes frequently, and websites can be updated or deleted quickly. Web archiving solves this by saving websites as they were at a particular time. Unlike regular backups, it keeps the Internet real. Archiving

lets organisations retain old information not found on current sites. It helps save important history and shows that an organisation manages information well, making it look good.

10.4 HISTORIES CONTEXT OF WEB ARCHIVING

Back in the early 1990s, the World Wide Web had quickly reached the hands of all and sundry. In 1996, Brewster Kahle founded the Internet Archive. His team was the first to save websites using slow Internet. National libraries like the Library of Congress and the British Library started saving important websites in the late 1990s. A bunch of people, including Herbert Van de Sompel, teamed up in 2001 for the "Harvesting for the Future" project, making saving websites easier. In 2003, the Memento project started capturing web pages at specific times, making it better to see how the web changed.

10.4.1 Early efforts and challenges

Facing technical challenges, projects like the Memento and WARC format deal with the need to save tricky web content. They also manage copyright and big data with tools like Hadoop. The International Internet Preservation Consortium promotes global teamwork and shares the best ways to do things. Despite early issues, these projects lay the groundwork for web archiving, showing how working together is powerful. Web archiving keeps changing with new technologies, securing our digital history for the future. Early challenges build the foundation, like deciding how much to save, storing data, technological problems, legal issues, ethical questions, money struggles, and infrastructure issues. This shows how much dedication it takes to save the web's history.

10.4.2 Milestones in the development of web archiving

The evolution of web archiving has been a captivating journey with important milestones shaping the capture and preservation of our dynamic digital world:

Early Beginnings (1990s):

1996: The Internet Wayback Machine pioneers public web archiving.

1998: The US National Digital Library Program begins archiving government websites.

1999: The first International Web Archiving Workshop (IWAW) brings together archivists and technologists.

Building Infrastructure (2000s):

2001: The Archive-It project launches a scalable web archiving service.

2003: The LOCKSS (Lots of Copies Keep Stuff Safe) program promotes distributed web archiving for redundancy and preservation.

2007: The Warc file format is introduced for standardised web archives.

Addressing Complexity (2010s):

2012: The Memento protocol enables time-traveling access to archived web pages.

2013: The first International Networking Conference focuses on web archiving challenges in developing countries.

2015: The Pandora project explores archiving social media and the dynamic web.

Recent Advancements (2020s):

2020: The Web Archiving Testbed launches for testing and developing new web archiving tools.

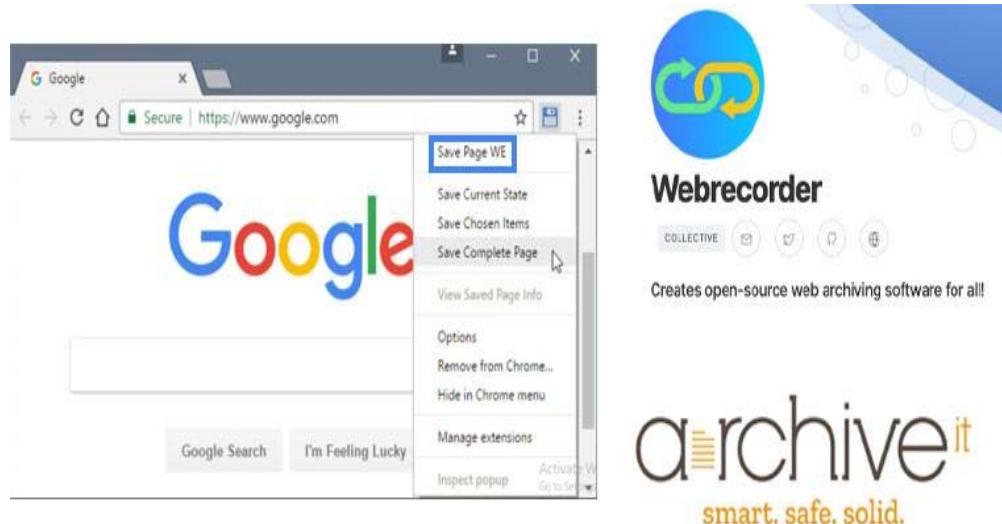
2022: The IIPC (International Internet Preservation Consortium) Web Archiving Working Group releases guidelines for ethical web archiving.

2023: The focus shifts towards preserving emerging web formats like interactive content and Web3 applications.

These milestones underscore web archiving's adaptability to technological innovations and the evolving Internet landscape.

10.5 BASICS OF WEB ARCHIVING TOOLS

Exploring the realm of web archiving tools, we find a range of user-friendly options catering to diverse needs, from casual researchers to seasoned professionals. Browser extensions like Webrecorder and Save Page WE simplify the archiving process, enabling one-click capture even for those with minimal technical knowledge. Additionally, platforms like Archive-It offer intuitive interfaces, allowing users to create custom collections and schedule regular captures effortlessly. These tools prioritise accessibility, encouraging individuals from various backgrounds to actively participate in preserving our digital heritage.



Web Archiving Tools

10.5.1 Introduction to User-friendly Tools for Archiving (e.g., Browser Extensions)

In our journey into the realm of web archiving tools, we are introduced to user-friendly options like Webrecorder and Save Page WE, serving a diverse audience from beginners to experts. These tools simplify the archiving process with one-click functionality and extend to advanced features such as selective archiving, customisable metadata, and collaborative capabilities. Recognising the paramount importance of digital preservation, institutions have adopted dedicated tools like Archive-It, Webrecorder, and Heritrix. This collective effort emphasises accessibility, welcoming individuals from various backgrounds to actively contribute to preserving our evolving digital landscape.

Webrecorder: Features: Webrecorder simplifies web page capture, allowing users to record and archive dynamic, interactive content. **Accessibility:** Designed for minimal technical expertise, its user-friendly interface and browser extension enables effortless archiving with a single click.

Save Page WE: User-Friendly Extension: Save Page WE is a browser extension for easy web page archiving suitable for casual users and researchers. **Single-Click Archiving:** Known for simplicity, the extension streamlines archiving with a one-click solution.

Archive-It: Custom Collections: Beyond individual page archiving, Archive-It enables custom collections for specific themes or topics. **Scheduled Captures:** Notably, users can schedule regular captures, maintaining updated archives over time.

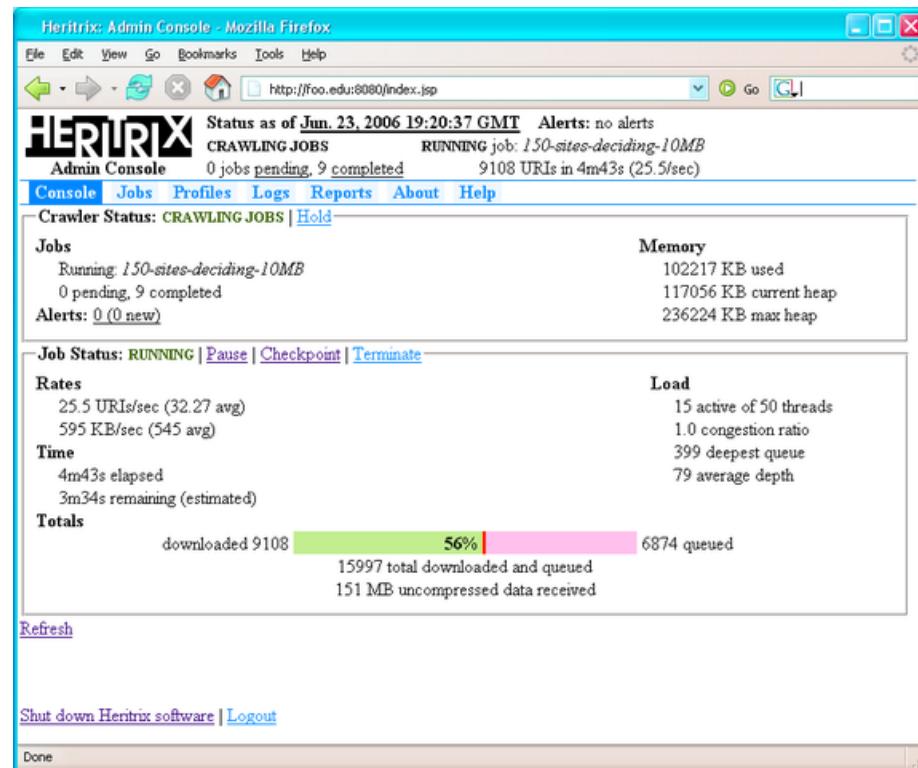
Community Engagement: Diverse Participation: User-friendly tools encourage varied engagement in preserving digital heritage, from casual users to professionals. **Promoting Accessibility:** As this emphasises accessibility, it allows users from diverse backgrounds to contribute.

Educational Initiatives: Tutorials and Resources: Tools offer tutorials, lowering entry barriers and empowering users with knowledge for web archiving contributions.

10.5.2 How Professionals can Integrate Archiving into their Workflow

Embarking on our exploration of web archiving, we find professionals spanning roles, such as librarians, archivists, media experts, and researchers, seamlessly integrating web archiving into their workflows. This practice, crucial for information preservation, sees widespread adoption of dedicated tools like Archive-It, Webrecorder, and Heritrix by institutions. These tools, equipped with advanced features like selective archiving, metadata customisation, and collaborative capabilities, empower professionals to curate comprehensive digital archives. As they navigate legal and ethical considerations, including copyright, terms of service, and privacy, professionals ensure compliance with regulations. Web archiving transcends

technicality, evolving into a strategic and ethical practice within the professional domain.



Heritrix

10.5.3 Web archiving features, strengths, and limitations.

The journey into web archiving tools reveals their diverse features, enriching their effectiveness in capturing and preserving web content. These tools capture entire websites and specific pages, navigate intricate sites, and replicate archived content within its original context. Utilising advanced crawling algorithms, they ensure comprehensive coverage of dynamic web content, making them important for studying website evolution, digital culture, and societal trends. However, challenges persist, particularly in accurately capturing interactive elements like forms and JavaScript-based interactivity. Additionally, difficulties may arise when capturing content from password-protected areas or websites with intricate structures.

Check Your Progress: 2

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

2) Compare your answers with those given at this Unit's end.

1. Analyse how the interface, content, and user interactions have transformed over time, revealing the platform's journey in the digital landscape.

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

2. Identify a web archive platform and explore its contents. Share your findings on how the archive preserves online content and its potential benefits for researchers.
-
-
-

10.6 RECORDS AND INFORMATION MANAGEMENT

Web archiving emerges as a central element in contemporary records and information management, tackling the distinctive challenges presented by online content's dynamic and constantly evolving nature. As a fundamental aspect of information governance, web archiving encompasses capturing, preserving, and managing web-based records. This all-inclusive approach guarantees the enduring accessibility, authenticity, and integrity of these records over time, highlighting its significance in pointing out the intricacies of modern digital information landscapes.

10.6.1 Websites as records

Incorporating effective website management into a holistic information and records management strategy aligns with practices for various organisational records. Similar to traditional files, websites play a crucial role in current and future activities, necessitating management for ongoing use, legal compliance, evidential support, and preservation of historical or cultural significance. Ensuring the accuracy of web content before publication is vital, given that even briefly live information may be archived or cached by search engines. Treating websites as records, involving the capture, management, and retrieval of information, significantly contributes to overall records management. The appearance and content of websites at specific times serve as valuable records, providing crucial evidence for legal or evidential purposes. Embracing such practices safeguards organisations and enhances the integrity of disseminated online information.

10.6.2 Selecting and collecting

Websites serve as visual and informational records, reflecting an organisation's intended public presentation and conveyed messages. Documents like board minutes, reports, policies, and plans embedded within websites act as concrete records of the activities they represent, signifying the time and resources invested by creators. Evaluating a website's value involves determining if it contains important business information or historical significance. Questions about content duplication or preservation in other locations, like shared network drives or electronic records management systems, aid informed decisions on retention periods. Adopting a format-neutral approach ensures uniform treatment of records, managing them based on purpose and utility rather than file format, encompassing spreadsheets, PDF documents, images, or websites consistently. Another thing pertinent to this is determining the archiving frequency. Depending on the rate of website

and content changes, along with the relative importance of the content, archiving may be needed once a year, twice a year, every three months, or on a customised schedule for accurately capturing evolving content. Non-technical considerations, especially regarding the scope and scale of collecting, impact the choice of a technical approach aligned with preservation needs. Regular archiving of valuable website content ensures capturing and preserving the evolving media landscape for future reference.

Check Your Progress: 3

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

2) Compare your answers with those given at this Unit's end.

1. Explore how web archiving can enhance online storytelling.

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

10.7 ETHICAL CONSIDERATIONS AND LEGAL FRAMEWORK

Safeguarding digital content is a major challenge for libraries worldwide. As digital materials can get lost or deleted easily, we must ensure they are preserved and stored appropriately. Libraries are well-equipped for this task, but creating a long-lasting plan poses significant challenges. Legal aspects, such as responsibility and abiding by the rules, are key concerns in this process. We must address these issues to ensure a robust plan for saving digital materials. It is important to understand and handle the complexities responsibly. Keep in mind that libraries play a crucial role in this, and we can contribute to preserving digital content for the future.

10.7.1 Legal and Ethical Issues in Web Archiving

Regarding web archiving, we must be mindful of legal and ethical considerations. We must respect copyrighting. So, understanding who owns the content and getting permissions are the keys. We must balance our preservation goals with the website's terms of service. Ethical practices, like anonymisation and consent, protect personal privacy must be followed. Being mindful of what we choose to archive minimises bias and ensures inclusivity. Accuracy requires careful verification and documentation. Balancing public access with safeguarding sensitive information may mean access restrictions. For long-term preservation, clear ownership and ongoing funding are essential. As the web is global, complying with international laws is a must. We must stay informed and collaborate with archivists, legal experts, and policymakers for a responsible approach.

10.7.2 Copyright, fair use, and permissions in the digital age

Copyright laws protect creators' rights, overseeing reproduction, distribution, and display. Fair use allows limited usage for purposes like criticism,

commentary, news reporting, teaching, scholarship, or research without explicit permission. However, clear-cut consent from copyright holders is necessary for various applications beyond fair use. It is crucial to walk through these principles to uphold intellectual property rights, nurture creativity, and responsibly share information in the digital domain. Always remember that these guidelines ensure a fair and respectful digital environment as we explore various opportunities in our online activities.

Copyright: Copyright is a legal framework that grants exclusive rights to creators of original works, encompassing literature, art, music, and digital content. These rights, applicable to online materials like blogs, images, videos, and software, provide copyright holders exclusive control over reproduction, distribution, display, and performance. The duration of these rights varies by jurisdiction, extending the creator's lifetime plus a specified number of years.

Fair Use: Fair use allows limited use of copyrighted material without seeking permission or paying the copyright holder. It balances copyright holders' rights with the public's interest, considering factors like the purpose of use, the nature of the work, the amount used, and its impact on the original work's market value.

Permissions: Permissions necessitate explicit consent from copyright holders for uses beyond fair use. This involves obtaining written permission, negotiating terms, or paying fees. Permissions are crucial for commercial use, creating derived works, or whenever you go beyond fair use allowances. In the digital age, obtaining permission can be done through online platforms or directly from copyright holders.

Digital Age Challenges: Copyright concerns intensify in the digital age due to the ease of copying and distributing content online. Issues like digital piracy, unauthorised sharing, and determining fair use in the digital domain have gained prominence. Digital rights management (DRM) technologies control access to digital content, and creative licenses provide creators with a framework to specify permissions for their work.

Technological Developments: Technological advancements, particularly in artificial intelligence and machine learning, have introduced new challenges to copyright. Questions arise regarding automated content creation, the incorporation of copyrighted material in algorithms, and the evolving impact on fair use considerations.

10.7.3 Balancing the public's right to know with privacy concerns

Balancing the public's right to know is crucial, especially in today's digital age, which has privacy concerns. Ensuring public access to information is fundamental for a transparent and informed society, maintaining democratic principles and accountability.

But we need to carefully deal with it. Privacy rights, protecting personal autonomy, dignity, and security, must be respected. This delicate balance often gets tested in media and journalism, especially when covering sensitive

topics. Journalists strive to disseminate accurate information while considering ethical and legal guidelines, sometimes involving redaction or anonymisation.

In our digital world, where personal information is easily accessible, maintaining this balance becomes quite the challenge. Concerns like private details disclosure, surveillance, and the risk of misinformation highlight the need for responsible journalism and robust privacy safeguards. Laws and regulations aim to guide us, but their interpretation sparks ongoing debates and legal challenges. So, buckle up as we navigate this intriguing terrain together!

Check Your Progress: 4

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

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

1. Highlight how web archiving plays a role in reporting for news media.

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

2. Emphasise the benefits of open-source reporting and distributive reporting in the context of web archiving.

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

10.8 WEB ARCHIVING FOR MEDIA PROFESSIONALS

Media archive management is all about organising and storing various media types, like videos and audio files, to make them easy to find and use later on. This involves adding descriptions or metadata, keeping files safe, and ensuring that media can be quickly retrieved. Our main goal is a system where media assets are well-organised, secure, and ready for future use. So, as we explore the world of media, remember the importance of maintaining an organised and accessible archive for a smooth and efficient workflow in the future.

10.8.1 Why is it important for media professionals?

Effective media archive management in the dynamic world of media and entertainment. Maintaining a well-organised archive is indispensable whether you're in production houses, news agencies, or journalism organisations. This

ensures easy access and repurposing of vast content volumes for new projects, re-releases, or licensing opportunities, reducing the Total Cost of Ownership (TCO). Robust media archive management fosters efficient collaboration among diverse project teams, ensuring a seamless workflow. News agencies and broadcasters heavily rely on this approach to curate extensive libraries, providing swift access to historical content for current stories and future reference. It aids journalists and editors in analysing historical trends and crafting compelling visual narratives. Remember, adherence to copyright regulations and licensing requirements is crucial for legal compliance in the ever-evolving landscape of professional media work.

10.8.2 Introduction to Preserving Digital Content for Journalistic and Media Purposes

In the dynamic field of digital preservation, the goal is to ensure that digital materials remain accessible for future media initiatives. Preserving digital content is crucial in the fast-paced world of journalism and media. It's about ensuring that today's stories, visuals, and narratives stay relevant for future generations. This preservation goes beyond safeguarding files and data; it is about preserving our collective media heritage and recognising the dynamic nature of the digital landscape. For journalistic and media purposes, digital preservation is a commitment to preserving the authenticity and accessibility of content that forms the basis of storytelling. It involves meticulous efforts to recreate the user experience, allowing future journalists and creators to access the rich tapestry of information and visuals that have shaped our worldview. In this digital era, where news unfolds in real-time and multimedia storytelling is key, the challenge is preserving bits and bytes and capturing the spirit and context of narratives woven into the digital fabric.

10.8.3 The need for Archiving in the Context of Digital Media

Archiving is crucial to an organisation's records management strategy, especially in digital media. It involves systematically storing digital documents securely in a long-term physical or digital repository, safeguarding them from damage or destruction to ensure their longevity and accessibility.

The primary goal of digital media archiving extends beyond preservation; it aims to store data over an extended period, aligning with the ever-evolving digital landscape. This strategic effort goes further by making records available to the public and future generations, fostering accessibility and historical continuity.

Much like their physical counterparts, archived digital media documents become easier to locate, protect, and maintain throughout their lifecycle. Neglecting this essential archiving process can obstruct a business's efficiency in the digital realm, underscoring the critical role digital archiving plays in the seamless functioning and sustainability of a media-centric organisation.

Digital media archiving is not just about safeguarding files; it's a commitment to preserving our collective media heritage, ensuring that today's stories and

visuals remain accessible and pertinent for future generations in the fast-paced world of journalism and media.

10.9 FUTURE OF WEB ARCHIVING IN MEDIA

Web archiving in media holds great promise as it adapts to the dynamic and evolving digital world. We look forward to trends beyond preserving static pages, including dynamic and interactive content with multimedia elements.

As our hunger for real-time information grows, web archiving in the media sector is expected to capture breaking news and live events more immediately. This advancement will demand innovative technologies for timely and comprehensive archiving, ensuring the preservation of rapidly changing digital narratives that define the future media landscape.

However, challenges arise amidst these opportunities. For a responsible and effective future in web archiving for media, considerations about the ethical use of archived content, evolving copyright issues, and robust preservation strategies must be fully appreciated. This process involves embracing possibilities and addressing ethical and legal considerations that come with the evolving digital landscape.

10.9.1 Emerging Technologies and Trends in Archiving

Emerging technologies significantly influence archiving, introducing new possibilities and challenges. Blockchain technology is gaining support and interest for its potential to enhance the integrity and authenticity of archived records through decentralised and secure ledgers. Artificial intelligence (AI) and machine learning transform archival processes by enabling automated sorting, categorising, and analysing vast amounts of data. Cloud-based archiving solutions provide measurable and accessible storage, allowing seamless collaboration and remote access. Additionally, advancements in digital preservation techniques, such as format migration and emulation, ensure long-term accessibility of archived content. However, these innovations also bring challenges, including ethical considerations in AI use, data security concerns in cloud storage, and the need for standardised practices in adopting emerging technologies for archiving.

10.9.2 Opportunities for Media Professionals in the Evolving Digital Landscape

Media professionals find abundant opportunities in the continually evolving digital landscape. The rise of online platforms and streaming services has created diverse avenues for creating, distributing, and engaging content. Social media platforms offer new outlets for storytelling, audience interaction, and brand building. Virtual and augmented reality technologies provide immersive storytelling possibilities, creating novel experiences for audiences. Data analytics and artificial intelligence enable media professionals to understand audience behaviour, personalise content, and optimise strategies for engagement. Podcasting and independent content creation have become more accessible, allowing media professionals to reach specific groups of audiences. Additionally, the global nature of the digital

landscape opens up opportunities for collaboration and content localisation on a broader scale. As the digital ecosystem evolves, media professionals can embrace innovation, expand their skill sets, and explore a dynamic landscape rich with creative and professional potential.

Check Your Progress: 5

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

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

1. Discuss potential challenges related to privacy and responsible use of archived material.

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

2. Share your thoughts on the role of web histories in understanding the digital landscape.

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

10.10 LET US SUM UP

In the intricate tapestry of Internet Histories, we compared the experience to enjoying a colossal online storybook, discovering how people and cultures intricately shape the Internet. This digital narrative led researchers into vast digital archives resembling online libraries, revealing how societal and cultural influences penetrate the Internet's evolution. Our interdisciplinary journey integrated insights from history, communication, media studies, and technology, with web archiving as a central tool. Beyond preserving content for researchers, it stood as a foundation stone for rich resources in history, sociology, and computer science, surpassing academia to impact legal cases, business continuity, and cultural preservation.

Delving into web archiving basics, we gained an understanding of diverse archiving methods, which ensure accessibility amid the Internet's constant instability. Exploring the historical context, we witnessed milestones securing the preservation of our dynamic digital world. User-friendly tools like Webrecorder and Save Page WE democratised archiving, emphasising integration into professional progress. This emphasised the strategic and ethical importance of web archiving, ensuring continuous accessibility and integrity of web-based records.

Looking closely into the future, we projected trends and technological advancements in web archiving for media, acknowledging challenges of ethics, copyright, and preservation strategies. In essence, web archiving emerged as a guardian of our collective media heritage, endowing media professionals responsible for maintaining an organised and accessible archive, a timeless repository for a flawless and efficient workflow in the ever-evolving media landscape.

10.11 KEYWORDS

Web archives are digital repositories that chronicle the evolving Internet by storing snapshots of web content, enabling users to revisit past versions of websites. They serve as essential historical records and resources for researchers, historians, and the public, employing diverse technologies to capture the dynamic nature of the online landscape.

Internet Histories: Internet Histories involve studying the evolution, development, and societal impact by delving into digital archives, using historical data to reveal how online activities and cultures have shaped their course, and combining perspectives from history, communication, media studies, and technological studies.

Online Information management involves systematically organising, storing, and disseminating information to support decision-making and operational efficiency. It also encompasses practices like data governance and knowledge management.

Legal Issues: Online legal issues are because of challenges and disputes in the digital realm, such as cyber security, digital privacy, intellectual property, and regulatory compliance within the online domain. Resolving these matters involves navigating the complex intersection of law and technology.

10.12 FURTHER READINGS

1. Brügger, N. (2016). Introduction: The web's first 25 years. *New Media & Society*.
2. Brügger, N. (2019). Understanding the archived web as a historical source. In *The SAGE Handbook of Web History*. SAGE Publications Ltd, <https://doi.org/10.4135/9781526470546>
3. Brügger, Niels and Schroeder, Ralph (Eds.). (2017) *The Web as History*. [Book]. UCL Press: London
4. Daniel Gomes, Elena Demidova. (2021). The past web. In Springer eBooks. <https://doi.org/10.1007/978-3-030-63291-5>
5. Katja Müller. (2023). Digital Archives and Collections: Creating online access to cultural heritage. *BERGHAHN BOOKS*.
6. Milligan, I. (2017). Welcome to the Web: The online community of GeoCities during the early years of the World Wide Web. In N. Brügger & R. Schroeder (Eds.), *The Web as history* (pp. 137–158). London: UCL Press.

7. Niels Brügger. *The Archived Web: Doing History in the Digital Age*. Cambridge, MA: The MIT Press, (2018).
8. Rajan, Mira T. Sundaran. (2012). ‘At the Frontiers of Lawmaking: Copyright and the Protection of Culture in India’, *The WIPO Journal* 4(1): 111–20.
9. Utpal K Banerjee. (2014). Towards an Online Audio-Visual Cultural Archive (01 ed., Vol. 01). [Google]. Indira Gandhi National Centre for the Arts.
10. Venlet, Jessica, Karen Stoll Farrell, Tammy Kim, Allison Jai O’Dell, and Jackie Dooley. (2018). *Descriptive Metadata for Web Archiving: Literature Review of User Needs*. Dublin, OH: OCLC Research.

10.13 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Web archiving plays a crucial role in preserving online content by capturing and storing web pages, websites, and digital materials for future access and reference. It ensures that valuable digital information remains accessible despite changes in technology, website structures, or the expiration of web content, safeguarding cultural heritage and historical records.

Check Your Progress: 2

1. Over time, digital platforms have undergone significant transformations in interface design, content delivery, and user interactions, reflecting the evolving landscape of technology and user preferences. Early websites featured basic interfaces with static content, limited interactivity, and text-heavy layouts. As technology advanced, platforms embraced multimedia elements, responsive design, and personalised user experiences, enhancing engagement and usability. Social media platforms revolutionised user interactions, facilitating real-time communication, user-generated content, and community building. These developments illustrate the platform's journey from static web pages to dynamic, interactive ecosystems that cater to diverse user needs and preferences.
2. Web archive platforms like the Internet Archive provide invaluable resources for preserving online content and documenting the evolution of the digital landscape. Through web crawling and archiving, these platforms capture snapshots of websites at different points in time, ensuring the preservation of digital artefacts for future generations. Researchers can explore archived websites to study historical trends, track changes in content and design, and analyse the impact of technological advancements on online communication. Additionally, web archives serve as repositories of cultural heritage, facilitating research in fields such as digital humanities, media studies, and social

sciences. Overall, web archives play a crucial role in preserving digital history and fostering academic inquiry into the ever-changing digital landscape.

Check Your Progress: 3

1. Web archiving enriches online storytelling by preserving historical web content, ensuring its accessibility for future generations. By capturing websites, social media, and multimedia elements, web archives provide a comprehensive view of digital narratives, enabling researchers, journalists, and storytellers to analyse trends, trace evolutions, and contextualise online discourse over time.

Check Your Progress: 4

1. Web archiving is crucial for news media reporting as it preserves online content, ensuring access to historical data for journalists. It safeguards against content alteration or removal, maintaining the integrity of news stories and allowing for accurate referencing and fact-checking, thus enhancing the credibility of news reporting.
2. Open-source reporting and distributive reporting are empowered by web archiving. Open-source reporting leverages publicly available information archived online, fostering transparency and collaboration among journalists. Distributive reporting involves decentralised investigation, where journalists worldwide contribute findings, facilitated by archived web data. These approaches democratise journalism, enriching reporting with diverse perspectives and sources.

Check Your Progress: 5

1. Challenges related to privacy and responsible use of archived material include ensuring compliance with data protection regulations, safeguarding sensitive information, and respecting individuals' rights to privacy. Additionally, ethical considerations regarding consent, anonymization, and data retention must be addressed to prevent misuse or unauthorised access to archived data.
2. Web histories play a crucial role in understanding the digital landscape by providing insights into users' online behaviours, preferences, and interactions. Analysing web histories can reveal trends, patterns, and shifts in digital consumption, facilitating research on information dissemination, user engagement, and online communities. However, ethical considerations regarding data privacy and consent are paramount when accessing and analysing web histories to ensure responsible use and mitigate potential risks of data exploitation.

UNIT 11 TEXTUAL ANALYSIS

Structure

- 11.0 Introduction
 - 11.1 Learning Outcomes
 - 11.2 Perspectives of Textual Analysis
 - 11.3 Types of text and methods of collection
 - 11.4 Presenting textual analysis.
 - 11.5 Further Readings
 - 11.6 Check Your Progress: Possible Answer
-

11.0 INTRODUCTION

In this Unit, you will learn about textual analysis as a research method. Before understanding what textual analysis is, let us understand what text means in textual analysis. Text in the textual analysis is something we interpret to draw some meanings out of it. This can be a book, an article, a television programme, a film, a magazine, ornaments, furniture, digital devices, and so on. Text can also be seen as social events in which people act or interact with each other. The interaction can be 'linguistic or non-linguistic' (Fairclough, 2003, p. 21). These forms of text are analysed in a way that provides an impression of how people make sense of things or events. If these sources are considered for interpretation and given some meanings, they can be called text in textual analysis. This means that text is something that is analysed for the production of meanings (McKee, 2003). Considering this, textual analysis can be defined as a method in which a researcher collects information concerning how people make sense of the world, interprets, or analyses it and attaches some meanings to it based on the research question. It is a method that helps understand how members of various cultural backgrounds fit into the setting in which they are a part. This method is employed by researchers in various disciplines, such as cultural studies, media studies, communication studies, sociology, philosophy, and so on (McKee, 2003).

The researcher tries to attach meanings to some selected text in textual analysis. The meanings can be given to the text in several ways, each being equally valid since it depends on the context in which the meaning is given to it. This means that the researcher considers everyone's interpretation as valid but does not rely on their meanings; instead, the researcher attaches his own meanings to this text based on the context of his analysis (VanderStoep & Johnston, 2009).

Researchers use textual analysis in communication studies to explain and analyse audio-visual text. During the process, the researcher elucidates the text and its structure. However, prior to this, important matters for the researcher are selecting the appropriate text, acquiring it, and deciding which method to use in its analysis (Frey et al., 1999).

11.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Understand the basic nuances of textual analysis methods;
- Describe steps involved in textual analysis; and
- Devise your own research question and methods under textual analysis method.

11.2 PERSPECTIVES OF TEXTUAL ANALYSIS

Central to textual analysis is meaning, which can be interpreted from various perspectives. There are three broad perspectives of textual analysis: the *rhetorical perspective*, the *critical studies perspective*, and the *discourse analysis perspective*. Each of these perspectives on meaning will yield a different interpretation of the text (VanderStoep & Johnston, 2009).

a) ***The rhetoric perspective:***

This perspective involves employing language in exploring the interactions between the actual text, the producer of the text (an author), and the readers of the text (audience). Through this perspective, the researcher explains, analyses, interprets, and evaluates what is persuasive within the text(s). The rhetorical perspective assumes that texts always carry a kind of meaning with them that influences people. Vander Stoep and Johnston (2009) claim that those engaging with textual analysis believe texts influence people's beliefs, attitudes, and behaviours. However, the extent of such an influence is always unknown. They further believe that culture is experienced through text, and through the analysis of cultural texts, we understand our traditions, values and even our identity (Vander et al., 2009). Rhetorical perspective serves important functions in textual analysis. It focuses on the persuasive ideas within the text and aids in understanding historical, social, and cultural contexts. Through this perspective, society can be evaluated by social criticism. One important function of rhetorical perspective is the pedagogical function, which teaches people what persuasion is and how it works (Andrews, 1983).

b) ***Critical studies perspective:***

The critical studies perspective is common among researchers from various disciplines, such as media and communication studies, cultural studies, sociology, and philosophy. In this perspective, the text is seen as a place of power struggle. All popular culture media content, such as films, news, music, etc., carries messages which reveal who has power and who does not. Researchers explore texts through the critical studies perspective to see open and hidden domination and bias. There is always an idea underlying the text that creates and strengthens some specific group's power. Through this perspective, a researcher examines how ideas and dominance are expressed in language. The critical studies perspective involves investigating language in texts to show how

particular phenomena in languages (such as selection of words, structure of sentences, metaphors, strategies, etc.) are used to represent a particular viewpoint. Through this perspective, the researcher binds analysis of language with different contexts, such as social, historical, and political contexts (Baker et al., 2013).

c) ***Discourse analytic perspective:***

The discourse analytic perspective entails understanding a phenomenon by interpreting discussions and narratives. What lies at the base of this perspective is that meanings are given to a discussion through the analysis of communicative action. Vander Stoep and Johnston (2009) have mentioned two specific kinds of analysis: *conversational* and *narrative*. *Conversational analysis*, as an approach, is used to study human interaction in society. It entails the interpretation of discourse taking place in a natural environment. Through this approach, verbal, written or physical communications (sometimes the behaviour of those engaging in discourses) are analysed to examine the structure of conversations as well as the functions that a particular utterance serves (Vander et al., 2009; Gee & Handford, 2012).

On the other hand, narrative analysis involves analysing naturally occurring storytelling. In this type of analysis, the researcher starts using a "set of principles and seeks to exhaust the meaning of the text using specified rules and principles but maintains a qualitative textual approach" (Lune & Berg, 2017, p. 186). Through narrative analysis, the researcher explores the story's structure and how stories help in community building, forming, and maintaining relationships, establishing identities and values and so on (VanderStoep and Johnston, 2009).

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. Describe the benefits and drawbacks of textual analysis?

.....

2. Enumerate the applications of textual analysis in digital media studies?

.....

11.3 TYPES OF TEXT AND METHODS OF COLLECTION

In textual analysis, two text types are *primary* and *secondary*. The *primary text* comprises original information collected for analysis by the researcher. It includes the text collected for the first time by the researcher, considering the research objectives. Several methods are employed to collect primary data, but in textual analysis, two common methods of primary text collection are interviewing and observation. On the other hand, the *secondary text* includes the text used for the descriptive study, which helps to understand the primary text or clarify its analysis. Secondary text includes the reference works selected from existing literature, such as articles, lectures, published research works, etc. These sources are commonly termed documents (Bauer et al., 2014; Sahu, 2013).

a) ***Interviewing:***

Interviewing is one of the most common methods of data collection in social sciences, media and communication studies, and behavioural sciences. It involves listening skills on the interviewer's part, wherein the interviewer records the interviewee's responses. These recorded responses are later transcribed into a text stream (Bauer et al., 2014). This means that in the interview method, communication takes place orally. In the interview method, the interviewer asks different questions to the respondents and records their responses. Interviews are of three types – personal interview, telephonic interview and chatting (Sahu, 2013). *Personal interviews* are face-to-face interviews conducted at places convenient to the respondents, such as the respondents' houses or office locations. They are costly, but they enable the respondents to obtain clarifications when the questions are unclear and allow the interviewer to follow up when the responses are incomplete (Shaughnessy et al., 2012). In *telephonic interviews*, respondents are contacted over the telephone, questions are asked, and responses are recorded. Through telephonic interviews, a researcher can have access to respondents who are spread over a large geographical area at a low cost, unlike personal interviews. In this type of interview, the respondents enjoy greater anonymity, encouraging them to be fair and honest in their responses without any influence from the interviewers' physical presence. The third type is *chatting*, which has become the most popular way of interviewing people and collecting their responses. Chatting is the exchange of words or communication between two or more persons through the Internet. People can communicate their ideas, thoughts, and beliefs to the interviewer through chat, even in the remotest places. Chatting can be of three types – written, audio and video chat. Through written chats, responses are collected in the form of text. In audio chats, responses are collected through audio recording devices, which are later transcribed into texts. With the advancement of technology, researchers are relying on various phone applications and video recording devices during interviews to collect data (Crano& Brewer, 2002).

b) Observation

Observation collects primary information, such as people, behaviour, or other events in each setting. Researchers observe how people interact in a given setting, communicate, and deal with events through this method. These observations are recorded in a variety of forms, such as photographs, audio, and videos, which are later converted into text by the researcher. However, observation does not involve only vision; all our senses are to be used while collecting text through observation (Fox, 1998). As such, the observation method can be used in two ways: *participant* and *unobtrusive*. *Participant observation*, having its roots in anthropology and sociology, is a method of data collection in which a researcher becomes part of the group or community under study. The researcher interacts with the people in their social environment and notes, recording voices and other sounds, clicking pictures, and asking questions to the group members, which helps unfold the meanings behind various texts (Driscoll, 2011).

Three important elements of participant observation are gaining access to the setting, building rapport with the participants, and spending longer time within the group. The first is gaining access to the setting, which entails researchers' admission into the study setting. This setting could be a group, a community, homes, places of work, play fields, commercial entities, sacred places, media institutions, etc. The second element is building rapport. Rapport building means establishing a warm relationship with the participants. The relationships must be built on trust and acceptance for fair responses from the participants. The third is spending longer periods within the group. This is the nature of participant observation, which requires the observer to spend around a year or more within the group to have larger conversations with the participants and better understand the group, people, events, and situations concerning the group. However, the amount of time spent in the group depends on the nature of the research. Thus, the time may vary from a few days to weeks to months and even years. This also involves multiple visits by the observer. The other is *unobtrusive observation*. It is a kind of covert observation in which the observer does not necessarily interact with the participants but only records their actions within the setting, such as their interaction and conversation with other group members, their behaviour, and so on. Most often, participants volunteer to be part of the research during observation, but in unobtrusive observation, participants are unaware that they are being observed (Troike, 2003).

c) Documents

Using primary sources by researchers in textual analysis poses certain challenges in critically analysing it, such as critical thinking, interpreting the text, finding a link between various texts, etc. Using documents included in the secondary data is key to enhancing our understanding or knowledge about the issue. These documents are made available by various sources such as academia, government establishments, private

organisations, business companies, media sectors, researchers, or other voluntary organisations. The documents made available by these sources include research papers/articles, booklets, books, journals, magazines, yearly reports, census reports, other technical reports, and so on (Sahu, 2013). However, researchers must ascertain the originality of these documents before utilising them. Now, documents can be available in visual and alphabetical text. Visual texts include graphics, maps, images, videos, etc. Alphabetical texts include anything in the secondary data and are available in written form. Thus, there are diverse documents produced by other researchers in their contexts but used in textual analysis in an entirely different context. Historians introduced documents as evidence of past events, but their use is widespread, especially in social sciences, media, and communication studies (Bauer et al., 2014).

11.4 PRESENTING TEXTUAL ANALYSIS

One of the best ways of presenting the findings of textual analysis is to consider the audience, whom you will present it, and the objectives of the analysis. As such, the analysis of texts is presented in three stages –the introduction, the body, and the conclusion. The *introduction* must be written to grab the readers' attention and provide necessary information about the objectives of textual analysis, the methods of text collection, and the ways of analysing and interpreting the text and the audience. It must also provide information about the kind of text that was analysed. The second stage is *the body*. This involves the interpretation and analysis of the text. All the text collected for analysis is scrutinised so that the researcher tries to locate the emerging themes within the texts, which are later utilised in different sections throughout the writing process. Different text forms, as discussed earlier, are utilised in this stage. Each chapter describes the link between the text and the researchers' analysis. At the start of each chapter, the main theme of the chapter is discussed, followed by the original text and then the researchers' interpretation of the specific quotes from the text, which connects the analysed text with the original quotes. Visuals are also used during the analysis. The third stage is the *conclusion*. In this section, all the findings of the analysis are collected and presented persuasively. The researcher must take the utmost care in bringing new information to the conclusion stage. However, the researcher must ensure that all the information presented earlier will provide readers with sufficient information to satisfy their needs (Hoey, 2001).

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. Briefly explain the different kinds of textual analysis

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

2. Describe the steps involved in presenting findings of textual analysis.
-
.....
.....

Bibliography:

1. Andrews, J. R. (1983). *The practice of rhetorical criticism*, Collier Macmillan Baker, P., Gabrielatos, C. & McEnery, T. (2013) *Discourse analysis and media*
2. *attitudes: The representation of Islam in the British press*, Cambridge University Press
3. Bauer, M.W., Suerdem, A.K.&Bicquelet, A. (2014). *Text analysis: An introductory manifesto*. In, M. W. Bauer, A. Bicquelet and A. K. Suerdem(eds.) *Textual analysis: Sage benchmarks in social research methods*, (pp. xxi-xlvii) London, UK: Sage
4. Crano, W. D. & Brewer, M. B. (2002). *Principles and methods of social research*
5. (2ndEdition), Lawrence Erlbaum Associates
6. Driscoll, D. L. (2011). *Introduction to Primary Research: Observations, Surveys, and Interviews*. In Charles Lowe and Pavel Zemliansky (Eds.) *Writing Spaces: Readings on Writing*, Volume 2(pp. 160–161) Parlor Press
7. Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*, London, NY: Routledge
8. Fox, N. (1998). *How to use observations in a research project*, Sheffield: Trent Focus Group
9. Frey, L., Botan, C., & Kreps, G. (1999). *Investigating communication: An introduction to research methods*, (2nd ed.), Boston: Allyn& Bacon
10. Gee, J. P.&Handford, M. (2012). *The Routledge Handbook of Discourse Analysis*, London, NY: Routledge
11. Hoey, M. (2001). *Textual Interaction: An introduction to written discourse analysis*, London, NY: Routledge
12. Lune, H. & Berg, B. L. (2017). *Qualitative Research Methods for the Social Sciences*, Pearson
13. McKee, A. (2003). *Textual analysis: A beginner's guide*. London: Sage.
14. Sahu, P. K. (2013). *Research Methodology: A guide for researchers in agricultural science, social science and other related fields*, India: Springer
15. Shaughnessy, J. J., Zechmeister, E. B. &Zechmeister, J. S. (2012). *Research Methods in Psychology* (9th Ed.), McGraw-Hill
16. Troike, M. S. (2003). *The Ethnography of Communication: An Introduction* (3rd Edition). Blackwell Publishing
17. VanderStoep, S. W.& Johnston, D. D. (2009). *Research methods for everyday life: Blending qualitative and quantitative approaches*. Wiley: Jossey-Bass

11.5 LET US SUM UP

In this Unit, you learnt about textual analysis as a research method and how the method is employed for the analysis of the text. What texts are and what the different forms of text used in textual analysis are elucidated in this Unit. Every researcher adopts a perspective while engaging in textual analysis. Three different perspectives of analysing text are discussed in detail. On a broader note, two text sources are used in textual analysis, and there are different text forms under these two sources. This Unit mentions and details the methods used to collect text from these two broader sources. Finally, the procedure to present the textual analysis is bifurcated in three parts and elucidated in the last section. This Unit will guide the students who engage in textual analysis in various disciplines, especially social sciences and media and communication studies.

11.6 FURTHER READINGS

1. Gourlay, L., Hamilton, M., & Lea, M. R. (2014). Textual practices in the new media digital landscape: Messing with digital literacies. *Research in Learning Technology*, 21(4). <https://oro.open.ac.uk/39372/2/Messing%20with%20digital%20literacies.pdf>
2. Günther, E., & Quandt, T. (2018). Word counts and topic models: Automated text analysis methods for digital journalism research. In *Rethinking research methods in an age of digital journalism* (pp. 75-88). Routledge.
3. Ignatow, G. (2016). Theoretical foundations for digital text analysis. *Journal for the Theory of Social Behaviour*, 46(1), 104-120. <https://www.dhi.ac.uk/san/waysofbeing/data/data-crone-ignatow-2016.pdf>
4. Patton, D. U., Blandfort, P., Frey, W. R., Schifanella, R., McGregor, K., & Chang, S. F. U. (2020). Vatas: An open-source web platform for visual and textual analysis of social media. *Journal of the Society for Social Work and Research*, 11(1), 133-155.
5. Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Exploring the boundaries of open innovation: Evidence from social media mining. *Technovation*, 119, 102447. <https://www.sciencedirect.com/science/article/pii/S0166497221002285>

11.7 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Textual analysis offers deep insights into meaning, context, and cultural significance within texts, aiding in understanding societal trends and ideologies. It allows researchers to uncover implicit messages, symbolism, and discourse structures, facilitating critical interpretation. However, it may be subjective, time-consuming, and reliant on researcher interpretation, potentially leading to bias and limited

generalizability.

2. Textual analysis finds diverse applications in digital media studies, including analysing online content for sentiment analysis, understanding audience engagement through comment sections, identifying patterns in social media discourse, examining linguistic features in online communication, and evaluating the impact of digital storytelling techniques on audience perceptions and behaviours.

Check Your Progress: 2

1. Textual analysis encompasses various methods for interpreting and understanding written or spoken texts. Content analysis involves quantifying textual data to identify patterns or themes. Discourse analysis examines language use within social contexts to uncover underlying ideologies. Semiotic analysis explores symbols and signs to reveal deeper meanings encoded within texts.
2. First, organise the findings by themes or patterns identified during analysis. Then, provide textual evidence to support each finding, including quotes or examples from the text. Next, contextualise the findings within relevant theoretical frameworks or literature. Finally, draw conclusions and implications based on the analysed data, highlighting significant insights.

UNIT 12 ONLINE ETHNOGRAPHY

Structure

- 12.0 Introduction
 - 12.1 Learning Outcomes
 - 12.2 Introduction to Ethnography
 - 12.2.1 Traversing Digital Landscape
 - 12.2.2 Need for Online Ethnography
 - 12.3 Methods of Online Ethnography
 - 12.3.1 Qualitative Methods
 - 12.3.2 Quantitative Methods
 - 12.4 Advantages and Disadvantages
 - 12.5 Ethics and Responsibilities
 - 12.6 Let Us Sum Up
 - 12.7 Keywords
 - 12.8 Further Readings
 - 12.9 Check Your Progress: Possible Answers
-

12.0 INTRODUCTION

In this Unit, you will be introduced to one of the most engaging research methods: Ethnography, the study of human communities. In the era of computer-mediated communication, we will discuss how online communities are adapting from and slowly replacing offline communities. We will then explore the need to study these communities and look at how ethnography has been adapted for online spaces.

Going further, we will also discuss various methods of conducting online ethnography, ranging from qualitative to quantitative, and which cases warrant which method of data collection. Lastly, we will discuss the advantages and disadvantages of online ethnography as a research method and consider the code of ethics applied to digital ethnographers.

12.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Describe Ethnography as a research practice;
- Adapt Ethnography for the digital space;
- Identify the research methodologies associated with Online Ethnography and
- Appreciate the ethics and responsibilities of an online ethnographer.

12.2 INTRODUCTION TO ETHNOGRAPHY

You may have read about travellers in ancient times who visited faraway lands on foot and recorded their accounts of the different places, cultures, customs, and traditions in their written memoirs. Chinese traveller Fa-Hien visited India in 405 AD, visiting the Buddhist communities in various cities and even learning Sanskrit to understand the culture. On the other hand, Moroccan traveller Ibn Battuta visited India in the early 14th century, and his account is full of tales of bandits in the hinterlands. Their written accounts were the first to document the vast Indian culture, the courts of erstwhile kings and the traditions of the society through an alien eye. While Ethnography had not developed as a research methodology back then, the books written by these travellers chronicled the culture of a community from an outsider's perspective, which is basically what Ethnography stands for today.

To put it precisely, Ethnography is the study of human communities and their indigenous cultures. It involves observation of individual behaviour and the collective reaction of the community, which forms longstanding cultural practices. Each community has its own set of norms and beliefs and moral code of practice forming the law of the land. The community elders usually set the code, and the younger population abides by it. Behaviour not accepted by the community usually leads to punishment or, in extreme cases, banishment of sorts.

In this case, the researcher or the ethnographer joins a particular community to study. He/she then employs various methods to gain information about it. Think of scientists spending years in a forest, observing a native tribe, and gathering data to know more about them. The data may include a brief history of the community, starting from its origins to evolution, to estimating the impact of the geographical and environmental conditions on the rituals and practices of the community and the emerging indigenous beliefs and culture that define that community.

12.2.1 Traversing Digital Landscape

Technology is spreading its roots to many aspects of society. Communication, which forms the core of human experience, is also being mediated through technology. Be it a text message you receive or an email you send, it is a digital form of communication. Platforms on the Internet are not just being used to communicate with individuals but also within groups. Social media platforms facilitate various groups' assembling and communicating online.

Digital Ethnography is a research method involving the exploration and study of communities assembled online with digital tools. These communities develop their social practices, contributing to a consolidated culture, much like the offline real-world communities. The people coming together to form these communities are not brought together due to their geographical proximity but rather due to common interests. The interests could range from sports to celebrities to science and education, and content based on these

topics is posted and shared in the group. With time, these groups formulate their dos and don'ts, applying to all participating individuals. These groups may include chat groups, group pages on social media, gaming communities, and other websites that allow for cross-user interaction.

The researcher or the online ethnographer joins a virtual community, most hosted on social media. They then observe the practices and interactions between the participants within the community. He or she may also gather field notes and record them using a digital tool, in this case, like a word processor. On the other hand, quantitative data can be easily tabulated in Excel sheets. Instead of photographs, the researcher may save the weblink to a page, aided with a screenshot or screen recording. The web link is useful for accessing the information at the data analysis stage. However, noting the access date could also prove beneficial as the information is dynamic on the web and can be taken down. As a last resort, screenshots and screen recordings also prove beneficial later in the analysis.

Collecting data in digital format helps in many ways. It is easy to archive without losing data quality and can be accessed and edited remotely. Several digital tools are also available for qualitative and quantitative analysis, such as social media analytics, data visualisation software, and qualitative analysis software. The data also becomes easier to handle and share with other collaborating researchers. It can also be digitally encrypted to ensure its safety.

The researcher may also interact with the content posted in the community, posting comments to initiate a conversation with the users. Individual participants may be contacted based on their level and frequency of contribution to the community, and one-on-one interviews may be conducted to gather richer qualitative data. The interviews can be carried off the platform, to a video conferencing one, or even an offline space, in case the interviewee consents.

12.2.2 Need for Online Ethnography

The amount of time humans spend online has been increasing gradually over the years due to factors related to work or leisure. Thus, as a wider discipline, anthropology, the study of human beings, must also look at the digital vistas to expand itself. With the advent of the Internet and social media platforms, diverse online communities are mushrooming. These communities have been observed to nurture their unique norms and practices, much like their offline counterparts.

Platformed on the World Wide Web, these communities attract a worldwide audience. They allow members from diverse geographical, economic, and social backgrounds to join and assimilate into the community, making them much more heterogeneous than offline communities. These platforms also offer varying levels of participation without the societal pressure to contribute. A person's contribution to an online community would be due to their interest, as it is a highly individualistic and personal activity. The users also have the option of anonymity. Thus, these communities also see uncensored participation compared to offline participation.

Social media platforms are also becoming highly instrumental in shaping public opinions and forming cultural trends. Online Ethnography, thus, helps researchers understand online behaviour and practices. Studying human behaviour online would help not only gauge the influence of technology but also search for new avenues of market research. These communities also see seeds of social movements planted, and thus, Online ethnography could explore the issues and concerns of virtual populations as well.

Digital ethnographers have also noted that social performances in online groups may differ from those in offline groups. This underlines the phenomenon that an individual's online persona differs from the offline one. Online ethnography can also help in understanding why this change occurs. An analysis of this sort is especially needed as computer-mediated communication is rapidly replacing the face-to-face format. Emergent online spaces such as manosphere also hint at a subaltern dormant in the offline spaces of society. It is still a matter of consideration whether online personalities should be studied as separate personas or as a continuation or result of offline personalities.

All these factors denote a rising and pressing need for adapting the traditional ethnographic methods to suit the digital platforms and inhabiting virtual populations, resulting in the practice of Online Ethnography.

Activity: 1

Join an online community based on any hobby or interests you may have. Figure out the avenues where you can gather information. You may have access to content posted in the public space. Frequency of posting, type of posts, frequency of reaction, and quality of reaction are a few things that you can observe. Identify the ways you can observe the participant interaction. List them down.

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 online communities different from offline communities?

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

2. What are the pros of collecting data in digital format?

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

12.3 METHODS OF ONLINE ETHNOGRAPHY

You may remember your time in a classroom. You may recall observing your classmates and finding a few notorious backbenchers, disciplined frontbenchers, teacher's pets, and uninterested genius kids. How did you figure out their characteristics? You might have observed everything from their behaviour to their uniform to their interactions with classmates and teachers. Apart from the language and words, nonverbal aspects such as clothing, body language, and regularity in attending school would also be contributing factors while classifying or coding the students. What would be your 'thick description' of your class?

Now imagine meeting all the classmates in a virtual group, all joining with their social media profiles. How would you identify them now? What will you be able to observe? What data points will you gather? Can you extract more or less information compared to meeting them offline? How would the 'thick description' vary in this case?

Let us look at the ways data can be collected in the case of Online Ethnography. Firstly, the level of access available to the researcher may vary depending on the community. Some communities on the deep web may allow transparent access to all members, given their anonymity. In contrast, some other communities on accessible social media platforms may need an existing member's referral to gain access. Some researchers choose to be lurkers - or mere observance in the case of online groups. Lurkers do not interact with other participants of the group and rather just observe the interaction and other social practices. On the other hand, participant observation needs them to interact with the content posted in the group and with other participants. They have a prolonged engagement and become deeply immersed in the native community, much like their offline counterparts.

Secondly, the interactions between the participants in online communities are largely based on textual exchange. As an online ethnographer, I find these texts provide most of the data to collect and analyse. Thus, recording and analysing the posts and comments becomes an integral part of digital ethnography. Thirdly, recording the communication between the participants and the researcher in the offline world may not be possible, although it is sometimes advisable. This makes referring back to the data impossible, as one needs to rely on memory, which can be prone to biases. In the case of an online interaction, it is mostly recorded and archived automatically, ready for referral as and when needed.

Thus, online ethnography incorporates a range of methodologies to overcome the challenges posed by virtual populations. The choice of method depends on two major factors: the aim of the study and the availability of data.

12.3.1 Qualitative Methods

Qualitative methods elicit qualitative data. Simply put, this data is not represented in numbers but reflects on the qualities of the researched attributes. It is helpful when the researcher aims to understand human experiences, perspectives, and behaviour. A few methods associated with Online qualitative research are:

Participant Observation Method: In this method, the researcher joins in and immerses themselves in the practices of the online community. They observe the emerging patterns and record them as 'Field notes'. At times, the observations are categorised according to codes pre-decided by the researcher, but most of the time, the researcher needs to be active and agile to gather unprecedented data as well.

Observations and notes help consolidate the findings in a 'thick description' of the community. The observations could be made on types of social media profiles, types of posts, types of interaction incited by the posts, and the traffic generated on the page. Participant observation helps find discrepancies between what is said and what is seen as a practice in the community.

Online Interviews: Online interviews are much like their offline counterparts; they are one-to-one conversations only mediated through a computer. They could be conducted through Instant Messaging or Chat, email, and Audio/Video calling over the web. These interviews help gather data on a participant's views and habits in an online space. They are also helpful in gathering any form of data that cannot be observed and needs to be elicited from individuals. They also shed light on the form of communication practised in the group.

There are two basic types of online interviews: Synchronous interviews are those where questions and responses are conducted simultaneously. Both interviewer and interviewee are present online at the same time. In this case, the interviewer can base their questions on the responses garnered. There is more spontaneity in the process, thus eliciting more candid data. On the other hand, asynchronous interviews are those where interviewer and interviewee can participate as and when one is free. While providing freedom of time, these interviews give the interviewee time to respond to the questions after giving them due thought. These interviews elicit more formal data, as deliberation and thought have been put into the conversation.

Online Content Analysis: This method is also called Online Textual analysis; like its offline counterpart, it includes gathering and analysing the online content through extensive coding. As discussed earlier, most communication is text-based in the online sphere. Thus, the researcher can analyse the posts, comments and other forms of text posted by the community users.

Sampling can be tricky for online content, as the universe is vast and ever-expanding. In this case, one must filter the data intensively to extrapolate and analyse. Also, the data is dynamic in this case. The person posting may edit it later or remove it altogether. Thus, the key analysis factors must also be in place much before the researcher starts to work, offering less spontaneity and greater planning than offline ethnography.

Online Focus Group Discussions: Like the offline counterpart, a Focus Group Discussion involves a researcher interacting with a group of pre-screened respondents qualified to represent the research sample. In the online format, the researcher needs to establish a base with all the respondents and set an online meeting time that suits all of them. Choosing a time to suit all

can be problematic, given that the online communities have members living in different time zones. Prepared with points to discuss, the researcher may conduct the discussion in a group chat, using text, or in audio-video format using video conferencing platforms.

Offline Participant Observation: After gathering the available data online, the researcher may move the data collection to the physical world to consolidate the research. The researcher may try to conduct a face-to-face interview with the samples and observe their body language, dressing habits, environment, etc., as well, a feat impossible in the virtual world. Like offline ethnography, the more comfortable the respondent feels, the more information can be gathered and the richer the data will be.

12.3.2 Quantitative Methods

Web Usage Mining: Web Usage Mining or Data Mining is the process of extracting and analysing data from the web to study emerging patterns. Various types of tools are available today to make data extraction an automated process. According to some researchers, data mining is a misnomer because the activity aims not just to extract the data but to analyse it to identify emerging patterns that meet the research objectives. The process usually involves five major steps:

1. **Selection of data**—The researcher identifies the data sets needed for the study. These could include comments on a post, details of user profiles, and even the movement of web traffic from one website to another.
2. **Data pre-processing:** The data is then brought into a format suitable for analysis according to the available tools.
3. **Coding of data** - The data is now classified and segmented into various categories for analysis.
4. **Pattern assessment** - Patterns are studied in the classified and coded data. These patterns are analysed again with the help of useful statistical tools.
5. **Findings** - The findings are now summarised and applied to the broader universe of the sample.

Social Network Analysis: Social Networks are intricate webs of people connected, sometimes through an acquaintance, sometimes through a colleague or even a distant cousin. Social Network Analysis is the analysis of structures of relationships within an online community. Simply put, it studies how humans are connected in the web world. This analysis helps identify common nodes and points to the opinion leaders of the group. This also helps in etching out the patterns of acceptance and popularity and, thus, the connection emerging within the group. Studying these patterns is instrumental in discovering how information is circulated within a group. The findings are usually represented through a graphical representation known as sociograms, with people represented as nodes and lines connecting them.

Online Content Analysis: Quantitative Content Analysis on the web works by extracting the content, coding the data sets, and then applying statistical

tools to them for analysis. Several tools can automate the process, as the data sets are huge in this case. Imagine if you had to manually code every comment an article on a popular news website would generate!

The data could be text-based, visual/graphic-based, or even audio-visual. Thus, the tools required to analyse the data would also differ. A combination of all these tools and the qualitative ones is known as Multimodal analysis.

Activity: 2

Let us return to the community you joined in the previous section. Now explore the type of content posted on the group, the reactions it accumulates, and the frequency. Record your field notes in digital format for a week. Check which posts gather more traffic compared to the ones that do not. At the end of the week, check what conclusions you can draw for the community.

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 two types of online interviews?

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

2. What are the processes involved with data mining?

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

12.4 ADVANTAGES AND DISADVANTAGES

As with any other research method, digital ethnography also has to offer its advantages and disadvantages to the researchers who opt for it. Let us discuss a few of the advantages first:

1. **Global reach:** The web can now be accessed from any corner of the world. While the cost of connecting to the Internet may vary, it has become a staple across social classes. Thus, the virtual population is not limited to any geographical area but is spread across physical, economic, and social boundaries.

2. **Flexibility in data:** While on the one hand, the researchers can access archives on the web to access data from the past, on the other hand, they can also witness real-time conversations and gather real-time data. In this case, one is also able to compare the data from the past and present.
3. **Versatility in data:** Data can be gathered in any format, from text to visual to audio files, each warranting a different analysis tool and method. This also allows the researcher to access qualitative as well as quantitative data.
4. **Economical:** Using digital tools to study online communities proves to be much more cost-effective. It saves considerably on the budget for offline travelling, lodging, and spending time in the field.
5. **Data preservation:** Data can be saved and accessed easily for future reference. There is no generation loss in copying the data, and by storing it digitally, it is safe from physical harm.
6. **Ease of research:** One can interact with anyone from their home without disturbing the community. Another factor is that one may stand out from the community due to their physical appearance in the real world. Still, in the case of online communities, physical appearance can be very well hidden. Think of how you would stand out from a native African tribe in Kenya if you wanted to study them and how your appearance would affect your interactions with that community.
7. **Cross-platform access:** Unlike in the real world, cyber ethnography can also help the researcher track a community member outside that community. Using profile information, usernames, and email addresses, the researcher can identify key members and track their activities on other social media platforms and websites.

Despite so many advantages, the method also has some disadvantages as well:

1. **Computer skills**—the researcher needs to have at least basic computer skills to start Digital Ethnography. But the more they know, the better tools they can manage, like data mining software, and elicit richer data.
2. **Temporal nature of online data**—The data does not remain static in the case of onion space. The researcher needs to keep their own records in case it is removed later.
3. **Fighting the algorithm**—The algorithm on social media websites tracks your content preferences and curates the content it thinks you may like. This could bias your feed immensely and show you only a fraction of the content that exists.
4. **Bias of the participants**—Like other community participants, the researcher also needs to create an online avatar or profile to start interacting with other users. In this case, your choice of image and the information you provide can affect the participants' behaviour towards you, thus affecting your data.

5. **Privacy concerns**—Digital Researchers disagree regarding the concept of personal and public on the Internet. Each website takes a different approach to the content posted by users on their public platforms. It becomes the researcher's responsibility to make sure the participants are informed that they are being studied and their consent is obtained when using their content as data.
6. **Electronic eavesdropping** - many researchers opine that online ethnography is nothing but eavesdropping on individuals using technological tools.
7. **No clear set of guidelines**—Due to the rapid succession of technology and thus updates in web platforms, no clear set of guidelines has yet been established by any authoritative bodies. A broad framework has been curated by AoIR (Association of Internet Researchers), but there is no abiding factor connected to it.

12.5 ETHICS AND RESPONSIBILITIES

While allowing for anonymity and a covert functioning method, digital ethnography comes with its ethics. Following are a few of the set rules and responsibilities to be followed by an ethnographer:

1. **Consent**—The researcher needs to ensure that the participants know that they are being observed. Researchers suggest creating a profile that clearly states the purpose or posting this information on your profile page so that it is in the public space.
2. **Privacy**—If the researcher stumbles upon information that the individual may not want to be publicised, they should respect their right to privacy, withhold the information, and not make it part of their research.
3. **Secure data handling**—data pertaining to individuals should be handled safely and should not compromise an individual's safety. There are several digital tools available to encrypt your data for safer handling, and they should be used.
4. **Respecting community norms**—each community has its own set of rules. A researcher should not act disruptively and should abide by the community guidelines.
5. **Avoid bias** - an individual can't be completely devoid of bias. However, it is good practice for a researcher to be aware of their own biases and avoid them while observing, collecting data, and writing a report for scientific research.

Like any other research method dealing with respondents, ethnography also calls for the researcher to be mindful, responsible, and respectful towards all being observed, even if they may disagree with them individually. Following these practices enriches the data and, thus, the 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 are the various advantages of online ethnography?

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

2. List a few responsibilities of an online ethnographer.

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

12.6 LET US SUM UP

In this Unit, you were introduced to Ethnography as a research practice. You were taken through the features of the digital landscape and learnt how Ethnography is adapted for virtual populations. We also covered various methods of Digital Ethnography, ranging from Qualitative to Quantitative, and pondered the advantages and disadvantages of each method. Lastly, we discussed the ethics and moral code associated with Ethnography and how it adapts to digital platforms.

12.7 KEYWORDS

Cyberculture: Culture of a group or community formed with the help of digital technologies.

Cyberspace: Non-physical environment, which enables individuals to interact with each other using digital tools

Mobile Ethnography: This type of ethnography where mobile becomes the primary tool for observation and data collection.

Field notes: A compilation of observations made by the researcher while observing the community.

Informed consent: Consent obtained from the participants being observed after explaining the purpose of the study.

Online Community: a group of people congregating in the online space, bound by certain values.

Sociogram: Graphical representation of findings of Social Network Analysis

12.8 FURTHER READINGS

1. Hine, C. (2008). Virtual ethnography: Modes, varieties, affordances. *The SAGE handbook of online research methods*, 257-270.
2. Kozinets, R. V. (2015). *Netnography: redefined*. Sage.
3. Madden, R. (2022). Being ethnographic: A guide to the theory and practice of ethnography. *Being Ethnographic*, 1-100.
4. Srivastav, S., & Rai, S. (2022). Netnography in Social Networking Sites—An Exploration of Cybercultures in Consumer Groups. *International Journal of Media and Information Literacy*, 7(2), 572-577.
5. Varis, P. (2015). Digital ethnography. In *The Routledge Handbook of Language and Digital Communication* (pp. 55-68). Routledge.

12.9 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Online communities prove to be more heterogeneous than offline communities, as they have members from diverse geographical, economic, and social backgrounds. With the option of anonymity, online communities see uncensored participation compared to offline ones. Lastly, participation in online communities is due to a person's interest and less peer pressure, as it is a highly individualistic activity.
2. Collecting data in digital format helps in many ways. It is easy to archive without any loss in quality and can be accessed and edited from remote locations. There are also several digital tools available for qualitative and quantitative data analysis. The data also becomes easier to handle and share with other collaborating researchers. It can also be digitally encrypted to ensure its safety.

Check Your Progress: 2

1. There are two types of online interviews: Synchronous and Asynchronous. Synchronous Interviews are those where questions and responses occur simultaneously. Both interviewer and interviewee are present online at the same time. Asynchronous interviews allow the interviewer and interviewee to participate as and when free.
2. The process of data mining usually involves five major steps:
 - a. Selection of data
 - b. Pre-processing of data
 - c. Classification of data
 - d. Pattern assessment
 - e. Reporting the findings

Check Your Progress: 3

1. The following are the advantages of Online Ethnography as a research method.
 - a. Global reach
 - b. Flexibility in data
 - c. Versatility in data
 - d. Economical
 - e. Data preservation
 - f. Ease of researcher
 - g. Cross-platform access
2. The responsibilities of an ethnographer include:
 - a. Obtaining the consent of the participants
 - b. Ensuring the privacy of the individuals
 - c. Handling data securely
 - d. Respecting the norms of the community studied
 - e. Avoid bias in data collection and reporting

UNIT 13 NARRATIVE ANALYSIS

Structure

- 13.0 Introduction
 - 13.1 Learning Outcomes
 - 13.2 History and Definition of Narrative Analysis
 - 13.3 Functions of Narrative Analysis
 - 13.4 Social Dimensions of Narrative
 - 13.5 Collecting Narratives
 - 13.6 Analysing Narratives
 - 13.7 An example of narrative analysis
 - 13.8 Let Us Sum Up
 - 13.9 Further Readings
 - 13.10 Check Your Progress: Possible Answers
-

13.0 INTRODUCTION

Narrative lies at the heart of being human (Byatt, 2000). It pervades our everyday life. We are born into a narrative world, live our lives through narratives, and afterwards are described in terms of narrative. Narratives concern the human means of making sense of an ever-changing world. It is through narrative that we can begin to define ourselves as having some sense of temporal continuity and as distinct from others.

13.1 LEARNING OUTCOMES

After completing this module, you should be able to

- Know about the definition and history of narrative analysis.
 - Learn the functions and social dimensions of narrative analysis
 - Learn about how to analyse narratives.
-

13.2 HISTORY AND DEFINITION

According to narrative theory (Murray, 1999 & Sarbin, 1986), we are born into a storied world and live our lives through the creation and exchange of narratives. A narrative can be defined as an organised interpretation of a sequence of events. This involves attributing agency to the characters in the narrative and inferring causal links between the events. In the classic formulation, a narrative is an account with three components: a beginning, a middle and an end. Unlike an open-ended piece of discourse, a narrative has a finished structure. (Murray, 1999 & Sarbin, 1986).

History of Narrative Psychology

Within psychology, three classic texts marked the specific narrative turn.

1. **Narrative Psychology:** *The Storied Nature of Human Conduct*, edited by Theodore Sarbin (1986). This collection amounted to a manifesto for the transformation of psychology. He summarised the implications of this alternative model:

When we give accounts of ourselves or others, we are guided by narrative plots. Whether for formal biographies or autobiographies, psychotherapy, self-disclosure, or entertainment, we do much more than catalogue a series of events. Rather, we render the events into a story.

Narratives are not just ways of seeing the world; we actively construct the world through narratives and live through the stories told by others and by ourselves. They have ontological status.

The book edited by Sarbin (1986) also contains a chapter by Ken and Mary Gergen (1986), arguing that narratives are social constructions developed in everyday social interaction. They are a shared means of making sense of the world. Gergen and Gergen, 1984 identified three structures that organised many narratives –progressive, in which there is a movement towards the goal; regressive, in which the reverse occurs; and stable, in which there is little change.

2. **Acts of Meaning by Jerome Bruner (1990)** followed his earlier **Actual Minds: Possible Worlds** (Bruner, 1986).

In these books, Bruner argued that there are two forms of thinking: the paradigmatic and the narrative. The former is the scientific method based on classification and categorization. The alternative narrative approach organises everyday interpretations of the world in storied form. Bruner identified several defining properties of narrative, including the following:

- It is composed of a unique sequence of events, mental states, and happenings that involve human beings as characters or actors.
- It can be ‘real’ or ‘imaginary’.
- It specialises in the forging of links between the exceptional and the ordinary.

These properties help us understand narrative as a way of constructing reality, of making sense of something obscure or unusual.

1. **Narrative Knowing and the Human Sciences** by Donald Polkinghorne (1988). The book's most important feature was the opening of hermeneutic philosophy, particularly the work of Paul Ricoeur, to more widespread discussion within psychology.

In his classic work *Time and Narrative*, Ricoeur (1984) argued that since we live in a temporal world, we must create narratives to bring order and meaning to the constantly changing flux. Not only do we create narratives about the world, but narratives are also central to how we conceive of ourselves and our identity. Through narrative, we not only construct a particular connectedness in our actions but also distinguish ourselves from others.

During the 1980s and 1990s, the study of narrative became much more extensive within various fields of psychology within personality and human development studies; Dan McAdams (1985) argued that narrative is central to our self-definition. Within clinical psychology, there was a movement towards developing narrative therapy (Mair, 1989 & Neimeyer, 1995) based on exploring alternative stories. Of particular note, the study of narrative within psychology encouraged the growth of greater contact with the humanities (Fulford, 1999 & Joy, 1997) and with the other social sciences (Maines, 1993).

13.3 FUNCTION OF NARRATIVE ANALYSIS

According to Jonathan A. Smith, there are three functions of narrative analysis, which are as follows:

The primary function of narrative is that it brings **order to disorder**. In telling a story, the narrator tries to organise the disorganised and give it meaning.

The use of narrative is particularly pronounced in everyday understandings of disruption (Becker, 1997). We all encounter disruptions to our everyday routines. Such disruptions include personal, family, financial, and health problems. These challenges to our daily routines encourage attempts by us to restore some sense of order. Narrative is the primary means of restoring this sense of order.

The need to restore a sense of order following disruption is especially pronounced in Western society, which is bounded by order and rationality. Gaylene Becker (1997) has argued that Western ideas about the life course emphasise linearity. Living in such a world, we try to make sense of inconsistencies. (Becker, G.1997).

Narratives not only bring order and meaning to our everyday lives but also **structure our very sense of selfhood**. We tell stories about our lives to ourselves and others. As such, we create a **narrative identity**. ‘Subjects recognize themselves in the stories they tell about themselves’ (Ricoeur, 1988). Narrative identity connects to different social relationships. It also provides a sense of localised coherence and stability.

Through narrative, we begin to define ourselves and clarify the continuity in our lives. In constructing a personal narrative, we select certain aspects of our lives and connect them with others. This process of narrative identity formation is dynamic and occurs in a changing social and personal context. Ricoeur (1988) said, ' We learn to become the narrator of our own story without completely becoming the author of our life.' (Ricoeur, 1988).

13.4 SOCIAL DIMENSIONS OF NARRATIVES

Narrative accounts are not emitted in a vacuum but are encouraged and shaped by certain social contexts. However, the narrator tells the story, and the story's character will depend upon whom the story is being told to, the relationship between the narrator and the audience, and the broader social

context (Murray, 1997). Although narrative is often considered in individual terms, one can also consider group, community, or social narratives. These are the narratives that collectives talk about themselves, their histories, and aspirations. As personal narratives are involved in creating personal identities, the social narratives define the history of a collective and distinguish it from other collectives. Further, these collective narratives overlap with personal narratives so individuals can define themselves as part of the group (Murray, 2000). In sum, we are enmeshed in a narrative world; we understand our world and ourselves through narrative.

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. Explain the basic tenets of narrative analysis?

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

2. Enumerate the steps involved in narrative analysis?

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

13.5 COLLECTING NARRATIVES

The primary source of material for the narrative researcher is the interview. Unlike the traditional structured interview with details of questions to be answered, the narrative interview is designed to provide an opportunity for the participant to give a detailed narrative account of a particular experience. The life story interview is the most extended version of the personal narrative interview. As its name implies, a life-story interview aims to encourage the participants to provide an extended account of their lives. At the outset of the interview, the researcher will explain that the study aims to learn about the person's life. However, narratives are not just life stories in the most general sense but also stories about everyday experiences, especially disruptions of daily life. Given the time and the opportunity, participants are often willing to provide extended narrative accounts of different experiences.

The challenge is for the researchers to convince the participants that they are interested in their narrative accounts. Thus, the researcher should reflect upon what the participants say and introduce supplementary questions designed to obtain clarification, such as, 'Why do you think that is the case?' or 'Could you give an example of that?' The interviewer can also use other methods

such as encouraging the participants to keep a personal journal to collect photographs or make a video. The aim is always to find a technique with which the participants are comfortable, and which will allow them to develop narrative accounts.

The researcher should collect background material about the central participants and details about the interviewer. Such

information is important when we begin to analyse the narrative accounts. A useful strategy is for the researcher to keep a detailed log of each interview. This could include some basic demographic details of the participant and when and where the interview occurred. (Smith, J.A. 2003).

13.6 ANALYSING NARRATIVES

The analysis of narrative accounts can be divided into two broad phases:

- Descriptive
- Interpretive

A thorough reading of the narrative precedes both these phases. In reading the narrative accounts, the aim is to familiarise oneself with both their structure and their content. A useful strategy is to prepare a summary of the narratives that will identify the key features, such as the beginning, the middle and the end. The analyst can highlight key issues and discern subplots. In reading the summaries, it is then possible to begin to get an idea of what the main issues being raised are (Mishler, 1986). Through this process of close reading, a coding frame can be developed that can be applied to the various narratives. The coding frame is designed to capture the overall meaning of the narratives and the various issues raised within each.

The second step is to connect the narrative with the broader theoretical literature used to interpret the story. Thus, the researcher goes beyond the descriptive phase to develop the interpretation. This requires a simultaneous familiarity with the narrative accounts and with the relevant literature such that one can begin to connect with the other.

A particular concern in narrative analysis is how the narrative is structured or organised. Various schemes have been developed to convey the temporal quality of narratives. The three-fold classification scheme developed by Gergen and Gergen (1984) is a useful analytic tool, but it is important not to apply it in a schematic way but rather in a flexible manner to encapsulate the various shifts in any narrative account. For example, the tragic narrative begins with a progressive structure, but then, despite the struggle, the central character is overcome, and the narrative becomes regressive. This regression can be overcome by changing the broad interpretive dimensions used to frame the event.

Gee (1991) described the value of exploring the poetic structure of popular narrative accounts. He argued that versus is an intrinsic part of everyday narrative accounting and that poetry is merely a more developed form of that accounting. In particular, he was concerned about using rhythm and metaphor

in popular narratives. The study by Becker (1999) is an example of the successful use of strategy to explore personal narratives. In reading through the pain narrative of an older adult, she noted that it had a certain poetic quality. She then recast the narrative account as a series of poetic stanzas with a similar structure. In recasting the narrative, the interviewer's questions are omitted, and the researcher organises the text into verses. This analysis requires attention to the overall rhythm underlying the narrative and the metaphors used to describe experiences.

The researcher can also consider the personal, interpersonal, group and social contexts (Murray, 2000). The personal context is concerned with how narrative draws on the individual's experience; the interpersonal and group context considers the audience and the co-construction of the narrative, and the societal context considers the broader societal narratives which structure our everyday accounts. While it is difficult to integrate all these contextual levels into a single analysis attention to one or the other may be particularly important in understanding the structure of certain narrative accounts.

13.7 AN EXAMPLE OF NARRATIVES

An Example: A Breast Cancer Story

The example is taken from a study of how women handle the disruption of their lives because of having breast cancer (Murray, 2002). Researchers were interested in how the women integrated the disease into their everyday lives. Also, how stories were constructed in a particular social and interpersonal context; all the women interviewed had had surgery for breast cancer. There was no sign of recurrence at their last check-up, and they had agreed to be interviewed about the experience. For many of the women, the interview was an emotional experience. Several of them mentioned that they had had limited opportunity to discuss the operation with others.

A summary of each of the narrative accounts was prepared. Certain commonalities in all the stories gave them the standard narrative structure.

1. **Beginning:** This was life before cancer. Different women emphasised aspects of their lives – family life, marriage, work etc. The main thing was that cancer did not play a part in their lives.
2. **Middle:** A major part of the story centres on the diagnosis of cancer, the surgery, and the reaction of the patient and that of their family or friends.
3. **End:** This involved looking back on the disruption in their lives, how they began to redefine themselves as the survivors of the disease, and how their life expectations and experiences changed.

Having developed the analysis of the narratives, we can then proceed to write a report grounded in the interviews. This example of a narrative account illustrates how people can use narratives to forge links between the exceptional and the ordinary (Bruner, 1990). When given an opportunity in the interview, the women were eager to provide detailed narrative accounts. In terms of structure, women's stories have a classic beginning, middle, and end. The beginning sets the scene; the middle details the experience of breast

cancer, and the end concerns the impact of the disease on their lives.

At the personal level of analysis, the narrative reflects the different experiences of the women. At the interpersonal level, the narrative analyst is interested in how the participant conveys her story to the interviewer. In her story, what issues does she emphasise? At the societal level, these narrative accounts mesh with the broader moral universes of women. From a moral perspective (Noblitt & Dempsey, 1996), narrative ties can be seen as tales of right and wrong and attempts to do the virtuous thing during times of challenge.

Narrative research offers extensive opportunities and is still being developed. When conducting a study, researchers should ask what they are trying to understand, what the participants are trying to say, and why they are trying to say that. The aim is to reveal the underlying structure of narrative accounts that shapes not only the way we account for our actions and those of others but also our very identity.

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. Enumerate the advantages and disadvantages of narrative analysis?

.....

2. Explain the application of narrative techniques in digital media research?

.....

13.8 LET US SUM UP

At the core of human existence lies narrative, which serves as a fundamental tool for comprehending the dynamic nature of our surroundings. Through narrative, we construct a framework that allows us to establish a sense of continuity within our lives and delineate our individual identities amidst the diversity of human experiences. Narrative theory posits that from birth, we are immersed in a world woven with stories, and our existence unfolds through the crafting and sharing of narratives, which serve as organized interpretations of sequential events. These narratives play a pivotal role in imposing order upon chaos, thereby furnishing structure to our understanding of selfhood. Moreover, it's imperative to recognize that narrative accounts are not crafted in isolation but are deeply influenced by the social milieus in which they emerge and evolve. In the realm of narrative analysis, scholars undertake a dual approach, comprising descriptive and interpretive phases, to unravel the intricate layers of meaning embedded within these narratives.

13.9 FURTHER READINGS

1. Bamberg, M. (2020). Narrative analysis: An integrative approach. Qualitative analysis: Eight approaches for the social sciences, 243-264.
2. Bandeli, K. K., Hussain, M. N., & Agarwal, N. (2020). A Framework towards Computational Narrative Analysis on Blogs. In Text2Story@ ECIR (pp. 63-69). <https://ceur-ws.org/Vol-2593/paper8.pdf>
3. Cowger, T., & Tritz, J. (2021). Narrative analysis research: A tool for extension educators. The Journal of Extension, 57(6), 1. <https://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=1259&context=joe>
4. Colussi, J., & Reis, T. A. (2020). Immersive journalism: a narrative analysis in virtual reality apps. Revista latina de comunicación social, (77), 19-32.
5. Josselson, R., & Hammack, P. L. (2021). Essentials of narrative analysis. American Psychological Association.
6. Pasquinelli, C., & Trunfio, M. (2020). Overtouristified cities: An online news media narrative analysis. Journal of Sustainable Tourism, 28(11), 1805-1824.

13.10 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Narrative analysis examines stories to understand underlying meanings, structures, and cultural contexts. It focuses on the narrative's construction, including plot, characters, and themes, and explores how narratives shape individual and collective identities, beliefs, and experiences. This approach acknowledges narratives as powerful tools for sense-making and communication in society.
2. Steps in Narrative Analysis:
 1. Data Collection: Gather narrative texts, such as stories, interviews, or media, relevant to the research question.
 2. Familiarisation: Read or listen to narratives multiple times to become familiar with content and context.
 3. Coding: Identify recurring themes, characters, and plot elements using coding techniques.
 4. Pattern Recognition: Analyse coded data to identify patterns, connections, and variations within narratives.
 5. Interpretation: Interpret findings by considering cultural, historical, and social contexts, and examining how narratives construct meaning and influence perceptions.
 6. Validation: Validate interpretations through peer review, participant feedback, or triangulation with other data sources.

1. Narrative analysis offers deep insights into individual experiences and societal discourses, allowing researchers to uncover complex meanings embedded within narratives. It facilitates understanding of cultural contexts and power dynamics. However, it can be time-consuming, subjective, and challenging to generalise findings beyond specific contexts.
2. In digital media research, narrative techniques are applied to analyse user-generated content, online storytelling, and digital narratives. Researchers examine how individuals construct and share narratives through social media, blogs, and online forums. This approach helps understand identity formation, community building, and cultural expressions in digital environments.

Block

4

ANALYSING AND INTERPRETING DATA

UNIT 14

Fundamentals of Data Analysis

UNIT 15

Digital Data Analysis

UNIT 16

Data Interpretation

UNIT 17

Report Writing

UNIT 18

Ethics in Internet Research

BLOCK 4 INTRODUCTION

In the first Block, you learned about the basics of Internet research. The subsequent two blocks are meant to help you understand different data collection tools. Now that the required data is being collected, this Block will help you understand the data analysis and interpretation process, professionally present your findings to a broader audience, and, more importantly, how to maintain ethical standards in your entire research journey.

Unit 14 Fundamentals of Data Analysis, where we learn about the core concepts shaping modern research methodologies. We explore diverse research perspectives, from handling quantitative data to qualitative analysis techniques. We unravel the intricacies of organising, organising, summarising, and testing hypotheses, culminating in drawing robust conclusions through varied analytical strategies and approaches.

Unit 15 Digital Data Analysis explains the intricate world of data retrieval and harvesting techniques within Digital Media Research Methods (DMRMs). Navigating the vast expanse of digital information empowers researchers to maintain focus and objectivity.

Unit 16 Data Interpretation discusses the fundamental principles of statistics, exploring populations versus samples, scales of measurement, and types of variables. We'll navigate through frequency distribution, measures of central tendency, and variability, empowering you to interpret data confidently and precisely.

Unit 17 Report Writing explores the essential stages of crafting a comprehensive report, from structuring the beginning with a captivating title to effectively presenting findings, including the main body and final components.

Unit 18 Ethics in Internet Research explains the fundamental principles of research ethics and defines the research process. Exploring digital platforms as research sources, we'll scrutinise source credibility and maintain source confidentiality. Lastly, we'll assess the advantages and disadvantages of Internet research.

These four Blocks of this course were designed to facilitate a better understanding of undertaking research on specific aspects of the digital media world. The ever-changing online platform would inspire you to explore nuances and understandings of human interactions.

UNIT 14 FUNDAMENTALS OF DATA ANALYSIS

Structure

- 14.0 Introduction
 - 14.1 Learning Outcomes
 - 14.2 Different Research Perspectives
 - 14.3 Handling Quantitative Data
 - 14.3.1 Organising and Coding Data
 - 14.3.2 Descriptive Statistics
 - 14.3.3 Summary Statistics
 - 14.3.4 Descriptive Measures
 - 14.4 Testing Hypotheses
 - 14.5 Qualitative Data Analysis
 - 14.6 Drawing Conclusion Through Data Analysis
 - 14.6.1 Constant Comparative Technique
 - 14.6.2 Analytical Induction Strategy
 - 14.6.3 Generalising About Generalisation
 - 14.6.4 Converging Practices and Approaches
 - 14.7 Let Us Sum Up
 - 14.8 Keywords
 - 14.9 Further Readings
 - 14.10 Check Your Progress: Possible Answers
-

14.0 INTRODUCTION

Data analysis involves processing raw data and converting it into information useful for users' decision-making. The data are analysed to answer questions, test hypotheses, and prove or disprove theories. The methods cover diverse operations of categorising, segmenting, and interpreting evidence, or data sets.

The process of data analysis by the researcher begins when an entire set of data has been collected. Organising the data correctly can save a lot of time and prevent mistakes. Most researchers use a database or statistical analysis programme to format data to fit their needs and organise their data effectively. Each data analysis software has its requirements for how data are to be entered. A good researcher enters all the data in the same format and the same database, as doing otherwise might lead to confusion and difficulty with the statistical analysis later. Sometimes, there are issues of interoperability between different software. Once the data have been entered, the researcher must check the data for accuracy. This can be accomplished by spot-checking a random assortment of participant data groups. Still, this method is less effective than re-entering the data a second time and searching for

discrepancies. This method is particularly easy to use when using numerical data because the researcher can simply use the database programme to sum the spreadsheet's columns and then look for total differences. One of the best methods of accuracy checking is to use a specialised computer programme to cross-check double-entered data for discrepancies (as this method is error-free). However, these programs can be hard to come by and may require extra training.

In this unit, we will focus on data analysis in research and the process of organising and coding data for effective analysis. We will also discuss how to distinguish between quantitative and qualitative data analysis so that you can apply appropriate data analysis techniques in your research work.

14.1 LEARNING OUTCOMES

After completing this unit, you should be able to:

- Describe the importance of data analysis in research;
- Organise data for effective analysis;
- Explain the importance of data analysis;
- Distinguish between quantitative and qualitative data analysis;
- Explain how to handle quantitative data;
- Discuss how to handle qualitative data; and

14.2 DIFFERENT RESEARCH PERSPECTIVES

The understanding of the term 'analyses' has tended to divide qualitative and quantitative research traditions. In quantitative media studies, rule-governed procedures of segmenting and categorising the components of content form the basis of interpretation. On the other hand, qualitative studies typically emphasise meaning as a process unfolding in contexts.

Using the words 'quantitative' or 'qualitative' can refer to a broad approach to the research problem, a methodology for the research and a defined set of research techniques. When we refer to quantification, the research analysis focuses on numbers and, in the case of content analysis, on measurable numerical values for 'space' and 'time'. These appear more systematic and replicable. Survey research also relies on greater quantification and greater measurement sophistication. However, this sophistication comes with a price: increasing quantification narrows the types of questions that can be addressed. This means that research depth is sacrificed to get research breadth. In quantitative data, we get a lot of findings that can be generalised to a larger population. In contrast, in qualitative data analysis, we get depth and underlying causes without being able to generalise to a larger population.

For this reason, there is no commonly accepted definition of the term 'qualitative'. Qualitative research relies mainly on observations and verbal data that reflect everyday experiences. On the other hand, and consequently, qualitative data must be analysed differently.

Check Your Progress: 1

- Note:** 1) Use the space below for your answer.
2) Compare your answers with those given at the end of this unit.

A. Fill in the blanks:

1. is defined as the processing of collecting raw data and converting it into information useful for users' decision-making.
2. The data is analysed to answer questions and tests. Or disprove theories.
3. Once the data has been entered, the researcher must check the data for.....
4. In quantitative media studies, rule-governed procedures of and , the components of content, form the basis of interpretation.
5. Qualitative research relies mainly on verbal data that reflects everyday experiences.

14.3 HANDLING QUANTITATIVE DATA

Quantitative research methodologies generate numerical data. Surveys (whether of audiences or content) and experiments are the basic data collection methods. Once numerical data have been collected, they need to be analysed through statistical techniques. These mathematical techniques describe, organise, and explore relationships within the data. In epistemological terms, quantitative research is typically grounded in a hypothetico-deductive approach, in which investigators mount hypotheses or predictions about the expected associations or cause-effect relationships between variables. The quantitative data collection and analysis aim to produce findings that lead to the acceptance or rejection of a specified hypothesis. Numerical data analysis through statistical procedures represents a systematic and objective way of determining whether significant patterns of relationships exist among those phenomena that have been measured in data collection.

14.3.1 Organising and Coding Data

Numerical data is collected through survey questionnaires, content-coding frames, or experimental instruments. These must be systematically coded and entered a database for further analysis. This process of organising and coding data is a crucial part of the analysis process. This is where most inexperienced researchers tend to make mistakes. If there are errors here, they will be reflected in the problems faced while analysing data, making analysis difficult, if not impossible.

The first stage of organising data is the creation of a coding manual. Often, researchers think that data can be entered directly into a database. While, yes, it can be entered, the coding manual anticipates any problems with entering, e.g.,

When multiple answers are possible or when there are double-digit replies. Simply stated, a coding manual is a set of instructions on how the data are to be tabulated. It is an extremely important document when organising data, especially if more than one person is entering data onto a coding sheet or database.

You may ask what a coding manual looks like. A brief design is given below:

Variable	Numerical value	Column
Location	1) Rural 2) Semi-urban 3) Urban 9) No Answer	1
Sex	1) Male 2) Female 3) Transgender 9) No Answer	2
Level of Education	1) Illiterate 2) Up to 8th standard 3) Matriculate 4) High School 5) Graduate and Above 9) No answer	3

Note how one column is assigned to one variable. And since a single digit cannot exceed 9, 9 is the code for a 'no answer'. If there are more than nine possible answers to a question, assign two columns, with 99 for a 'no answer'.

Sometimes, a question may have multiple answers because that is how it has been asked. Then, one column must be assigned for each possible answer.

A sample coding sheet is provided in the Annexure. In a coding sheet, one row is assigned to one respondent. One or two columns (as the need may be) are assigned to a question or variable. Through this coding process, key variables can be identified easily, and errors, duplication, repetition, etc., can be addressed. Once the data are coded numerically entered a coding sheet, and transferred onto a computerised database, data can be checked for accuracy, and statistical tests can be performed. The accuracy with which the data transfer takes place and the application of statistical procedures that are appropriate for the data are crucial to the entire quantitative research work—errors in the early stages as it can skew or invalidate the results.

Different software packages have different methods for entering data into columns. For instance, in Excel, it is possible to enter two numbers in one cell, separating them into 1 and 3. In SPSS, one cell in a column can have only one number. Follow the instructions of the software package in designing your coding manual and entering your data.

14.3.2 Descriptive Statistics

Descriptive statistics are discussed in detail in Unit 14, and we will merely summarise different types of statistics here. Quantitative data analysis often begins by adopting a simple descriptive approach to establish initial patterns in the findings. For instance, a public opinion survey about the performance of elected political leaders might present data in percentages of respondents who agreed or disagreed that the political leaders were competent. In a further analysis, results can be described in terms of gender, age, social class, and political affiliations of the respondents. Such results can be visually displayed in a bar chart or summarised in a table. Some of the ways of displaying results are described.

Another type of study might be conducting a survey to ask 1000 respondents to state how many hours they spend watching television each week. In this case, descriptive statistics can be applied to show how many respondents viewed nothing or less than an hour, between one to two hours and so on. Next, a frequency distribution could be generated, showing how the respondents were distributed across different viewing volumes. Such data can be visually represented in a line graph bar chart, or table.

14.3.3 Summary Statistics

Data can be further analysed in terms of summary statistics. Summary statistics measure two basic aspects of the distribution of ' scores or measurements in a data set: central tendency, dispersion, or variability. A central tendency measure indicates which out of a range of scores is the 'typical' one. This typical score in turn can be defined in three different ways:

- a) The mode is the most frequently occurring in a range of scores. For example, if in a set of ten scores, five score '4', three score '2 ', and two score' 1', the mode is 4.
- b) The median score is the midpoint in a range of scores. In the following scores, the median is "7": 2 4 5 6 (7) 9 10 11 12. The score '7' lies at the exact halfway point in this distribution of scores. In other cases where there is an even number of scores and therefore no exact mid-point, the median must be calculated by averaging between the two centre scores: 4 5 7 8 9 11 12 13. Here, the median is 8.5 or the average of '8' and '9'.
- c) The mean score is the average of the total range of scores. In the previous example, the eight scores totalled 62, divided by 8, giving a mean score of 7.75.

14.3.4 Descriptive Measures

Other fundamental descriptive measures are range, variance, and standard deviation.

Range, the simplest dispersion expression, is the difference between the highest and the lowest scores in a particular distribution.

Variance provides a mathematical index of the degree to which scores deviate from the mean score. It tends to be expressed not in terms of the original scores but as squared deviations from the mean. To compute the variance,

one subtracts the mean of a distribution from each score and then squares the result. These squared scores are then summed and divided by the number of original scores minus one. Variance is a powerful and widely applied measure.

Standard Deviation is a statistical procedure that enables us to see how much the group members differ from the mean value for the group. A low standard deviation means most group members are very close to the average, while a high standard deviation score means the numbers are spread out. Unit 14 explains the procedure for arriving at a standard deviation at length.

14.4 TESTING HYPOTHESES

Much quantitative research goes beyond the simple description of data and their distributions. The search for connections or correlations to understand the underlying causes of behaviour is an important part of any data analysis. For this, the researcher may have framed specific questions or hypotheses. The data must be tested to validate or disprove the hypotheses framed at the beginning of the research. In hypothesis testing, the researcher is interested in establishing whether two or more variables are associated or whether the scores obtained in two or more groups are different - and, if so, whether such association is in a 'significant' manner. Significance means simply that the association of the variables must not happen by chance or the laws of probability alone. When a relationship is highly significant, such an association would not have happened in the normal course of things, and there is an important association between the variables.

When deciding whether to accept or reject a hypothesis, the researcher must examine the statistical significance of the results. The starting point is to set up a null hypothesis, or a hypothesis which asserts that any statistical differences or relationships that emerge within the data are entirely due to chance fluctuations or random error. The research hypothesis puts up an alternative viewpoint- that the statistical relationships or differences are not due to chance but represent real phenomena that can be explained theoretically in the research.

The essential characteristic of quantitative research is that it reduces phenomena to numerical codes. Numerical measurement, however, can occur at more than one level, and understanding different levels of measurement is crucial to the correct use of statistical methods of data analysis and to the proper interpretation of data.

Within mass communication research, hypothesis testing is done using tests of association and significance. The term "association" is closely related to the term 'correlation'. A commonly used test of significance is the chi-square test for independence. Also called the Pearson's chi-square test or the chi-square test of association, this test is used to discover if there is a relationship between two categorical variables. Simply put, the test may show us if there is a relationship between the sex of a respondent and the amount of television s/he watches or his/her social media habits.

Tests of significance are also used in mass communication to tell us how strong a relationship between two variables is. The types of significance tests commonly used are the T-test, the F-test, and the analysis of variance (ANOVA). Essentially, these tests tell us about the significance of the difference between the mean of a small sample and the hypothetical mean of the population (expressed in terms of standard error).

You don't need to know the formulae to perform these tests and arrive at some scores at this stage. You can program the software you are using, which will calculate the scores for you and give you the score, the P value, and the significance level. However, it is important for you to know how to read and interpret the scores.

The first step is to determine if the association is significant. The second step is to determine which variables have the most impact. To do this, compare the p- value to the significance level on a chart available in any book on basic statistics. For a chi-square test, if the significance level is less than .05, you can conclude that an association between the variables exists. Different tests have different significance levels. You should either look up a book on basic statistics or consult a statistician to get an accurate picture.

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.

A. Fill in the blanks:

1. In epistemological terms, quantitative research is typically grounded in a approach.
2. The aim of the quantitative data collection and analysis, then, is to produce findings which lead to the..... or..... of a specified hypothesis.
3. The.....is the most frequently occurring in a range of scores.
4. The median score is the.....in a range of scores.
5.is the difference between the highest and the lowest scores in a particular distribution.

14.5 QUALITATIVE DATA ANALYSIS

Qualitative studies are different as they draw their strength not from numbers but from answering the questions 'why' and 'how'. A quantitative study will provide vast amounts of data for analysis, while a qualitative study will enable in-depth analysis.

Qualitative data comes in a variety of forms, such as field notes made while observing in the field, interview transcripts, documents, diaries, and journals. Also, during the course of the study, the researcher accumulates a large volume of data. Organising, analysing, and making sense of this data poses special challenges to the researcher.

Unlike the quantitative approach, where data analysis does not begin until all the numbers are collected, data analysis in qualitative research is a continuous process that begins as early as the time of data collection and continues throughout the project.

Another important difference between the qualitative and quantitative methods is that while in quantitative analysis, hypotheses are developed prior to the study, in qualitative analysis, researchers use an inductive method. The data collected using this method are grouped into appropriate and meaningful categories; explanations emerge from the data.

Organising the data

The researcher is the main instrument in qualitative data collection and analysis and must prepare before beginning the investigation. Some social scientists describe this preparation as *epoché*, the process by which the researcher tries to become aware of and remove, as far as possible, prejudices, viewpoints, or assumptions that might interfere with the analysis. *Epoché* helps the researcher put aside personal viewpoints so that the phenomenon under study may be seen for itself.

The researcher generally organises the information along a temporal dimension to facilitate working with large amounts of data generated by qualitative analysis. This means that the data are arranged in chronological order according to the sequence of events that occurred during the data collection process. Further, each piece of information is coded to identify the source, and multiple copies of notes and all other documents are gathered.

Working with vast amounts of data can be very difficult and time-consuming. Today, there are several software applications that help you collate and organise the data collected. However, the process, whether done by hand or through the use of software, remains more or less the same, and it needs to be understood.

Let us take this process further step by step.

Identify the main themes: Carefully go through all the descriptive responses given by respondents to each question and understand the meaning they communicate. Once these are understood, broad categories or themes can be developed to classify data and analyse responses. The researcher develops the broad themes that reflect these meanings from these responses. These themes become the basis for analysing the text of unstructured interviews.

Assign some codes to the main themes: If the researcher wants to count the number of times a theme has occurred in an interview, s/he needs to select a few responses to an open-ended question and identify the main themes. S/he continues to identify these themes from the same question till a saturation point is reached. Write these themes and assign a code to each of them using numbers or keywords.

Classify responses under the main themes: Go through them and fit them into the identified main themes.

Integrate themes and responses into the findings: Link the qualitative responses with quantitative measures - either validating the number or developing an understanding of why people responded in a particular way. While discussing the main themes that emerged from their study, some researchers use quotations to keep the feel of the response. Such quotations, especially from case studies, add depth to the understanding of the problem. The themes must connect to the objectives and research questions in the report.

It is important to note that, in qualitative research, hypotheses are not tested. It entirely depends on the way the qualitative data are analysed and reported.

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.

A. Fill in the blanks:

1. The three aspects of qualitative empirical research are and
2. In quantitative analysis, hypotheses are developed.....to the study.
3. In qualitative research, data analysis is a..... process.
4.is the process by which the researcher tries to remove or at least become aware of prejudices.
5. To facilitate working with large amounts of data, the researcher generally first organises the information along..... a dimension.

14.6 DRAWING CONCLUSION THROUGH DATA ANALYSIS

Qualitative data can be analysed using different techniques. The two popular techniques are constant comparative and analytical induction.

14.6.1 Constant Comparative Technique

This process involves the following four steps:

- Comparatively assigning incidents to categories
- Elaborating and refining categories
- Searching for relationships and themes among categories
- Simplifying and integrating data into a coherent theoretical structure.

After the data have been analysed, the researcher places each unit of analysis into a set of provisional categories. As each new unit is examined, it is compared to the other units previously assigned to that category to see whether its placement is appropriate. If some units of analysis do not fit into any pre-existing category, new classifications may be created. Units that fit into more than one category should be noted and placed into a new category

by finding similarities among units that fit into a category.

The researcher may write rules or propositions that attempt to describe the underlying meaning that defines a category. These rules or propositions may also be revised throughout the study. This helps the researcher focus and crystallise the study and explore the theoretical dimensions of the emerging category system.

In the next stage, the researcher may examine the propositional statements and find meaningful connections. Finally, the researcher writes or summarises the research findings and integrates them into a coherent explanation of the phenomenon of the study's objectives.

14.6.2 Analytical Induction Strategy

This method blends hypothesis construction and data analysis. This strategy defines the research topic, and the hypotheses are developed. The hypotheses may be refined or reformulated until they can be adequately tested. The researcher uses the analytical induction approach to explain a phenomenon.

Let us try to understand this with the help of an example. Suppose a researcher is interested in explaining why people watch home-shopping channels. The answer to this may be obvious: people do so to buy products. The researcher may not be satisfied but decides to use this as an initial hypothesis. S/he further looks out for a heavy viewer of these channels. During the interview, the respondent may say that the primary reason for watching the channels is to find new and unique products. With this input, the researcher can reformulate the hypothesis. Another viewer may answer this differently by emphasising that s/he uses the prices advertised on the channel to compare before shopping. The hypothesis is further redefined. As the researcher progresses, s/he finds cases that do not add to the existing reasons and hence cannot fit into the revised hypothesis. This method can be exhausting, making it difficult for the researcher to determine the exact stopping point.

Then what should be done? This is a tricky question for which there is no simple answer. One way, especially when researching an event such as a riot, is to limit one's findings to the main objectives of the research and to clearly state that this is being done. Another way is to define the limitations or scope of the research early in the designing phase so that it is easy to see what can be included and what cannot. Finally, if there is a time limit on the study (especially if funded by a donor) or as a time to submit the thesis draft, the time limit will indicate when to stop; remember that all research is finite in time - it does not go on endlessly, and the researcher must determine when it is best to stop.

14.6.3 Generalising about Generalisation

The literature and the theory of science recognise two different conceptions of 'generalisation'. Empirical or statistical generalisation refers to the capacity of quantitative methodologies to apply predefined (hypothetically deduced) categories to a representative set of empirical instances, thus supporting external validity. Theoretical or analytical generalisation refers to articulating

new concepts or categories, typically in qualitative inquiry, that conceive empirical instances in a more consistent and insightful manner, prioritising internal validity. Both these aspects are relevant for assessing findings and insights of qualitative and quantitative studies.

14.6.4 Converging Practices and Approaches

In recent decades, several research scholars have outlined ways to combine and mix quantitative and qualitative methodologies. Three principal forms have emerged as possible methods of combining qualitative and quantitative methodologies. These are facilitation, triangulation, and complementarity.

Facilitation: Traditionally, The most common practice has been treating qualitative and quantitative approaches as relatively separate steps in research. Facilitation aims to facilitate in-depth analyses of respondents by formulating interpretive categories. For example, in survey research, analytical strategies are arrived at that are conceptually precise and meaningful to the respondents. The mapping of respondents' social lives may be facilitated by subsequently drawing up interpretative categories after analysing the members in-depth.

Triangulation is a strategy for gaining several perspectives on the same phenomenon. It can be carried out through several data types, several investigators, and several methodologies. For example, experimental and observational approaches can be combined and analysed to interpret a phenomenon.

Complementarity: Here, different analytical categories and procedures may be appropriate for capturing aspects of the same empirical domain or for addressing two different domains that affect the same research question. In the final stage, two categorically different sets of findings can be joined concerning a common theoretical framework.

Check Your Progress: 4

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

A. Fill in the blanks:

1. The two popular techniques of qualitative data analysis are the..... and technique.
2. In the constant comparative technique, the researcher places each unit of analysis into a set of categories.
3. The method of analytical induction blends and data analysis.
4. Three principal forms of combining qualitative and quantitative methodologies are, and
5. is a strategy for gaining several perspectives on the same phenomenon.

14.7 LET US SUM UP

In this unit, we have discussed the purpose, importance, and significance of data analysis in research. Data analysis is defined as the processing of collecting raw data and converting it into information useful for users' decision-making. The data is analysed to answer questions, test hypotheses, or disprove theories. This chapter discusses various methods that are applied in data analysis, both quantitative and qualitative.

The understanding of analyses has tended to divide qualitative and quantitative research traditions. In quantitative media studies, rule-governed procedures of segmenting and categorising the components of content form the basis of interpretation. Meanwhile, qualitative studies typically emphasise meaning as a process unfolding in contexts. Ultimately, it is not an either/or situation. Each tradition must feed into and supplement each other to yield accurate results and logical explanations for phenomena. How data are gathered and analysed may differ, but both aim to understand a situation or solve a problem in the long run.

Quantitative data collection and analysis aim to produce findings that lead to the acceptance or rejection of a specified hypothesis. Numerical data analysis through statistical procedures represents a systematic and objective way of determining whether significant patterns of relationships exist among the phenomena measured in data collection. Descriptive and summary statistics and descriptive measures are used to analyse the data.

Qualitative data comes in various forms, such as notes made while observing in the field, interview transcripts, documents, diaries, and journals. The constant comparative technique and analytical deduction strategy are applied to analyse and draw conclusions from the data.

In recent decades, many scholars and researchers have suggested and outlined ways of mixing and combining quantitative and qualitative methodologies. These combined forms can enrich and reflect in-depth and breadth of the status of that knowledge.

With this unit, you have almost reached the end of the research process. The final stage is report writing or the presentation of your research for your peers, the academic community, the funding agency, or simply for your next academic degree. The last unit of this course explains how to write your report effectively.

14.8 KEYWORDS

Data analysis is the process of collecting raw data and converting it into information useful for users' decision-making.

External validity: The degree to which the results of a study are generalisable to other situations.

Epoche is the process by which the researcher tries to remove or at least become aware of prejudices.

Hypothesis: A tentative generalisation about the relationship between two or

more variables that predict an outcome.

Internal validity: The properties of a research study are such that the results are based on expected conditions rather than extraneous variables.

Operationalisation: A definition that specifies behaviour patterns and procedures to experience or measure a concept.

Summary statistics are statistics that summarise a great deal of numerical information about a distribution, such as mean and standard deviation.

Skewness is the degree of departure of a curve from the normal distribution (curves can be positively or negatively skewed).

Triangulation: It combines quantitative and qualitative approaches to solve a problem.

Qualitative research method: A research method that uses flexible questioning.

Quantitative research method: A research method that uses standard questioning.

14.9 FURTHER READINGS

1. Arthur Asa Berger, 2011, Media and Communication Research Methods, Sage Publications.
2. Anders Hansen & David Machin, 2013, Media and Communication Research Methods, Palgrave Macmillan.
3. Klaus Bruhn Jensen, 2012, A Handbook of Media and Communication Research: Qualitative and Quantitative Methodologies, Routledge, London.
4. Roger D. Wimmer and Joseph R. Dominic, 2010, Mass Media Research, An Introduction, Wadsworth.

14.10 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Data Analysis
2. Hypotheses
3. Accuracy
4. Categorising, Segmenting
5. Observations

Check Your Progress: 2

1. Hypothetico-deductive
2. Acceptance, rejection
3. Mode
4. Mid-point

5. Range

Check Your Progress: 3

1. Strategy, Tactics, Techniques
2. Prior
3. Continuous
4. Epoche
5. Temporal

Check Your Progress: 4

1. Constant comparative, the analytical induction
2. Provisional
3. Hypothesis construction
4. Facilitation, triangulation, and complementarity
5. Triangulation

CODING SHEET

UNIT 15 DIGITAL DATA HARVESTING AND ANALYSIS

Structure

- 15.0 Introduction
 - 15.1 Learning Outcome
 - 15.2 Digital Data Harvesting Techniques
 - 15.3 Digital Data Analysis Techniques
 - 15.4 Let Us Sum Up
 - 15.5 Further Readings
 - 15.6 Check Your Progress: Possible Answer
-

15.0 INTRODUCTION

In this Unit, you will learn about data retrieval (or extraction) and data harvesting techniques in Digital Media Research Methods (henceforth DMRMs). Understanding digital data retrieval techniques from the mess of information empowers a researcher to stay focused and objective. The information delivered through digital media platforms needs different information treatments. The techniques mentioned below would allow a researcher to collect results that are mostly below two thousand on Google and sometimes as low as sixty. However, the idea is not to have fewer results but to have the results according to a researcher's needs and research objectives.

First, you will learn the concept of data reduction, i.e. minimising the information overload and effortlessly sifting through the information maze on the Internet, thus arriving at your required information in Google. Secondly, you would become familiar with various websites and news sources' data extraction and harvesting tools. The chapter poses new challenges when designing your Google power search (henceforth GPS) commands to better tell Google or the search engine about your query. The availability of a few open-source resources and browser extensions would develop your sense of information discrimination. The development of the "art of navigation" in the information overload may be the biggest takeaway of this chapter.

15.1 LEARNING OUTCOME

After completing this Unit, you should be able to:

- Understand the different techniques of data harvesting
- Undertake a data collection exercise to extract information from the cyberspace
- Understand the different techniques of data analysis

15.2 DIGITAL DATA HARVESTING TECHNIQUES

Data is regarded as a fundamental unit and scientifically recorded factual piece of information which can be appropriated for further analysis. This way, the digital content available on online newspapers, websites, social media platforms, discussion forums, government documents, reports, etc., constitutes big data that can be harvested using relevant tools. According to Luo & 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 would acquaint yourself with the various data retrieval and harvesting techniques.

15.2.1 Digital Data Retrieval & Harvesting Techniques Amenable to Research

Data retrieval is defined as obtaining data from websites using commands and techniques. In contrast, data harvesting is defined as "scraping or downloading" data on our desktops for actual analysis. Many software and tools are available that scrape data as per your keywords from the entire web or the websites you desire, but most of them are paid, and the free ones may not serve your purpose fully. This section enumerates good free online resources. Once you have correctly identified the research problem and the objectives, you may proceed with the data retrieval or scrapping. Conventional methods like questionnaires, hard copies of newspapers or records offer little respite to a researcher in this information age. The ease of data retrieval through online questionnaires, news websites and freely available databases is astonishing. Web data retrieval techniques such as Google Power Searching, which apply across the web content, have reduced the data collection time to just a few minutes and the results to only a few hundred.

Like the power searching commands, many news websites have the option of "advanced search", wherein they provide a customised search for the news as per the time frame, keyword, word combination with keywords, an exact search of words in a sequence, deletion of non-desired words, the addition of desired words, news sources, region etc. thus giving you complete control over the content. In other words, these filters bypass the SEO (search engine optimisation) techniques and present data as per your query. A few useful Google power search (henceforth GPS) commands are mentioned here.

Basic search—First, entering a few words in Google, like **Word technology, power empowerment, and Internet police court judge**, would bring the results (documents) from around the world, with these words anywhere in the text, as shown in the image.

Google search results for the query "Word technology power empowerment internet police court judge". The search bar shows the query. Below it, the "All" tab is selected, followed by News, Images, Videos, and More. Settings and Tools options are on the right. The results page indicates about 11,70,000 results found in 1.07 seconds. The top result is a link to "Information Technology Act-Section 65. | Cyber Crimes Laws India" from https://cybercrimelawyer.wordpress.com/. The snippet describes Section 66 of the Information Technology Act, mentioning hacking, computer system, power to investigate offences, Senior Inspector of Police, Cyber Police Station, Crime Branch, CID, competent jurisdiction, and the Additional Metropolitan Magistrate, Court 37.

Exact search—The search "media censorship in India" (in double quotes) would bring results that have the exact sequence of words of the query. Without double quotes, every word is searched differently, and the results increase unnecessarily.

Google search results for the query "filetype:pdf technology AND media AND internet AND rural AND skill OR talent". The search bar shows the query. Below it, the "All" tab is selected, followed by News, Images, Videos, Maps, and More. Settings and Tools options are on the right. The results page indicates about 19,30,000 results found in 1.14 seconds. The top result is a link to "Scholarly articles for filetype:pdf technology AND media AND internet AND rural AND skill OR talent" from www.fao.org. The snippet discusses the Internet and rural agricultural development, preservice elementary teachers' technology skill, digital inequality, and Internet use as penetration.

Google search results for the query "\"media censorship in india\"". The search bar shows the query. Below it, the "All" tab is selected, followed by News, Images, Videos, Maps, and More. Settings and Tools options are on the right. The results page indicates about 27 results found in 0.63 seconds. The top result is a link to "Sec66A of IT Act and Social Media Censorship in India | My India" from www.mapsofindia.com. The snippet discusses Article 66A of the IT Act and social media censorship.

The image shows only 27 focused results, and a portion of the line "**“media censorship in India”**" is highlighted in bold, along with some context of words before and after. This would be the standard interpretation of the results for the other queries also as mentioned below. The double quotes put on the single word like "empower" would bring the results (documents) which have the "empower" word in pure form and not as "empowering" or "empowered", etc. Hence, it is advisable not to double quote a single word to get results with all the word derivatives.

site:www.hindustantimes.com technology AND rural AND scheme—This command, once typed in Google, would bring the news items within the Hindustan Times website, with the words "technology", "rural", and

"scheme" anywhere in the running text of news items. The selected words, such as "rural" and "technology," are highlighted in the result.

The "AND" is not a word but a Boolean operator, which must be in capital letters; otherwise, Google would count it as a word if it were used in a small letter. The continuity in the letters of the site command must be taken care of, and one space should be entered before and after the AND command, as shown in the image.

Filetype: pdf technology AND media AND Internet AND rural AND skill OR talent—This command would bring only pdf documents with the words "technology" and "media" and "Internet" and "skill" or the word "talent". The file type can be changed to include more, like docx, ppt, txt, xls, etc., as per your choice of data.

It should be noted that when we use the AND command, the results decrease as we limit the choices. When we use the OR command, the results increase because we give the choices. Also, as shown in the figure, when the search query tends to become academic, Google offers scholarly research articles on the query with good citations.

site:www.hindustantimes.com women AND technology—money—rural—
This would show news from the Hindustan Times' website with "women" and "technology" words only, thereby removing the news that had "money" and "rural" words in the text.

Here, the minus sign must be just before the next word and have a space after the previous word. It can also be used in multiples if you want to retrieve documents that don't have "money" and "rural" words. Another search with "nano"—car would not bring the results of nano car but the nanoparticles, etc.

site:www.tribuneindia.com in title crime AND women - The news would be from The Tribune newspaper, which only has "crime" and "woman" words in the news title. Hence, we have only 259 results which are highly focused.

Google search results for the query "site:www.tribuneindia.com intitle:crime AND women". The results show one main news article from The Tribune.

site:www.tribuneindia.com intitle:crime AND women

All Videos Images News More Settings Tools

About 259 results (0.81 seconds)

Conflict, modernisation fuel crime against Valley women - The Tribune
www.tribuneindia.com/news/community/conflict...fuel-crime...women/369842.html ▾
Feb 27, 2017 - The Tribune: SRINAGAR: "It is very unfortunate that there is high increase in the rape instances and ravenous maniacs are not even sparing ..."

site:www.tribuneindia.com in title crime intext: cyber OR Internet – In this, news from The Tribune would appear that has the word "crime" in their title and either "cyber" or "Internet" words in the running text of the news. Hence, a researcher can collect data from any publication's website and enter the desired query in the title and the text with the commands and can get the desired output. These 56 results, as shown in the figure, can manually (though not recommended) be opened, and saved as HTML files on the computer and printed for the study.

Google search results for the query "site:www.tribuneindia.com intitle:crime intext:cyber OR internet". The results show one news article from The Tribune.

site:www.tribuneindia.com intitle:crime intext:cyber OR internet

All News Videos Images Maps More Settings Tools

56 results (0.68 seconds)

Cyber crime police station opens in Shimla - The Tribune
www.tribuneindia.com/news/community/cyber-crime-police-station.../338235.html ▾
Dec 18, 2016 - Chief Minister Virbhadra Singh digitally inaugurated the state cyber crime police station at the police headquarters from his office here today.

inurl: MOOC AND course—This command brings all hyperlinks (URLs) in which the words "MOOC" and "course" appear, as shown in the first result in the picture.

Google search results for the query "inurl:MOOC AND course". The results show the MOOC List website.

inurl:MOOC AND course

All News Videos Images Maps More Settings Tools

About 2,63,000 results (1.02 seconds)

MOOC List | Massive Open Online Courses a complete list of free ...
<https://www.mooc-list.com/> ▾
A complete list of Massive Open Online Courses (free online courses) offered by the top universities and colleges in a wide range of subjects. MOOC it is an ...
Providers and Categories · Language · Find MOOCs By Multiple Criteria · Log in

site:.gov women AND empowerment—By entering this in Google, you will retrieve documents with the words "women" and "empowerment" within the .gov domain. Likewise, the search can be expanded to include .in, .edu, .au, .pk, and .sl (country-specific domain names).

site:.gov women AND empowerment

All Images News Videos Maps More Settings Tools

About 2,46,000 results (0.86 seconds)

Gender Equality and Women's Empowerment | U.S. Agency for ...
<https://www.usaid.gov/what-we-do/gender-equality-and-womens-empowerment> ▾
 Our Strategy. At USAID, we believe that gender equality and women's empowerment isn't a part of development but the core of development. Progress cannot be delivered in a vacuum. For societies to thrive, women and girls must have access to education, healthcare, and technology.

Here, it would be pertinent to know that by combining individual commands, e.g. site:.gov women AND empowerment OR skills -crime -police filetype:pdf, we can get much better and enriched results in Google.

Time filter – the data can be retrieved between a few dates, as shown in the figure. This command would be activated by clicking the "Tool" button, as shown in the figure.

site:.gov women AND empowerment

All Images News Videos Maps More Settings Tools

Any country ▾ Any time ▾ All results ▾

Gender E...	Any time	Empowerment U.S. Agency for ...
https://www.u...	Past hour	er-equality-and-womens-empowerment ▾
Our Strategy. ...	Past 24 hours	nder equality and women's empowerment isn't a part of
development t...	Past week	Progress cannot be delivered in a vacuum. For societies to
thrive, women...	Past month	to education, healthcare, and technology.
Empoweri...	Past year	nd Conflict Gender Equality and ...
https://www.u...	Custom range...	er...womens-empowerment/.../crisis-conflict ▾
Jan 18, 2017 -		in [PDF, 324K] lays out how the Agency will invest in gender
equality and w...		isis and conflict-affected countries to promote the rights and

Filetype: ppt women AROUND (5) technology – The search results in the form of PowerPoint presentations would appear in which the word "technology" appears within the proximity of 5 words left and five words right of the word "women". More commands, like OR, minus, and intext title, can be included here.

filetype:ppt internet AND women AROUND (5)technology

All News Images Videos Maps More Settings Tools

62 results (0.83 seconds)

Women in Technology - Read More at Internet Society
[Ad www.internetsociety.org/Women](http://www.internetsociety.org/Women) ▾
 Unsolicited Advice: How to Be a Fearless Girl in ICT. Find Out More
 Women in Tech & IT Empowering Women
 Become a Member of ISOC Internet Privacy in 2016

[PPT] Women In Technology Workshops AGENDA - Department of ...
www.cs.toronto.edu/~karolina/ibm-wit/facilitator.ppt ▾
 Women In Technology Workshops ... The Internet was invented in the 1960s. ____ 7. You should believe everything you see on the Internet. ____ 8. You have to ...

Related: www.census.gov—By applying this command, the other similar sites relating to the one entered (www.census.gov) will appear. This can be useful for finding similar stuff like databases, research publications, etc. It broadens the researcher's perspective.

Google search results for "Related:www.census.gov". The search bar shows the query. Below it, the "All" tab is selected. The results page displays the following information:

- Census.gov**
<https://www.census.gov/>▼
Learn how the U.S. Census Bureau serves America as the leading source of quality data about our people, business ... **Census** of Governments ... **Related Sites** ...
- Related Sites - U.S. Census Bureau**
<https://www.census.gov/topics/income-poverty/income/about/related-sites.html>▼
Mar 1, 2016 - Find sites related to this topic from other federal statistical agencies and trusted sources.
... **Census.gov** Income & Poverty Income About.

About 1,72,00,000 results (0.95 seconds)

All the above commands must be entered in Google, and the results are obtained. The command characters are currently limited to 2000. These GPS commands written here are simple and not complex, whereas the research data needs a complex command design (for effective data retrieval), as per our research objectives. The individual commands described above can be freely used in combinations. Hence, a typical complex search command may look as - **filetype:pdf inurl:"corpus linguistics" intext: computer AND Internet AND statistics OR valid**.

Google search results for "filetype:pdf inurl:"corpus linguistics" intext:computer AND internet AND s". The search bar shows the query. Below it, the "All" tab is selected. The results page displays the following information:

- What is Corpus Linguistics? - UCSB Linguistics - University of ...**
https://www.researchgate.net/...Corpus_Linguistics/.../What-is-Corpus-Linguistics.pdf▼
by ST Gries - Cited by 57 - Related articles
internet data'. Question: My marketing and advertising, **computer** and tech language – just google Java or Ruby and note how often the ties, scripting for web applications, and **statistics/statistical** graphics respectively, but all three are in ...

60 results (0.73 seconds)

The still more complex query can be designed as - **site:initial: technology-intext: crime AND women -inurl: cyber AND -bbc AND -journal AND -Internet AND -police AND -court AND -blog AND -science AND -news AND -student**.

Google search results for "site:.in intitle:technology intext:crime AND women -inurl:cyber AND -bbc AND -journal AND -Internet AND -police AND -court AND -blog AND -science AND -news AND -student". The search bar shows the query. Below it, the "All" tab is selected. The results page displays the following information:

- Indian Institute of Technology Bhubaneswar**
www.iitbbs.ac.in/wgrc-events.php▼
Workshop to End Violence against Women. APR 122013. Awareness Programme on Prevention of Crime against Women. MAR 082013. Observance of ...

18 results (1.14 seconds)

Country-Specific Search –By selecting the country from the tools (as shown), we can get the search results specific to that country. The picture shows the results from India.

"corpus linguistics" filetype:pdf

All Books News Images Videos More Settings Tools

Country: India Any time All results Clear

[PDF] Corpus Linguistics: An Introduction - eolss.net
www.eolss.net/sample-chapters/c04/e6-91-17.pdf ▾
by NS Dash - Cited by 30 - Related articles
LINGUISTICS - Corpus Linguistics: An Introduction - Niladri Sekhar Dash ... than half a century ago
Corpus Linguistics has started its journey as a field.

If we search The Hindu's website for simple words like **site:thehindu.com ombudsman Internet false journalism**, we will get the results (news) with these words.

site:thehindu.com ombudsman internet false journalism

All News Images Videos Maps More Settings Tools

26 results (0.51 seconds)

Self-criticism is integral to self-regulation - The Hindu
www.thehindu.com › Opinion › Readers' Editor ▾
Nov 7, 2016 - "Good journalism is not dying; it is getting better and bigger. It's just bad journalism makes lot more noise than it used to do five years ago," he ...

Searching the results in Hindi or another Non-English language –

If we use the command - महिला AND प्रौद्योगिकी AND इंटरनेट AND कौशल filetype:pdf then all results in Hindi would appear in pdf document format.

महिला AND प्रौद्योगिकी AND इंटरनेट AND कौशल filetype:pdf

All News Images Maps Videos More Settings Tools

73 results (0.83 seconds)

[PDF] NIELIT AR-2016 हिंदू-साल1.pdf
www.nielit.gov.in/sites/.../NIELITAnnualReportHindi2016.pdf ▾ Translate this page
मुझे राष्ट्रीय इलेक्ट्रॉनिकी एवं सूचना प्रौद्योगिकी संस्थान (नाइलिट) के 21वें वार्षिक प्रतिवेदन के साथ-साथ वर्ष 2015-16 के। लेखों का ... ग्रामीण महिलाओं तथा युवाओं और उद्यमियों व जनसामाज्य को आवश्यक कौशल प्राप्त करने में सक्रिय रूप से सहायता प्रदान कर रहे हैं ... चूँकि, हमारे जीवन के समर्त पहलुओं में इंटरनेट के प्रयोग में निरंतर बढ़ोतरी हो रही है, इसलिए नई...

The Google search needs non-English words (Hindi, Punjabi, etc.) in Unicode. If we go to www.translate.google.com, we can get the output in Unicode for the desired words.

Translate

Hindi English Spanish English - detected ▾

skill x

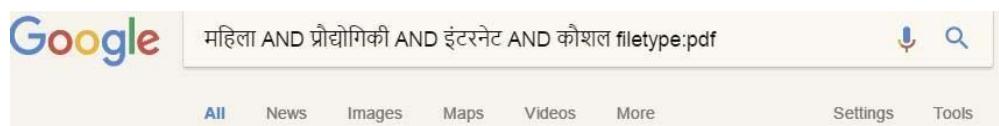
कौशल

6/5000

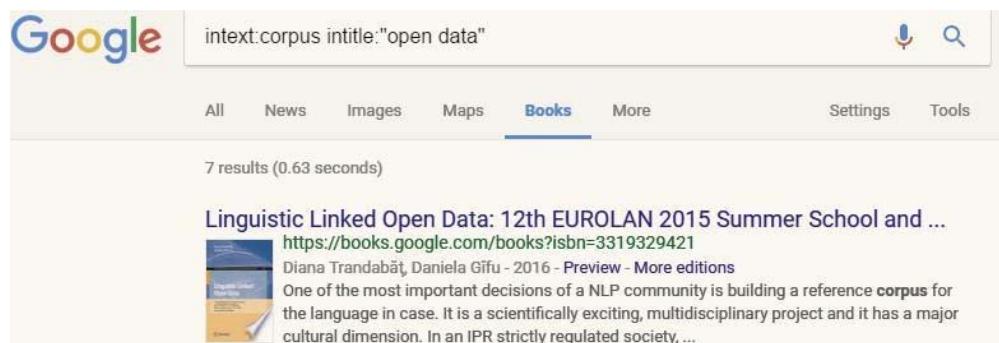
A few things may be taken care of while using GPS –

1. The commands can be combined irrespective of the sequence and length. However, the increased complexity may lose sight of the research objective.

2. You cannot self-design the commands as these are Boolean operators, and the Google servers recognise them as such (e.g. AND, not "and").
3. Spaces and continuity in the commands should be meticulously maintained, as the results may change drastically if the command is not entered properly.
4. Search can be done in regional languages, too.
5. One Google search releases 7 grams of carbon dioxide into the atmosphere; hence, it is advisable to use Google search with discretion.
6. The same set of commands would work on all Google products like www.scholar.google.com, YouTube, Google News (www.news.google.com) and Google Books (www.books.google.com)
7. Double quotes and AND & OR commands are universal commands like Ctrl V and Ctrl C. They work well in other search engines, too.
8. Increased complexity in commands beyond a limit is likely to produce wrong results, as we may not be able to comprehend the full design of the command.
9. Testing a few results after running the command is advisable to ensure the correct results.
10. If unsure, go to the "advanced search" in the "settings" and search from there, but it doesn't support all power commands.
11. Google, by default, searches for "All" (shown in the image). However, the search can be expanded to include images, videos, maps, books etc.



12. Google search with images is quite useful.
13. www.books.google.com would fetch highly beneficial results from millions of books available. The command written underneath produces only seven results, and when opened, the keywords are highlighted in the text to give a contextual snapshot for better understanding.



14. www.news.google.com is a very good resource for all online news (from all websites and sources). As shown in the figure, the site has an inbuilt search window for easy data retrieval.

The screenshot shows the Google News search interface. On the left, there's a sidebar with 'Top Stories' and various news categories like 'Arsenal F.C.', 'Rodrigo Duterte', etc. The main area has search filters for words, phrases, and dates. A search result for 'air Russia ties on' is displayed, showing a snippet about India repairing its partnership with Russia, links to related news, and small images of political figures.

Let's assume that the relevant content from the web has been obtained through Google search and needs to be harvested in the proper format. If the results are a few (in hundreds), then a manual approach of opening each result in a separate window can be tried, and data can be copied in separate text or word format. In case of a large number of results, researchers can always invoke the 5% error margin rule and arrive at the sample size of results out of the total population of results by entering the population size in the website www.raosoft.com/samplesize.html and the sample obtained.

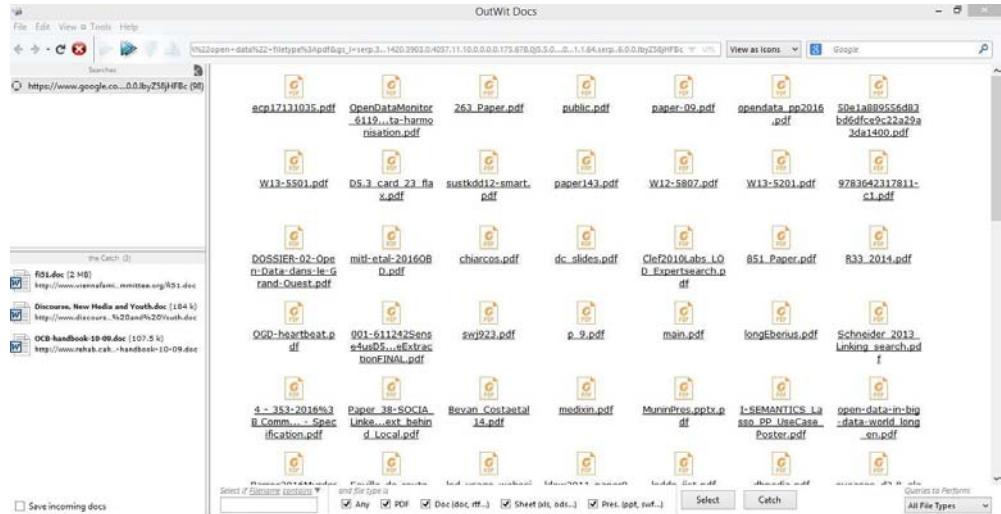
There are a few chrome extensions (e.g. linkclump) which, after applying the GPS command, can collect the headlines of the links and their URLs. The Headlines can be researched for content analysis. The output is shown in the image.

Book1 - Excel (Product Activation Failed)											
Clipboard		Font			Alignment			Number			Styles
FILE	HOME	INSERT	PAGE LAYOUT	FORMULAS	DATA	REVIEW	VIEW	DEVELOPER			
Cut	Copy	Format Painter	Font	Font	Font	Font	Font	Font	Font	Font	Font
Paste	Clipboard	Font	Font	Font	Font	Font	Font	Font	Font	Font	Font
1 Linguistic Linked Open Data: 12th EUROLAN 2015 Summer School and ...											
2 Using Open Data to Detect Organized Crime Threats: Factors Driving ...											
3 Open Data for Education: Linked, Shared, and Reusable Data for ...											
4 Open Data for Resilience Initiative Field Guide: - Page 3											
5 Open data: Accès, territoires, citoyenneté: des thématiques ...											
6 Open data e nativi digitali: Per un uso intelligente delle tecnologie											
7 Il Fenomeno Open Data: Indicazioni e Norme per un Mondo di Dati Aperti											

The individual URLs can be opened in www.textise.net to save the data and text file format, or they can be fed into Python (programming language beyond the scope of this chapter) and the text collected. Linkclump can be downloaded from the Google Play Store.

Application—Headlines can be collected and analysed using the corpus software.

. If you want to make a collection of PowerPoint files, use the "filetype: ppt" command (also pdf, doc, ppt, etc.) and run it in a freeware system called "outwit doc". All documents are downloaded effortlessly, which makes good content for researchers. This data can be easily read by the Free Desktop Search software (e.g., Copernic Desktop Search) in a similar way to Google commands. This immediately makes the context clear to the researcher.



Similarly, "outwit hub" freeware allows data extraction from RSS feeds and exporting them to Excel for research. It can also extract data from the webpage, harvest the tables inside, etc. It also generates the headline, a brief abstract of the news, and the news URLs. Along similar lines, images from www.images.google.com or www.flickr.com can be harvested by opening the image search URL in the "outwit image hub".

The screenshot shows the OutWit Hub Light application window. On the left is a sidebar with categories like page, links, documents, images, contacts, data, lists, guess, scraped, tot, words, news, source, automators, queries, scrapers, misc, jobs, and history. The main area shows a list of news items from 'http://rss.nytimes.com/services/xml/rss/nyt/Trump.xml' with columns for Id, Source Url, Feed Title, Title, and Article Url. The right side has tabs for Excel, Custom, Export, and Edit, with an Excel table view showing the same data.

Source Url	Feed Title	Title
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	U.K. Investigating Missed Signals Over Man...	U.K. Investigating Missed Signals Over Man...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	Francesco Totti Leaves the Field, and Romans...	Francesco Totti Leaves the Field, and Romans...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	Foot Soldiers in a Shadowy Battle Between R...	Foot Soldiers in a Shadowy Battle Between R...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	The Future of European Transit: Unraveling an...	The Future of European Transit: Unraveling an...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	News Analysis: Melania Trump, in Supporting...	News Analysis: Melania Trump, in Supporting...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	The Coat of Arms Said 'Integrity: Now It Says'...	The Coat of Arms Said 'Integrity: Now It Says'...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	U.K. Investigating Missed Signals Over Man...	U.K. Investigating Missed Signals Over Man...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	Francesco Totti Leaves the Field, and Romans...	Francesco Totti Leaves the Field, and Romans...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	Foot Soldiers in a Shadowy Battle Between R...	Foot Soldiers in a Shadowy Battle Between R...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	The Future of European Transit: Unraveling an...	The Future of European Transit: Unraveling an...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	News Analysis: Melania Trump, in Supporting...	News Analysis: Melania Trump, in Supporting...
http://rss.nytimes.com/services/xml/rss/nyt/Euro...	The Coat of Arms Said 'Integrity: Now It Says'...	The Coat of Arms Said 'Integrity: Now It Says'...

The RSS feeds are the easiest way to retrieve data from news websites. One need not chase the content; the reverse is true. The new sites have RSS feeds in XML format in the form of hyperlinks, which are read by the feed readers like Feedly. They are stored and refreshed at desired regular intervals to acquaint with the latest news on that feed's topic. However, harvesting data from the feeds is still an issue. We must apply manual methods for small research studies and proprietary software for the bigger ones. Opening RSS feeds in MS- Excel is easy and effective data collection. By activating the

"developer" tab in the ribbon, clicking "source", pressing "XML maps", pressing "add," and entering the feed URL in the "file name" field and opening the feed by going to the "data" tab in the ribbon and pressing "refresh all" button would refresh the data from RSS feeds whenever the news site updates it.

Metasearch engines –These engines search multiple search engines in one go and remove duplicate results while collating the result. www.searx.com is a very good free meta-search engine.



which can search the web and segregate information into images, music files, news, science stuff, social media, and videos. It has a built-in time filter, too. The results are downloadable in CSV file, RSS, and Jason formats. It has some built-in statistical functions, too, which calculate the relevance of the news stories. However, its use beyond its limit restricts your search and data download.

Regarding audio-visual media, every minute, 400 hrs of videos are uploaded on YouTube, adding to the already billions, but by applying the filters just under the address bar (HD, duration, time, type of video, view count, etc.) while searching, the results are reduced to a few hundred. This gives representative data to the researcher, which can be downloaded using "YouTube downloaders" for audio-visual analysis.

<https://ytdt.digitalmethods.net/> can harvest the video details and convert them into Excel format by selecting the desired variables, such as title, video description, comment text, date, likes, replies, duration, view count, top-rated comments, etc.

The screenshot shows a web browser window with the URL https://tools.digitalmethods.net/netvizz/youtube/mod_video_info.php?videohash=cLqjK3ddSy0%26t%3E. The page title is "YTDT Video Info and Comments". At the top, there are navigation links: Home, Channel Info, Channel Network, Video List, Video Network, Video Info, and FAQ. Below the title, a section titled "YTDT Video Info and Comments" contains descriptive text about the module's purpose and outputs. It lists four types of outputs: tabular files for video info, comments, authors, and interactions. A note says the first three can be shown directly in the browser by enabling HTML output. Below this, there are input fields: "video id:" with the value "cLqjK3ddSy0&t=532s" (with a note that video IDs can be found in URLs), and "HTML output" with a checked checkbox (with a note that it adds HTML result tables). A "Submit" button is at the bottom.

Apart from collecting comments and usual video parameters, the freeware downloads the pdf file, which maps user interactions in the comments section. This makes it a nice tool for video social network analysis. The comments thus collected can be harvested by using the "data" tab on the ribbon and pressing the "from web" button. A text box would prompt where you can put the webpage URL with the comments. After pressing go, the webpage would be imported into Excel in an editable format, which is amenable to data analysis. These data attributes can be stored in Excel or SPSS to analyse the categorical data.

Google SpreadSheets, too, is a very important tool for scraping web data from websites and social media. However, it is a slightly technical process, and there are very useful videos available on YouTube to help you understand how Google SpreadSheets work for harvesting website data.

It should be kept in mind that all major data harvesting tools are largely paid for with proper planning; freeware, too, can serve the purpose. Additionally, there are very good readymade collections of data on every topic collected and hosted by researchers for non-commercial and research purposes. GPS commands can be accessed and downloaded quickly.

Data harvesting from social media like WhatsApp, Facebook, and Twitter, a separate Unit is devoted to this topic, where the concepts of API (Application Programming Interface), access token, secret, consumer key, etc., have been discussed at length.

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. Create a keyword string for a research topic of your choice and use the appropriate online tool to extract the relevant data. Share your experience.

2. Lists the pros and cons of using digital resources for academic purposes.
-
-
-
-

15.3 DIGITAL DATA ANALYSIS TECHNIQUES

In this Unit, you will learn how to apply data analysis skills to the text/data under the Digital Media Research Methods (henceforth DMRMs). This understanding of data behaviour would acquaint you with the various technological intricacies of analysis. The information harvested through digital media platforms needs different analytical skills. For many researchers, the DMRMs is a completely technological process with very little manual data handling and analysis. For others, it may be a healthy combination of hassle-free online data collection and subsequent analysis in a conventional research protocol. Both approaches have pros and cons depending upon the nature and duration of the research enquiry. However, with technological advancement, the argument tends to favour maximum research automation, especially for content analysis studies.

You will learn a few new kinds of digital data analysis while simultaneously applying the DMRMs. These software-assisted analytical skills are new to the social sciences and media research. This chapter will help you make sense of the data by applying state-of-the-art data analytical techniques like corpus and sentiment analysis. For systematic and easy learning, this DMRMs section has been divided into three sub-parts, which cover the whole gamut of the topics, from the nature of the data to interpretational nuances that run deep into the data.

In the last section, you learn the powerful commands to search Google for your query with utmost precision and focus, thus weeding out unwanted and irrelevant results. The results are high in quality and very low in quantity. Thus, the data is easily manageable even with the manual method of downloading (copy and paste). Various tools and Apps were discussed to harvest this data in the format you need and, if not, convert it with a converter. Along with the text harvesting, you learn to download PowerPoint presentations (PPTs), doc files, PDF files, and images with a single click using freeware. RSS feeds were discussed as the push method rather than the pull method, along with a constant and less time-consuming method of refreshing the data stream. This chapter discusses the analysis of the data collected.

Methods to Analyse the Digital Data

It is assumed that the data harvested conforms to the format and parameters of the research objectives. It would be an injustice to discount the relevance of traditional research methods even on the electronically collected data. The DMRMs can supplement, not replace, the traditional methods. Here, two scenarios emerge. First, technology takes care of the data collection, and the researchers conduct further analysis and interpretation per established traditional norms and conventions. e. taking printouts of the collected news, entering them in the code sheet manually, and finally analysing them using statistical analysis software.

Second, the more technologically advanced option, the huge data collection (in text form), is subjected to "computational analysis" like corpus linguistics and textual analysis. The rationale is that this data, being huge, cannot be processed and analysed by the human mind. Both are perfect as per the scope of enquiry. The data collected is of two types – Quantitative (numerical) and Qualitative (text).

Quantitative Data Analysis—This data can be of nominal type, i.e. naming the variable (**Gender** – 1. Male 2. Female), ordinal i.e. in order (**Happiness level after Meditation** – 1. Very Bad 2. Bad 3. Good. 4. Very good 5. Excellent) and interval (**Age** – 10-15, 15-20, 20-25, 25-30). It simply means that the output can be quantified in numbers. Usually, data feeding in Excel or SPSS for analysis must follow a certain method, which is sometimes a challenge for researchers. However, the harvested datasets from open databases are mostly good, and descriptive statistics (frequency, mean, median, mode, standard deviation, etc.) can be generated in MS Excel. For this, open Excel, click the file, go to options, click add-ins, then Excel add-ins, add 'analysis toolpak' and click OK. In the Data tab of the menu, a new option 'data analysis' would appear on the right side. Now feed the numerical data in column click the "data analysis". A box would appear and select the "descriptive statistics" option. Now, define the input range by selecting the data and the output range. Click on the summary statistics, and OK. You will get the descriptive stat in the new column, as shown in column 1 of the image.

A	B	C	D
12	<i>Column1</i>		
12			
13	Mean	19.4375	
25	Standard Error	1.231424	
16	Median	21.5	
14	Mode	12	
18	Standard Deviation	4.925698	
16	Sample Variance	24.2625	
24	Kurtosis	-1.49362	
26	Skewness	-0.35829	
23	Range	14	
24	Minimum	12	
21	Maximum	26	
22	Sum	311	
22	Count	16	
23	Confidence Level(95.0%)	2.624719	

However, Excel's central tendency tests (mean, median, mode) can be generated only for the numeric (parametric) data. SPSS is required to analyse the categorical or nominal data (Non-parametric). For easier parametric data analysis, www.socscistatistics.com/tests/ can be accessed, and data can be fed, and output taken easily.

The present section will focus on qualitative data (collection of huge text) and its analysis through a corpus linguistics approach. A corpus is a "body of texts, utterances, or other specimens considered more or less representative of a language, and usually rendered in a machine-readable format" (Peng, H., Cambria, E., & Hussain, A. 2017). Earlier, we were taught to apply data analysis techniques only on nominal and numerical data, whereas the text obtained through open-ended questions was expected to be explained in a qualitative manner of essay writing. The DMRMs challenge this notion, and statistics apply to the text too. The corpus approach to the social sciences (CASS) is an innovative DMRM approach based on the corpus method. Huge data files in .txt format can be opened using a few free corpus analysis software (e.g. AntConc). The HTML PPT or Doc format data can be converted into .txt format and saved as UTF-8 encoding in Notepad for corpus analysis.

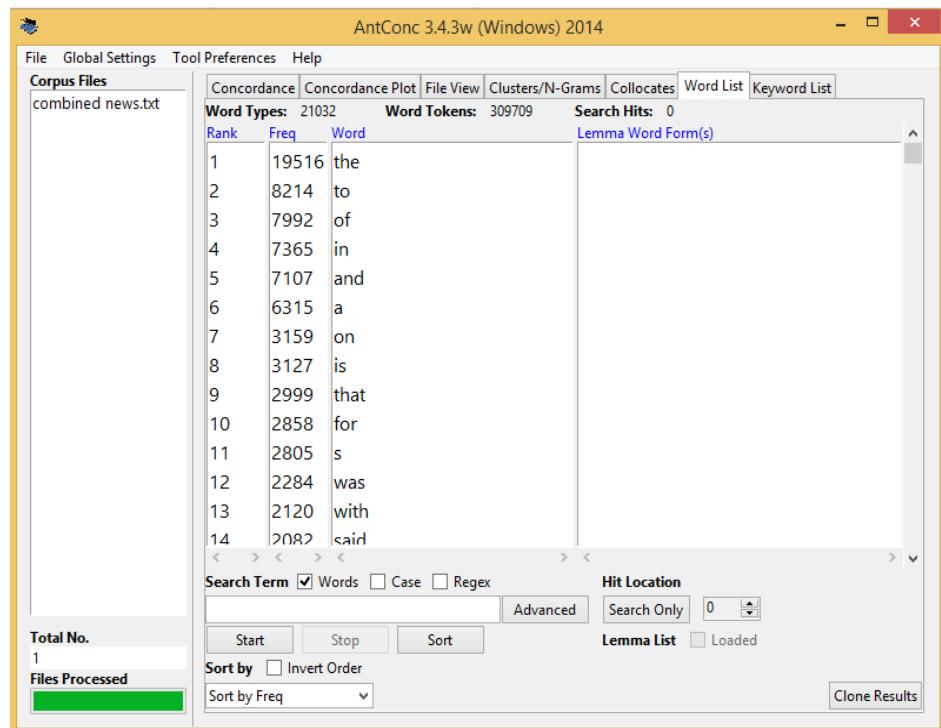
Before applying textual statistical analysis, it is very important to know the "nature of the data " in conventional research; likewise, in DMRMs, the basic characteristics of the text need to be obtained. Text analysis tools online (www.textalyzer.net) can analyse basic traits of the text, i.e. total word count (tokens), unique words (types), lexical density (textual diversity), sentence count, average sentence length and readability factor (easy, medium or hard). It can analyse text on a website and upload documents and the text on its website, as shown in the image.

Total word count :	467
Number of different words :	327
Complexity factor (Lexical Density) :	70%
Readability (Gunning-Fog Index) : (6-easy 20-hard)	10
Total number of characters :	4779
Number of characters without spaces :	3171
Average Syllables per Word :	1.91
Sentence count :	55
Average sentence length (words) :	14
Max sentence length (words) :	104
(in a recent corpus study of us presidential elections speeches of hillary clinton and donald trump it was found that trump's keywords include going very hillary i they are it i'm don't great china really billion nobody mexico immigration disaster islamic nice mean folks me somebody people ok tremendous problem thought totally radical anybody got amazing border rigged foreign money worst literally whereas hillary's keywords include women families economy together work american and america future rights create men growth equal fair deserve global challenges threats opportunity israel help stronger students income invest raising everyone god children abortion responsibility and vigorous)	
Min sentence length (words) :	1
(textalyzer)	
Readability (Alternative) beta : (100-easy 20-hard, optimal 60-70)	30.8

The Gunning Fog Index (also at www.gunning-fog-index.com) analyses the text's readability as 6-easy and 20-hard. Lexical density or complexity factor analyses the text's "richness" (i.e., more repeated words, less richness). The word's frequency can also be counted.

A free software, AntConc (www.laurenceanthony.net/software/antconc/), can be downloaded for more advanced corpus analysis.

Word List - The most basic corpus analysis of text involves the generation of "word lists", i.e., the total number of words (tokens) present in the text and the number of unique words (types). The type-to-token ratio gives a score of lexical density (i.e. richness of the language in text), e.g. if a text has a total of 120456 words and 15675 unique words, then the type-to-token ratio is 0.13 (13%).



A low score indicates non-informative text and vice versa. A token-to-type ratio indicates the "vocabulary density score", which in our case is 8th. This also indicates the diversity in the text by indicating that "at every 8th word in the text, a new word appears". However, the interpretation of textual scores can be highly subjective at times. The words can be clicked to see the context in which they are used in the text.

Collocations - The text can be further analysed deeply by using collocations. Sinclair (1991) defines collocation as the tendency for words to occur together and be identifiable by frequency of occurrence, e.g. "media habits", "triple talaq", and "social media". The software calculates the collocates of the "node word" (i.e. central word as defined by you; here, the node word is "media"). The figures show that the tendency of co-occurrence of the word "social" with "media" is statistically significant, with a mutual information ratio score of 9.7, which is high. Any score above 3 is significant. The negative score indicates that the two words shun each other. Collocation is calculated for the node word within a span of 5 words to the left and five to the right.

This word association forms the meaning in the readers' mind, and the nature of one word affects the other. A random view of your association of words at a mental level may tell you the significance of the "meanings" that you attribute to the objects in this world.

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits	32					
Hit	KWIC					
1		Get Net Neutrality Debate Right in India: Mark Zuckerberg NEW YORK: Social media giant Facebook has said it's important to get the				
2		of talented music director and have expressing their feeling on the social media. Over the course of his career, Shrivastava had composed music				
3		Asha, who has more than 1.89 million followers on her Twitter account. *Social media is here to stay and it's going to get				
4		's evident by the over 4.3 million followers that he has on social media. While in New Zealand, he is also upbeat about				
5		showing it all to the world by posting it on the social media has resulted in horrific deaths worldwide. Recently a 19-year				
6		and scholars to counter these forces through various avenues including on social media, so that the ideology of extremism does not take root				
7		US marshals takes motherhood to a different level altogether. Viral on social media, this picture clicked by an ex-servicewoman and a mother,				
8		damaging the typewriter of an elderly typist went viral on the social media. Krishna Kumar (65) had been sitting on the pavement outside				
9		smashed. As the pictures of the incident went viral on the social media, the CM asked senior officers to suspend the errant cop				
10		here traced the boy's family in the neighbouring country via social media. "Mohammad Ramzan got separated from his mother Begum Razia				
11		and narrated his ordeal across the border with the help of social media," he said. "What couldn't be done in two years,				
12		, as a mark of protest, post the pictures or videos on social media that may spark communal tension. The state government's move				

Concordances - Qualitative studies tend to gain a lot because of DMRMs. The latter has made various new approaches to text reading possible. The concordance approach allows the meaning derivation from contexts. Concordance is the display of a given word "together with a given amount of preceding and following context". It is, thus, contextualising the text.

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	
Total No. of Collocate Types: 656			Total No. of Collocate Tokens:		
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	94	56	38	3.41349	the
2	39	16	23	3.60165	and
3	38	24	14	3.39486	of
4	37	25	12	3.31685	to
5	33	17	16	3.30920	in
6	32	32	0	9.71895	social
7	23	5	18	4.08457	that
8	23	16	7	4.00958	on
9	23	17	6	3.01027	a

The word list function is not very useful unless the context is known. Even negative words can often be used positively, and vice versa, e.g., Let us save the world from **the crisis**. **Peace** talks derailed. As explained in the chapter, a detailed sentiment analysis can be conducted for these statements. The image shows the context of two collocates (social and media) in the corpus text. By describing the text, we can find the context in which they have been used. Hence, a whole range of discussion around two compatible words can be found and explained.

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. Enumerate the benefits of digital data analysis techniques.

.....

.....

.....

.....

2. Explain the utilisation of digital data analysis techniques.
-
-
-
-

Word Clusters -

DMRMs, through the corpus approach, also enable the generation of word clusters of one or more words. The clusters present in the text stress the "word sense" of the text. Word clusters are different from collocates as clusters are fixed words appearing together, whereas collocates may or may not be fixed (span of 5 words on either side), and their probability is statistically calculated.

AntConc 3.4.3w (M)			
File Preferences Help		Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List	Total No. of Cluster Types 93 Total No. of Cluster Tokens 140
Rank	Freq	Range	Cluster
1	10	1	media and
2	7	1	media reports
3	6	1	media reported
4	5	1	media baron
5	5	1	media report
6	4	1	media for
7	3	1	media events

The word cluster "media and", which has a frequency of 10, as shown in the image, can be clicked, and the context in which it appears can be read (concordance). A researcher must enter a word in the software to find the clusters.

AntConc 3.4.3w (M)			
File Preferences Help		Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List	Total No. of Cluster Types 93 Total No. of Cluster Tokens 140
Rank	Freq	Keyness	Keyword
1	175	515.003	polic
2	156	360.501	indrani
3	151	338.426	sheena
4	552	312.185	her
5	74	171.007	mukerjea
6	110	159.821	murder
7	163	152.758	accused
8	285	132.910	case
9	61	127.033	peter
10	60	124.724	bora
11	444	109.115	she
12	65	96.190	khanna
13	75	93.509	saudi
14	2284	89.730	was
15	37	85.503	raicad

N-grams—It is like word clusters, with the only difference being that N-grams are software-identifiable groups of words, and we cannot specify our own words to find N-grams.

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Total No. of N-Gram Types		154084	Total No. of N-Gram Tokens		309708		
Freq	Prob	N-gram					
365	0.241	from the					
340	0.269	will be					
337	0.194	has been					
337	0.934	prime minister					
320	0.158	it is					
301	0.145	said the					
296	0.095	is a					
274	0.703	per cent					
260	0.129	it was					
253	0.169	have been					
250	0.135	as a					
247	0.079	is the					
241	0.012	the first					

As explained above, this context can be subjected to different analysis conventions, such as concordance.

Keywords - Another important milestone is the method of keyword comparison between two sets of texts. DMRMs in this function allows us to statistically compare the unusually high frequency of certain words in one text and vice versa. The image shows the comparison of two corpora. The text on political news was compared to texts on crime news.

Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Types Before Cut:		21032	Types After Cut:		21032	Search Hits:	0
Rank	Freq	Keyness	Keyword				
1	675	111.493	modi				
2	1262	70.361	will				
3	3127	58.847	is				
4	1205	55.657	india				
5	1395	51.604	i				
6	408	43.950	bjp				
7	203	36.576	narendra				
8	709	34.105	minister				
9	256	32.509	bihar				
10	2022	32.394	it				
11	533	32.053	pakistan				
12	207	29.530	meeting				
13	154	27.748	kashmir				
14	633	27.515	you				
15	274	27.110	cent				

It shows that words like Modi, India, BJP, minister, Bihar, Pakistan, Kashmir, and meeting (all relevant words) are statistically (see the keyness value) more frequent in the political corpus than the crime corpus. When the crime corpus is compared to the political corpus, words like murder, Indrani, Sheena, accused, case, etc., were statistically more frequent in the crime corpus than in the political corpus.

In a recent corpus study of US Presidential elections speeches of Hillary Clinton and Donald Trump, it was found that Trump's keywords include going, very, Hillary, I, they, are, it, I'm, don't, great, China, really, billion, nobody, Mexico, immigration, disaster, Islamic, nice, mean, folks, me, somebody, people, OK, tremendous, problem, thought, totally, radical, anybody, got, amazing, border, rigged, foreign, money, worst, literally whereas Hillary's keywords include women, families, economy, together, work, American (and America), future, rights, create, men, growth, equal, fair, deserve, global, challenges, threats, opportunity, Israeli, help, stronger,

students, income, invest, raising, everyone, God, children, abortion, responsibility, and vigorous. (Kristin, 2016)

In addition to the above functions, a corpus can be analysed using the "stop word list" widely available on the Internet. This list filters out unimportant words like am, is, are, me, you, has, have, this, that, etc. Hence, the core words present a more coherent analysis.

Corpus can be searched using the specific words provided by the researchers in a vast amount of text. The word list function would throw only the results having those words, and this way, the research process can stay focused. Since dealing with the text is a fun exercise, a few researchers have done corpus analysis by analysing the text for good and bad words or, in other words, positive and negative words. A corpus study of tweets on Jat reservation agitation in Haryana revealed that negative words outnumber positive words by a huge margin.

However, the corpus approach has a few limitations. "The improper and skewed representation of the target language fails to serve properly or fails to meet the general requirements of the language investigators" (Landau 2001: 321). Language is an internal construct and hence cannot be regarded as the sole evidence in determining something.

Word Cloud—A graphic representation of the text's prominent words (in number) can be generated at www.tagcrowd.com and www.wordle.com by copying and pasting the text into the box.



Sentiment Analysis – "Sentiment analysis (sometimes known as opinion mining or emotion Artificial Intelligence) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information" (Wikipedia). "sentiment analysis aims to determine the attitude of a speaker, writer, or other subject concerning some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event" (ibid). It is done on a text (corpus) or individual statements collection. The exponential rise of social media (Facebook, Twitter, WhatsApp etc.) and growing access to users' feedback (YouTube comments on videos, blogs, and readers' comments on the news of online newspapers) have given rise to sentiment analysis. The criterion to which the text is subjected is polarity (positive, neutral, negative), subjectivity (people's opinions, views, discussions etc.) and objectivity (facts and figures, unbiased information). The degree and strength of polarity to matter as a sentence or paragraph may be mildly positive or negative. This is a relatively new branch

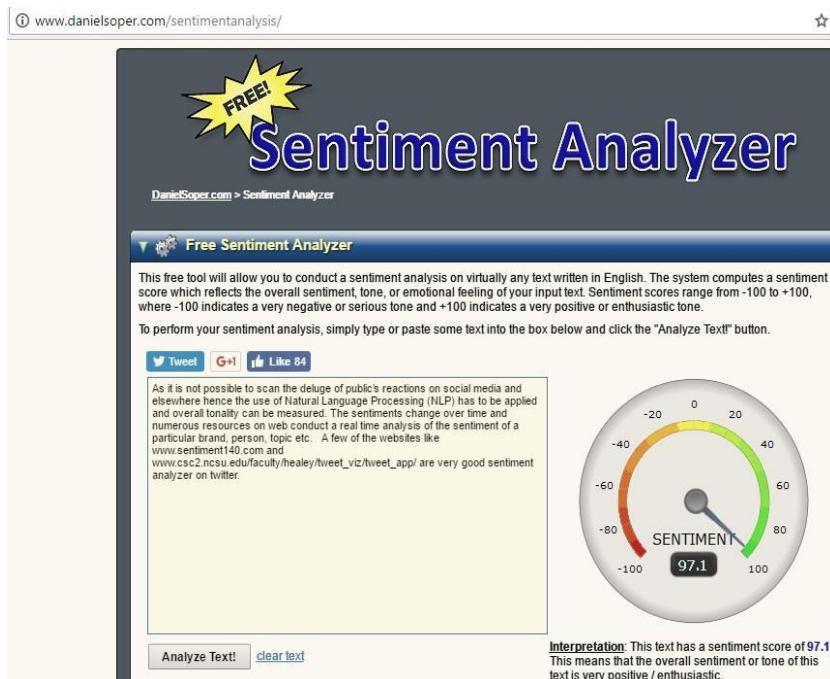
of DMRMs and is gaining popularity in social sciences. The most convenient way of analysing the statements (e.g. headlines, leaders' quotes, the gist of parliament speeches) is to use Google spreadsheets and install an extension known as "Text Analysis" by clicking add-ons. The statements to be analysed for sentiment are pasted in the spreadsheet as individual sentences, selected, and analysed by the App.

Statement	Tonality	Score	Polarity	Score
Six months of demonetisation: How Modi failed to keep the promises ...	objective	0.99	negative	0.83
Demonetisation and the Delusion of GDP Growth	objective	0.52	neutral	0.61
demonetisation: Every department should disclose demonetisation...	subjective	1.00	neutral	0.51
Buy Demonetisation and Black Money Book Online at Low Prices in ...	objective	0.54	neutral	0.67
Demonetisation - Latest News, Analysis, Opinion -	subjective	1.00	neutral	0.85
Demonetisation: Why no window to deposit banned notes after ...	objective	0.91	negative	0.70
Has demonetisation boosted income tax collections? -	subjective	1.00	neutral	0.82
Demonetisation News - Demonetisation Latest news on www.oneindia ...	subjective	0.86	neutral	0.80
India Has Added 91 Lakh Taxpayers Since Demonetisation: Arun Jaitley	subjective	0.98	neutral	0.69
Decipher the Demonetisation	subjective	0.84	neutral	0.50
Demonetisation: After SC criticism, the Centre will decide on the deadline.	objective	0.87	neutral	0.85
demonetisation: Latest News, Articles on demonetisation Business ...	objective	0.51	neutral	0.83
Demonetisation effect: The government collects Rs 6000 crore in tax on unexplained ...	objective	1.00	neutral	0.75
Demonetisation to usher in tax terrorism? -	subjective	0.87	neutral	0.83
Demonetisation: Critics Silenced as India moves ahead - Narendra Modi	objective	0.76	neutral	0.61
Macroeconomic Impact of Demonetisation- A Preliminary Assessment ...	objective	1.00	neutral	0.55
Infrastructure Today Magazine Impact of demonetisation on ...	objective	1.00	neutral	0.76
The Reserve Bank on Demonetisation	objective	0.98	neutral	0.62
Demonetisation: reality and the optics	subjective	1.00	neutral	0.42

Demonetisation will hurt the BJP very badly: Rahul Gandhi - Inc. in	subjective	0.99	negative	0.48
Social impact of demonetisation may have been greater: World Bank ...	objective	0.70	neutral	0.54
Luxury realty market takes hit post demonetisation -	subjective	0.56	neutral	0.72

The Table presents a few harvested headlines on "demonetisation" from Indian newspapers by applying the Google command – demonetisation site: in. You can analyse if the sentiment analysis done by the software makes sense. The score indicates confidence in determining the suitability of the value assigned to that statement. A score of 1 indicates 100 per cent confidence.

Sentiment analysis of a complete paragraph can be done by www.danielsoper.com/sentimentanalysis and its interpretation is very simple.



As it is impossible to scan the deluge of the public's reactions on social media and elsewhere, Natural Language Processing (NLP) must be applied, and overall tonality can be measured. Sentiments change over time, and numerous resources on the web conduct real-time analyses of the sentiment of a particular brand, person, topic, etc. Some websites, like www.sentiment140.com and www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/, are very good sentiment analysis on Twitter.

<https://www.lexalytics.com/semantria/excel> is also a powerful tool for analysing sentiment, assigning categories, using word clouds, etc., to the text. It's free for first analysing the first ten thousand statements. The sentiment of a single word (person, brand, issue, etc.) can be nicely analysed by <http://www.opinioncrawl.com/>

It must be noted that text analytics is a relatively new area in the computational aspect. Hence, the tools may or may not perform as per your

expectation of accuracy, but by and large, they give you a fair idea of the "behaviour" of the text.

15.4 LET US SUM UP

You have learnt the broad concept of Digital Media Research Methods, emphasising analysing the data threadbare for qualitative and quantitative attributes. Only DMRMs can make this dual treatment possible for the data, whereas other conventional techniques apply to either. The lesson empowers you to appreciate this innovation and challenge the strongly held opinions about the nature of data analysis. The researchers trained in qualitative research can expect their "thorough and thoughtful control" over the data.

15.5 FURTHER READINGS

1. Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J., Ogata, H., ... & Tsai, C. C. (2020). Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology*, 11, 580820.
2. Levy, R. E. (2021). Social media, news consumption, and polarisation: Evidence from a field experiment. *American Economic Review*, 111(3), 831-870.
3. Naeem, M., Jamal, T., Diaz-Martinez, J., Butt, S. A., Montesano, N., Tariq, M. I., ... & De-La-Hoz-Valdiris, E. (2022). Trends and future perspective challenges in big data. In *Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, 15–18 October 2019, Arad, Romania* (pp. 309-325). Springer Singapore.
4. Mahrt, M., & Scharkow, M. (2013). The value of big data in digital media research. *Journal of Broadcasting & Electronic Media*, 57(1), 20-33.
5. Penzel, T., Hirshkowitz, M., Harsh, J., Chervin, R. D., Butkov 5, N., Kryger, M., ... & Chesson, A. L. (2007). Digital analysis and technical specifications. *Journal of Clinical Sleep Medicine*, 3(02), 109-120.
6. Twenge, J. M., & Martin, G. N. (2020). Gender differences in associations between digital media use and psychological well-being: Evidence from three large datasets. *Journal of Adolescence*, 79, 91-102.

15.6 CHECK YOUR PROGRESS: POSSIBLE ANSWER

Check Your Progress: 1

1. Selecting the right keywords is pivotal for effective research. For instance, for a study on "Digital Inequalities in Online Education," relevant keywords could include "digital divide," "online learning," "accessibility," and "educational disparities." Using tools like Google Scholar or databases specific to the field can yield valuable data. Google Scholar used to extract scholarly articles, ensuring credibility and depth in the research. The process involved refining search terms iteratively to capture diverse perspectives and

current findings, enriching the investigation.

2. Pros and cons of employing digital resources in academia require careful consideration. On the positive side, digital resources offer unparalleled access to vast information, facilitate collaborative learning through online platforms, and promote flexibility in study schedules. However, challenges such as information overload, credibility issues with online sources, and distractions like social media can impede academic focus. Additionally, reliance on digital tools may exacerbate digital inequalities, disadvantaging students with limited access to technology or reliable internet connections. Balancing these factors is crucial for maximising the benefits of digital resources while mitigating their drawbacks in academic pursuits.

Check Your Progress: 2

1. Digital data analysis techniques offer a multitude of advantages. Firstly, they facilitate enhanced decision-making through data-driven insights, enabling organizations to make informed choices based on empirical evidence rather than intuition alone. Secondly, these techniques allow for the identification of trends and patterns within datasets, empowering businesses, and institutions to develop strategies that capitalize on emerging opportunities or mitigate potential risks. Additionally, digital data analysis enhances operational efficiency by streamlining processes and identifying areas for improvement, leading to cost reduction and resource optimization. Lastly, these techniques enable targeted marketing efforts and personalized customer experiences by leveraging data to understand consumer behaviour and preferences, ultimately driving customer satisfaction and loyalty.

2. The utilization of digital data analysis techniques spans across various industries and sectors, each benefiting from the insights derived from data analysis. In the realm of business, organizations employ these techniques for market research, customer segmentation, and predictive analytics to gain a competitive edge and drive growth. Similarly, in healthcare, digital data analysis supports patient monitoring, disease prediction, and drug development, ultimately improving patient outcomes and advancing medical research. Financial institutions utilize these techniques for risk assessment, fraud detection, and portfolio management, safeguarding assets and ensuring financial stability. In academia, researchers utilize digital data analysis for research data analysis, trend identification, and hypothesis testing, contributing to the advancement of knowledge and understanding in their respective fields. Moreover, in governance, policymakers leverage data analysis for policy formulation, public opinion analysis, and performance evaluation, facilitating evidence-based decision-making and effective governance practices.

UNIT 16 DATA INTERPRETATION

Structure

- 16.0 Introduction
- 16.1 Learning Outcomes
- 16.2 Introduction to Statistics
 - 16.2.1 Populations and Samples
 - 16.2.2 Scales of Measurement
 - 16.2.3 Discrete and Continuous Variables
- 16.3 Frequency Distribution
 - 16.3.1 Frequency Distribution Tables
 - 16.3.2 Continuous Variables and Real Limits
 - 16.3.3 Frequency Distribution Graphs
 - 16.3.4 Histograms and Bar Graphs
 - 16.3.5 Shape of a Frequency Distribution
- 16.4 Measures of Central Tendency
 - 16.4.1 Mean
 - 16.4.2 Median
 - 16.4.3 Mode
- 16.5 Variability
 - 16.5.1 Range
 - 16.5.2 Interquartile Range and Semi-Interquartile Range
 - 16.5.3 Standard Deviation and Variance for a Population
 - 16.5.4 Standard Deviation and Variance for Samples
- 16.6 Let Us Sum Up
- 16.7 Further Readings
- 16.8 Check Your Progress: Possible Answers

16.0 INTRODUCTION

Numbers are part of our daily life; we are used to coming across various sets of numbers, meaning something to us. For example, a mild earthquake measured 4 on the Richter scale hits a mountain region, but a severe 7-point Richter scale earthquake killed 100s of people. The difference between 4 and 7 numbers is just 3, but the difference in terms of magnitude - the former is a mild one, and the latter is a destructive force. So, numbers have meaning; when it comes to complex data sets, we need a scientific tool to classify and understand these data and make sense of them. The process of understanding data is data analysis, and the tool used to do this is statistics. In Communication Studies, statistical analysis is generally employed in quantitative settings, and it has a limited scope in qualitative research.

In this Unit, we shall discuss different tools of descriptive statistical analysis in detail, using exercises.

16.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Discuss the need for and importance of statistical analysis in communication research;
- Describe Populations, Samples and Scales of measurement;
- Explain Frequency Distribution;
- Define Variability;
- Discuss Standard Deviation and Variance for a population;
- Explain measures of Central Tendency such as mean, mode and median;
- Use the above statistical methods while analysing your data.

16.2 INTRODUCTION TO STATISTICS

Statistics is a branch of mathematical sciences which deals with numbers and data sets. A basic definition of statistics is "a process and tool involved in collection, analysis, interpretation, explanation and presentation of data." We can extract information from raw data through various statistical analysis tools and create a new understanding of numbers. Statistics help us to simplify complex data sets. Statistical tools help us to interpret and explain information. It helps us to make informed decisions and even helps us to predict a direction based on diverse data sets.

Statistics are broadly classified into two types: descriptive statistics and inferential statistics. Descriptive statistics, as the name suggests, are meant to classify, and summarise collected raw data. Major kinds of analysis under descriptive statistics are measures of frequency, measures of central tendency, and measures of dispersion. The second type, inferential statistics, helps us draw conclusions from associations or differences between variables.

As this Unit deals with the basic concepts of statistical applications, we shall focus on various analytical tools of descriptive statistics in detail through exercises.

16.2.1 Populations and Samples

Block 1, Unit 4 is devoted to sampling. You could go back to that Unit to refresh yourself on sampling. Here, just to recap, a population is the entire group of individuals that a researcher wishes to study. A sample is a subset of a population. It is drawn from out of the population as defined by the researcher according to the rules of sampling.

16.2.2 Scales of Measurement

Measurement is the heart and soul of research. One of the central characteristics of a quantitative study is the requirement to measure whatever you study: gender, occupation, height, weight, time, viewership, motivation, care, commitment, gratification, loyalty... etc. When measuring, we assign

numbers (1, 2, 3, etc) to variables and concepts under investigation. For example, suppose you have a variable such as gender with two categories (female and male). In that case, you will measure it quantitatively by assigning a unique number to each of the two gender categories (you may assign number 1 to the female category and 2 to the male category).

Numbers can also be placed into a scale based on their inherent characteristics, such as the kind of number. We will briefly look at numbers in terms of scale.

Nominal Scale

This is the lowest scale (or level) of measurement. At this level, each response category of a variable is assigned a unique number. For example, as discussed above, a variable such as gender with two response categories will assign a unique or distinct number to each category. At a nominal level, the category is used simply to describe.

No mathematical relationships are possible at the nominal scale of measurement. You cannot add, subtract, multiply or divide. You can only count.

Ordinal Scale

The word 'ordinal' is derived from the word 'order'. In an ordinal measurement scale, numbers indicate only the rank order of cases (respondents) on some variable.

For example, all students **who** pass an examination get an I, II or III class. In this case, you can say that a student who earned an I class fared better in the examination than a student who secured an II class; similarly, a student who passed in II class fared better than a student who passed in III class. That is, you can rank students according to their results. But you cannot say the distance (interval) between any two ranks.

Interval Scale

Interval measurement has the properties of the nominal and ordinal scales, plus the quality that equal intervals between numerals indicate equal distance in the measured variables. For example, if the temperature today is 20°C and on May 1 it was 40° , then you can say that today is 20°C lower than May 1. In other words, with an interval scale, you can determine precisely which observation is of greater magnitude and by how much it is greater. You can do so by subtracting one number from the other. This tells us something important about interval scales: measurements on an interval scale permit basic arithmetic operation such as subtraction and addition.

However, measurements on an interval scale do not allow you to multiply or divide. Let us explain this with the example of temperature: Assume the temperature is 20°C today and 40°C on a given day in summer. The only statement you can make in this case is that the temperature difference between the two given dates is 20°C . But you cannot say it is half as hot today as on a given day in summer. Why is it so? Because a ratio ($20^{\circ}\text{C} : 40^{\circ}\text{C}$) is meaningful only when the measurement scale has an absolute zero point

that indicates the total absence of the measured concept. For example, zero degrees on the Centigrade scale does not mean the absence of temperature. This point becomes clearer if you consider both temperatures measured on the Celsius and the Fahrenheit scales.

Interval scales consist of numbers that rank items such that numerically equal distances on the scale represent equal distances in the measured property. However, the 0 on the scale is arbitrary; 0 does not have a true or absolute value, and it does not indicate the absence of the property measured.

Ratio scale

The fourth level, called ratio measurement, retains the properties of the three preceding scales plus has the additional property of an absolute zero point. Because a ratio scale of measurement has an absolute zero point, ratios of numbers on this scale reflect ratios of magnitudes. For example, the variable income, measured in rupees, has this property. If one person's annual income is ten lakhs and another person's annual income is 20 lakhs, you can divide one with the other (that is, form a ratio) to state that one is twice (or half) as much as the other.

Ratio scales consist of numbers that rank items such that numerically equal distances on the scale represent equal distances in the measured property. Importantly, the 0 on the scale means the absence of the property being measured. Therefore, the 0 has real meaning.

16.2.3 Discrete and Continuous Variables

People who design research pay close attention to the classification of variables. Although one can classify variables differently, we will examine two types of variables here: discrete and continuous. Let us do so by looking at some questions first:

- How many siblings do you have?
- How many newspapers are published in your city? What is your marital status?

The answer to each of the three questions is a whole number. A discrete variable consists of separate, indivisible categories. It cannot be divided into fractional parts. You can have two or three siblings, but you can't have 2.5 siblings, for example.

A discrete variable has no intermediate values between two adjacent categories.

Now, examine another set of questions you generally read in a questionnaire. What is your monthly household income?

- What is your age?

Response to the two questions need not necessarily be whole numbers. Your income can be measured in rupees and paise, and your age can be measured in years, months, days, and even minutes. So, the two variables in this instance (monthly household income (MHI) and age) are not discrete. Variables such as MHI, age, time, and weight are not limited to a fixed set of separate, indivisible categories. These variables are called continuous because they can be divided into infinite fractional parts.

A continuous variable is divisible into an infinite number of fractional parts.

Check Your Progress: 1

Note: 1) Use the space below for your answer.
2) Compare your answers with those given at this Unit's end.
Indicate the level of measurement of the following variables:

1. Employment status: working full-time, part-time, as a student, housewife, or unemployed.

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

2. Religion: Christian, Muslim, Sikh, Jain, Hindu, other.

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

3. Gravity of crime: most serious, moderately serious, least serious.

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

4. Agreement or disagreement: Ask respondents to state their level of agreement on a scale of strongly agree, somewhat agree, neutral, somewhat disagree, or strongly disagree.

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

5. Number of items purchased: Shoppers in a mall indicate the number of items purchased there.

.....
.....

16.3 FREQUENCY DISTRIBUTION

Working with a small dataset makes it easy to make sense of it. However, it is difficult to make sense of the data or see any pattern or trend in a large dataset (let's say more than 100 respondents have answered 70 questions on a questionnaire). This is where frequency distribution helps. Simply put, a frequency distribution organises the data as a table or graph, making it easier to see a clear pattern or trend. One of the most common procedures for organising a data set is to place the scores in a frequency distribution.

A frequency distribution records the number of individuals located in each category on the scale of measurement. It allows the researcher to see "at a glance" the entire set of scores.

16.3.1 Frequency Distribution Tables

The simplest frequency distribution table presents the list of the individual scores in a column from highest to lowest. Besides each score, the frequency or number of times the scores occurred in the data is shown in another column. It is customary to use an X as the column heading for the scores and an f as the column heading for the frequencies.

Example: A class of 20 students' scores the following marks out of 10 ($N = 20$ scores).

We will organise these scores by constructing a frequency distribution table.

- First, identify the highest score in the set of scores. In this set, the highest score (X) is 10.
- Next, identify the lowest score in the set of scores. In the set, it is (X) 4.
- Write down the scores from the highest to the lowest (10 to 4) in a column, irrespective of whether a score appears in the set of scores (5 doesn't appear in the set of scores, but it is identified, as shown in the table). This is the first column in the table.
- In the fourth step, against each number, count and write down the frequency (f) with which it appears. For example, ten appears twice, nine appears seven times, and so on. This is the second column of the table.
- Finally, in the third column, multiply each X value by its frequency and then add these products. This gets the "X from a frequency distribution table. This sum may be expressed in symbols as " fX ". In addition, by adding up the frequencies, you will obtain the total number of individuals: " $f = N$ " (20 students or scores in the example below)

Score	Frequency Table Frequency	Sum
X	F	fX
10	2	20
9	7	63
8	5	40
7	2	14

6	3	18
5	0	0
4	1	4
	20	$jX=159$

16.3.2 Continuous Variables and Real Limits

The concept of a continuous variable and real limits are useful in measurement. Let us, therefore, try to understand them with a simple example. How do you measure time and the number of subjects you study in any degree programme? If you have sophisticated instruments, you could measure time in years or even break it down into months, days, hours, minutes, seconds, or even nanoseconds. But you cannot do the same thing with the number of subjects you study in your degree programme. In the above example, time is a continuous variable. A continuous variable has an infinite number of possible values. It has no gaps in its scale (that is why it is continuous). Therefore, it may be represented by a continuous number line containing infinite points (see below).

Let's take another example. When you measure body weight to the nearest kilogram, weights of 50.3 and 49.6 kgs would both be rounded to scores of $X = 50$. Note that a score of $X = 50$ is an approximate number; it is not a single point on the number line. The actual body weight is somewhere between 49.5 kgs and 50.5 kgs. Any measurement within this interval will be assigned a value of $X = 50$. The boundaries that form the interval are called the real limits of the interval corresponding to $X = 50$.

X	F
X	Y
12 - 13	4
10 - 11	5
8-9	3
6-7	3
4-5	2



Real Limits

Real limits are the boundaries of intervals for scores that are represented on a continuous number line. The real limits separating two adjacent scores are

located exactly between the scores. Each score has two real limits. The upper real limit is at the top, and the lower real limit is at the bottom of the interval. Real limits also apply to the class intervals of a grouped frequency distribution table. For example, a class interval of 40 - 49 contains scores from $X = 40$ to $X = 49$. These values are called the apparent limits of the interval because they appear to form the upper and lower boundaries for the class interval. But $X = 40$ is an interval from 39.5 to 40.5. Similarly, $X = 49$ is an interval from 48.5 to 49.5. Therefore, the real limits of the interval are 39.5 (the lower real limit) and 49.5 (the upper real limit).

16.3.3 Frequency Distribution Graphs

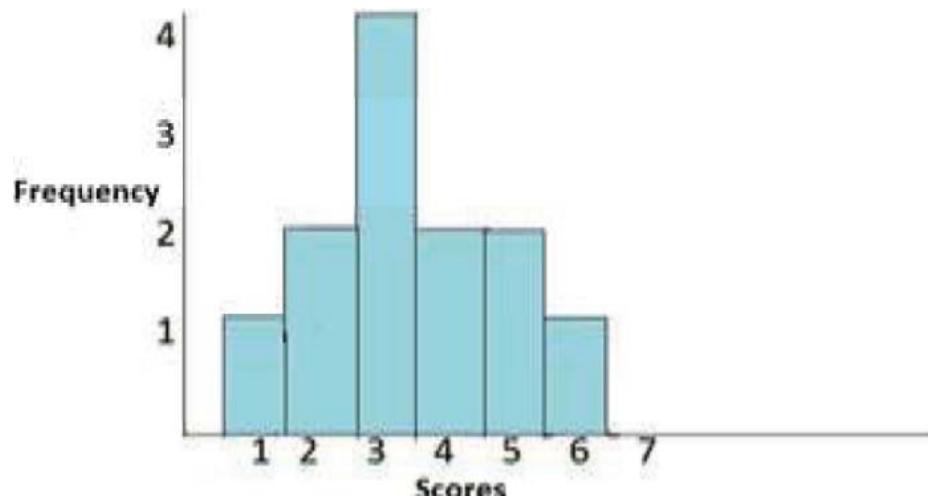
A frequency distribution graph is a pictorial representation of the data shown in a frequency distribution table. All the graphs start with perpendicular lines called axes. The horizontal line is the X-axis, and the vertical line is the Y-axis. The scores are listed along the X-axis, increasing value from left to right. The frequencies are listed on the Y-axis in increasing value from bottom to top. Generally, the point where the two axes intersect should have a zero value for both the scores and the frequencies. A final general rule is that the graph should be constructed so that its height (Y-axis) is approximately three-quarters of its length (X-axis).

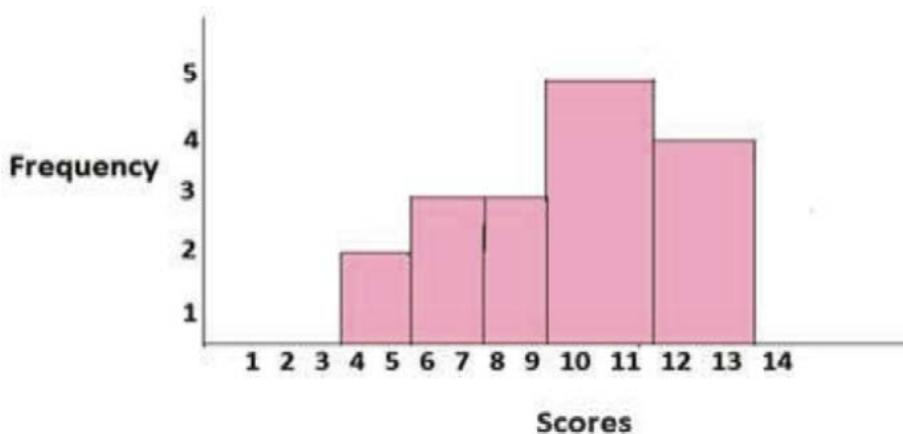
16.3.4 Histograms and Bar Graphs

For this type of graph, you simply draw a bar above each score so that the bar's height corresponds to the score's frequency. The measurement scale determines the choice between using a histogram or a bar graph. When a frequency distribution graph shows data from an interval or ratio scale, the bars are drawn so that adjacent bar touches each other. The touching bars produce a continuous figure, emphasising the variable's continuity. This type of frequency distribution graph is called a histogram. For a histogram, vertical bars are drawn above each score so that:

1. The height of the bar corresponds to the frequency.
2. The width of the bar extends to the real limits of the score.
3. E.g., Frequency Distribution Histogram.

For example, a frequency distribution histogram is used for grouped data.



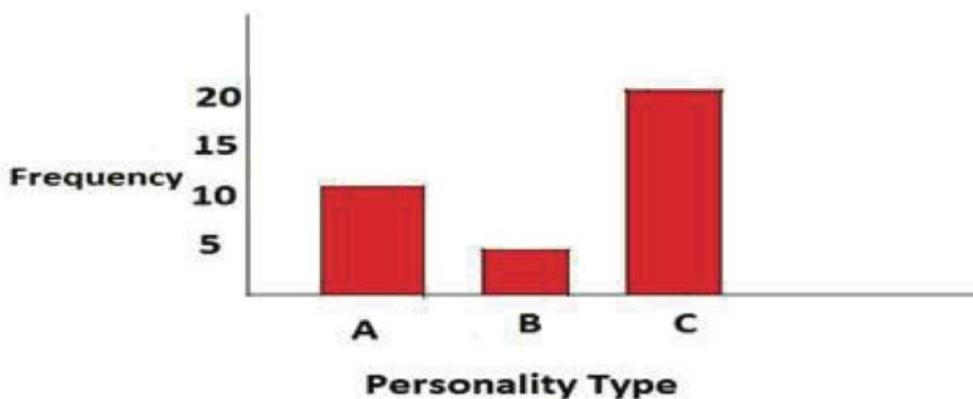


X	F
6	1
5	2
4	2
3	4
2	2
1	1

When presenting the frequency distribution for data from a nominal or ordinal scale, the graph is constructed so there is some space between the bars. In this case, the bars emphasise that the scale consists of separate, distinct categories. The resulting graph is called a bar graph. For a bar graph, a vertical bar is drawn above each score (or category) so that.

1. The height of the bar corresponds to the frequency
2. There is a space separating each bar from the text.

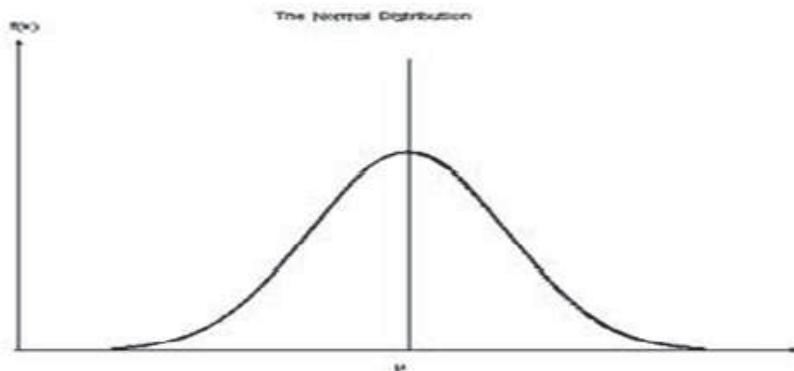
For example, a bar graph shows the distribution of personality types in a sample of college students. The graph is drawn with space between the bars because personality type is a discrete variable measured on a nominal scale.



16.3.5 Shape of a Frequency Distribution

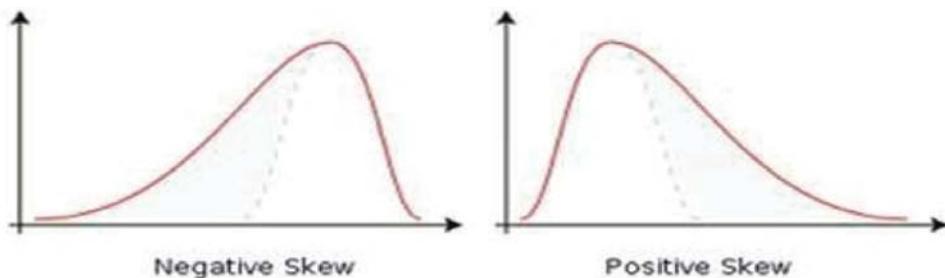
Three characteristics completely describe any distribution: Shape, central tendency, and variability. We will later examine central tendency and variability. Let's try to understand the shape of the distribution now. Nearly all distributions can be classified as being either symmetrical or skewed.

- In a symmetrical distribution, it is possible to draw a vertical line through the middle so that one side of the distribution is an exact mirror image of the other.
- In a skewed distribution, scores tend to pile up toward one end of the scale and taper off gradually at the other end. The section where the scores taper off towards one end of a distribution is called the tail of the distribution. A skewed distribution with the tail to the right-hand side is said to be positively skewed because the tail points toward the positive (above zero) end of the X-axis. If the tail points to the left, the distribution is said to be negatively skewed.



Symmetrical Distribution

Positive and Negative Skew



Check Your Progress: 2

Find frequency distribution for Male journalists, Female journalists, and Total Journalists.

S.No.	Age Intervals	Male	Female
1	Under20	0	100
2	21-25	17	83
3	26-30	40	60
4	31-35	35	65
5	36-40	51	49
6	41-45	48	52
7	46-50	55	45
8	51-55	56	44
9	56-60	53	47
10	61-65	56	44
11	Over 65	67	33

Find frequency distribution for Girls, Boys and Total Class. The dataset is about the time a class of students spends watching television programmes in a month.

S.No.	Time Intervals	Girls	Boys
1	140-149	0	0
2	130-139	0	32
3	120-129	0	48
4	110-119	1	29
5	100-109	0	18
6	90-99	3	14
7	80-89	5	5
8	70-79	6	5
9	60-69	14	0
10	50-59	7	1
11	40-49	11	0
12	30-39	4	0

16.4 MEASURES OF CENTRAL TENDENCY

You are quite familiar with what an average is and how it is calculated. When asked to calculate the average, most of us rush into calculating the mean. But this isn't exactly the only meaning of the 'average'. In statistics, you can calculate the so-called average in three different ways. Collectively, these three - mean, medium, and mode, are called measures of central tendency. You can pretty much guess what central tendency is. As the phrase connotes, the central tendency is an average that resides in the centre or represents the centre of the distribution. Measures of central tendency give us a number that describes where the centre lies. Knowing the centre of a distribution tells us a great deal about a distribution because most of the scores in a distribution tend to cluster around the centre.

16.4.1 Mean

The mean is the arithmetic centre of the distribution. You can find the mean by adding all the scores (OX) together and dividing the total by the number of scores (N). For example (X = scores 1, 2, 3, 4, 5; N=5):

$$\text{Xare } 1+2+3+4+5 = (\text{OX}) 15$$

$$N=5$$

The population mean(μ) is: $15/5 = 3$ ($\mu = \text{OX}/N$). The sample mean (n) is: $15/5$.

$$= 3 \ (\bar{x} = \text{OX}/n)$$

You may now note that the procedure we used to calculate the mean of the population and the sample is the same, but we have used different symbols for them.

We now know how the mean is calculated. Means have four properties. Here are four of them:

1. The mean is the most reliable and often used central tendency measure.
2. The mean need not be an actual score in the distribution.
3. Outliers strongly influence the mean.
4. The sum of the deviations about the mean must always be equal to zero.

You should understand the last two properties of the mean listed above. First, you should know that an outlier is an extreme score. It is a score that stands apart from most of the scores in the distribution. Thus, in a set of scores 1, 2, and 9, score 9 is an outlier.

Outliers tend to pull the mean towards them. Thus, a distribution of values 1, 2, 3, 4, and 5 averages 3. Three does a good job of describing where most of the scores in this distribution lie. However, if there is an outlier, say 25 (instead of 5) in the above distribution, the mean changes greatly. The new distribution 1, 2, 3, 4, 25 has a mean of 7. Seven is not close to most of the other values in the distribution. Thus, the mean is a poor measure of the centre when we have outliers.

16.4.2 Median

The median is the physical centre of the distribution. It is the value in the middle when the distribution values are arranged sequentially. The distribution 0, 1, 2, 2, 3, 4, 5, 5, 5, 7, 8 has a median value of 4 because five values are above this point and five values are below this point (0, 1, 2, 2, 3, 4, 5, 5, 5, 7, 8).

If you have an even set of numbers, then there will be two values at the centre, and you will average these two values together to determine the median. For example, if we take out one of the numbers in the distribution so that we have 0, 1, 2, 2, 3, 4, 5, 5, 5, 7, then the two values in the centre are 3 and 4 (0, 1,

2, 2, 3, 4, 5, 5, 5, 7). The average is 3.5, and that is the median. You would have noticed that the median is resistant to outliers. That is, outliers will generally not affect the median and will not be affected as much as the mean. The median might move slightly toward the skew or outliers in the distribution.

16.4.3 Mode

The mode is the most frequent score in the distribution. Examine the following scores. 1, 1, 2, 3, 3, 3, 4, 4, 4, 4, 4, 5, 6, 7. What score appears most often in the distribution? 4. Hence, 4 is the mode. This distribution of scores has only one mode (4).

Now examine another set of scores: 1, 1, 2, 3, 3, 3, 4, 4, 4, 4, 5, 6. Here, 3 and 4 appear four times each. Therefore, there are two modes (3 and 4) in the set of scores.

Now consider this set of scores: 1, 2, 3, 4, 5, 6, 7, 8, 9. Here each score appears only once; no score appears twice or more. Therefore, you conclude there is no mode in the set of scores. You would have again noticed another characteristic of mode: it remains unaffected by outliers.

1. students were asked to select their favourite subject in a class. 63 students chose English, 29 chose Maths and 9 chose Statistics. What is the mode of this distribution?
2. Why do extreme scores (outliers) not affect the median?

Check Your Progress: 3

Find the mode, median, and mean of the following:

a) 19,24,28,28,30,34,38,42,49,51,52,61,64

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

b) 53,43, 77,44,35, 14,56,43, 75,45,80

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

c) 14,25,47,56,56,58,58,66,82,89,91

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

d) 61,52,81,56,35,17,56,51,82,35

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

e) 31,44,45,46,43,55,56,56,57,62,64,66,69

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

16.5 VARIABILITY

Our education and experience enable us to understand what a mean is. However, understanding the concept of variability doesn't come to us so easily.

Variability, in simple English, refers to difference (or distance). So, if a population of 10 students' scores 10 out of 10, it certainly is a great performance, but there is no difference in the score of the 10 students. They

have all scored the same marks! The statistician would say the same thing in a different language (there is no variability in the distribution of (10) scores, she would say). On the other hand, if the ten students score 1 through 10, there would be great variability.

In statistics, our goal is to measure the amount of variability for a particular set of scores. If the scores in a distribution are all the same, then there is no variability. If differences between scores are small, then the variability is small, and if differences between scores are large, then the variability is large.

In our everyday language, 'small' and 'big' are qualitative expressions. Each person has their own subjective measure of what these two words mean. In statistics, though, variability is quantitatively measured and expressed.

Why do we measure variability? We measure variability to determine how spread out a distribution of scores is. The purpose is to find out whether the scores are all close to each other (for example: 2, 4, 4, 4, 5, 5) or are scattered over a wide range of values (0, 0, 1, 2, 6, 9). Variability does much more than describe a set of scores. It helps us understand two additional characteristics of the distribution. First, variability tells us how well the mean (average) describes the distribution. If the variability is small, the scores are all close together, and each score is close to the mean (calculate the mean of the first set of scores and see this for yourself). In this situation, the mean is a good representative of all the scores in the distribution. On the other hand, when variability is large, the scores are all spread out and are not necessarily close to the mean (calculate the mean of the second set of scores). Is the mean of the second set of scores a good representative of the distribution as the mean of the first set?

The second characteristic of variability is kind of related to the first. Variability also tells us how well an individual score (or group of scores) represents the entire distribution. In the first set of scores (2, 4, 4, 4, 5, 5), the distance of any individual score (2, 4 or 5) is not too far from the rest of the scores. In this case, 2 or 4 or 5 nearly accurately represent the entire distribution. Can you say the same thing about the second distribution? Does any individual score represent the entire distribution? We can't be so certain here or as certain as we are with the first set of scores. Now, with this little theoretical background, let's determine how variability is calculated. There are three different measures of variability: the range, the interquartile range, and the standard deviation. Let's look at each of them.

16.5.1 Range

The range is the distance (or difference) between the largest and smallest scores in the distribution. So, the range of the two scores above is 3 and 9, respectively. (To be accurate, the range is the difference between the upper real limit for the largest (maximum) X value and the lower real limit for the smallest (minimum) X value).

You can perhaps see the difficulty of using range as a measure of variability. It seeks inputs from only two extreme values (scores) at the cost of all other

values in the distribution. In other words, it often does not accurately describe the variability for the entire distribution. For this reason, the range is considered a crude and unreliable measure of variability.

16.5.2 Interquartile Range and Semi-Interquartile Range

The second measure of variability is the interquartile range (or semi-quartile range). Using quartiles, you can divide a distribution into four equal parts (for easier understanding, think of dividing 100 into four equal parts!).

The first quartile (Q1) is the score that separates the lower 25% of the distribution from the remaining 75%.

The second quartile (Q2) is the score with exactly two quartiles or 50% of the distribution below it. The second quartile is the median (they are the same).

Finally, the third quartile (Q3) is the score that divides the bottom 75 per cent of the distribution from the top 25 per cent.

The semi-quartile range is one-half of the Interquartile range. It provides a descriptive measure of scores' "typical" distance from the median (Q2).

The Interquartile range is the distance between the first and third quartiles.

$$\text{Interquartile range} = Q3 - Q1$$

When the Interquartile range is used to describe variability, it is commonly transformed into the semi-interquartile range, which is simply one-half of the Interquartile range.

$$\text{Semi-interquartile range} = \frac{(Q3 - Q1)}{2}$$

Because the semi-interquartile range focuses on the middle 50% of the distribution, it is less likely to be influenced by extreme scores, giving a more stable measure of variability than the range. However, much like the range, the semi-interquartile range does not consider the actual distances between individual scores. In other words, it doesn't tell us how scattered or clustered the scores are.

For example, Q2 (the middle quartile) is the median.

Q1 (the lower quartile) is the median of the numbers to the left of or below Q2. Q3 (the upper quartile) is the median of the numbers to the right of or above Q2. Example: Find the lower, middle, and upper quartiles of the data

12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32

Since the data is already in ascending order, find the median. 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32

22 is the median; therefore, Q2= 22

The median of the numbers to the left of Q2: 12, 14, 16, 18, 20 16 is the median; therefore, Q1 = 16

The median of the numbers to the right of Q2: 24, 26, 28, 30, 32 28 is the median; therefore, Q3 = 28

The inter-quartile range of a distribution is the difference between the upper and lower quartiles.

That is, inter quartile range = Q3 - Q1

Since, Q3 = 28 and Q1 = 16

Inter quartile range (Q2) = 28 - 16 = 12

The semi-interquartile distribution range is half the difference between the upper and lower quartiles or half the interquartile range.

Therefore, from the example above, the interquartile range is 12.

$$\text{Therefore, Semi Interquartile range} = \frac{(Q3 - Q1)}{2} = \frac{12}{2} = 6$$

16.5.3 Standard Deviation and Variance for a Population

Now, we come to the third measure of variability: the standard deviation. Unlike the first two measures of variability, the standard deviation considers the distances between individual scores. So, naturally, it tells us something important that the two other previous measures don't: how scattered or clustered the scores are in a distribution.

Well, by now, you must be wondering how one calculates the distance between individual scores. Think of it this way. How is the distance measured in a city? The standard way of measuring distance is from one central point to another. The same logic is used in calculating the standard deviation. The central point, in this instance, is the mean. So, we measure the deviation of the scores from the mean.

Standard means typical; the deviation is the distance from the mean

The standard deviation can be measured in six simple steps. Let us illustrate this using a set of scores (2, 4, 4, 4, 5, 5).

Step 1 is to add up the scores. $2 + 4 + 4 + 4 + 5 + 5 = 24$

Step 2 is to compute the mean of the set of scores. You do this by dividing the total (24) by the number of scores ($N = 6$)

$$24 \div 6 = 4$$

Step 3 is to calculate the mean of the deviation scores. To do this, subtract the mean from each score and add the deviation scores.

$$2-4 = -2$$

$$4-4 = 0$$

$$4-4 = 0$$

$$4-4 = 0$$

$$5-4 = 1$$

$$\underline{5-4 = 1}$$

$$0$$

When you do this, you end up with zero (because -2 and + 2 cancel each other out). You will observe that the distances above the mean are equal to those below the mean. This will always be so. Therefore, the mean deviation score provides no information about variability.

Step 4 allows you to overcome this by squaring the difference between the mean and individual scores (squaring helps eliminate the + and—signs) and adding them up. You get when you do this for the set of scores in the example.

$$\begin{array}{r} 2-4 = -2 = 4 \\ 4-4 = 0 = 0 \\ 4-4 = 0 = 0 \\ 4-4 = 0 = 0 \\ 5-4 = 1 = 1 \\ \underline{5-4 = 1 = 1} \\ \hline 6 \end{array}$$

At this step, you have eliminated the signs and can add up the squared deviation scores to get a value that is not zero (in this case, 6). Several textbooks call this the Sum of Squares or SS. The sum of Squares is an important concept used in several statistical tests. You must remember what it is and how it is calculated.

The sum of squares (SS) is the sum of the squared deviation scores.

Step 5 takes us to another important concept called variance. You compute variance by dividing the sum of squares by N. In the example,

$6 / 6 = 1$. This is the variance.

Variance is the mean of the squared deviation scores.

The final step, Step 6, requires taking the square root of the variance.
Standard Deviation = $\sqrt{\text{variance}}$

Standard deviation is the square root of the variance

In the example, $\sqrt{1} = 1$

Why do you take the square root of the variance? You do it because you had squared the deviation scores in Step 4. Remember, you had done it because you would have otherwise ended up with zero as variability. You then squared the deviation scores to get over this difficulty. Now, you would like to remove the influence of squaring the deviation scores to obtain the standard deviation. You do this by taking the square root of variance in Step 6.

The concept of standard deviation (or variance) is the same for a sample as for a population; however, the details of calculations differ slightly depending on whether you have sample data or a complete population. Below are the formulas for calculating the sum of squares, variance, and standard deviation of a population of scores.

The sum of squares (SS) is the sum of the squared deviation scores.

$$SS = I(X - \mu)^2 \text{ or } SS = \frac{\sum X^2 - (\sum X)^2}{N}$$

$$\begin{aligned} \text{Population variance} &= \frac{SS}{N} \\ \text{Population standard deviation} &= \sqrt{\frac{SS}{N}} \end{aligned}$$

E.g. For the following population of scores, calculate SS, Variance and Standard Deviation

2, 10, 6, 9, 8, 7

16.5.4 Standard Deviation and Variance for Samples

Variance and standard deviation for sample data have the same basic definitions as they do for populations: To compute these, we first need to find SS, the sum of squared deviations. The formulas we use to compute sample SS are identical to those used for populations.

Definitional formula: $SS = I(X - \bar{X})^2$

$$\text{Computational formula: } SS = \frac{\sum X^2 - (\sum X)^2}{N}$$

Note that the difference between these formulas and the population formulas is minor. We have substituted \bar{X} in place of μ and n in place of N . The difference in notation will not affect the calculations.

It is necessary to adjust the formulas for sample variance and standard deviation to correct the bias in sample variability. This is done by replacing n with $n-1$:

$$\text{Sample variance} = S^2 = \frac{SS}{n-1}$$

Sample standard Deviation (S) is simply the square root of the variance.

$$\text{Sample standard Deviation} = S = \sqrt{\frac{SS}{n-1}}$$

Dividing the sum of squares in the numerator by $n-1$ instead of n increases the value you will obtain. The adjustment makes sample variability an accurate or unbiased estimator of population variability. For this reason, sample variance is often called estimated population variance, and the sample

standard deviation is called estimated population standard deviation.

Check Your Progress: 4

Find the Range, Quartile Range, Variance and Standard Deviation of the following:

1) 5,14,13, 9, 7, 7, 8, 6,12

.....
.....

2) 15, 18,9,23, 18

.....
.....

3) 9,3,9,6,6,9

.....
.....

4) 28,39,42,29,39,40,36,46,41,30

.....
.....

5) 133,215,250,108,206,159,206,178

.....
.....

6) 76,94, 76,82,78,86,90

.....
.....

7) 52,61,49,52,49,52,41, 58

.....
.....

8) 58,59,63,45,44,41,36,36

.....
.....

9) 29,29,31,41,42,43,31,29,23,25

.....
.....

10) 59,54,37,45,31,60,32,34

.....
.....

16.6 LET US SUM UP

In this Unit, we explained the basic analytical tools under the descriptive statistics category. We discussed the fundamental concepts of quantitative research. The basic concept of quantitative techniques is to make generalisations of research findings based on a group of samples. Statistics plays a crucial role in the generalisation process.

We also discussed the building blocks of measurements - Nominal, Ordinal, Interval and Ratio. Before collecting any quantitative data for a study, the researcher should decide the nature of the statistical applications s/he will employ. This determines the nature and need of the scale of measurements in data collection instruments. These measurements would generate a different data set, and how it can be classified as continuous and discrete variables were discussed.

It was further outlined how these data can be classified and summarised in the form of frequency distributions, with the nature and types of graphical presentation of data, measures of central tendency—mean, median, and mode—and measures of dispersion—range, quartile, variance, and standard deviation.

It is hoped that you will be able to use the above statistical methods to analyse your data when embarking on your research journey.

16.7 FURTHER READINGS

1. Dayal, M. (2017). Media Metrics: An Introduction to Quantitative Research in Mass Communication, SAGE Texts
2. Rajagopalan (2012), Statistical Inference, Prentice Hall India Learning Private Limited
3. Rajaratnam T. (2015), Statistics for Social Sciences, SAGE Publishing India Kanji, G. K. (2006), 100 statistical tests, Sage

16.8 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Nominal
2. Ordinal
3. Ratio

Check Your Progress: 2

- 1.

S.No.	Age Intervals	Male	Male%	Female	Female%
1	Under 20	0	0	100	16
2	21-25	17	4	83	13

3	26-30	40	8	60	10
4	31-35	35	7	65	10
5	36-40	51	11	49	8
6	41-45	48	10	52	8
7	46-50	55	12	45	7
8	51-55	56	12	44	7
9	56-60	53	11	47	8
10	61-65	56	12	44	7
11	Over 65	67	14	33	5
	Total	478	100	622	100
	478+622=1100		43.5		56.5

2.

S.No.	Time Intervals	Girls	Girls%	Boys	Boys%
1	140-149	0	0.00	0	0.00
2	130-139	0	0.00	32	21.05
3	120-129	0	0.00	48	31.58
4	110-119	1	1.96	29	19.08
5	100-109	0	0.00	18	11.84
6	90-99	3	5.88	14	9.21
7	80-89	5	9.80	5	3.29
8	70-79	6	11.76	5	3.29
9	60-69	14	27.45	0	0.00
10	50-59	7	13.73	1	0.66
11	40-49	11	21.57	0	0.00
12	30-39	4	7.84	0	0.00
	Total	51	100.00	152	100.00
	51 + 152 = 203		25.12		74.88

Check Your Progress: 3

Q. No	Mean	median	mode	Q. No
1	40	38	28	1
2	51.4	45	53	2
3	58.4	58	56, 58	3
4	52.6	54	56	4
5	53.4	56	56	5

Check Your Progress: 4

Q. No	Range	Quartile Range 0%o,25%o,50%o, 75%o,100%o	Variance	Standard Deviation
1	5, 14	5, 7, 8, 12, 14	10.5	3.24
2	9,23	9, 15, 18, 18, 23	26.3	5.13
3	3,9	3, 6, 7.5, 9, 9	6	2.45
4	28,46	28,31.5,39,40.75,46	37.1	6.1
5	108,250	108, 152.5, 192,208.25, 250	2183.84	46.73
6	76,94	76, 77,82,88,94	50.48	7.10
7	41,61	41, 49, 52, 53.5, 61	36.5	6.04
8	36,63	36, 39.75, 44.5, 58.25, 63	115.36	10.74
9	23,43	23, 29, 30, 38.5, 43	51.12	7.15
10	31,60	31, 33.50, 41, 55.25, 60	149.14	12.21

UNIT 17 REPORT WRITING

Structure

- 17.0 Introduction
 - 17.1 Learning Outcomes
 - 17.2 Stages in Report Writing
 - 17.3 The Beginning
 - 17.3.1 The Title
 - 17.3.2 Abstract
 - 17.3.3 Acknowledgements
 - 17.3.4 List of contents
 - 17.3.5 Introduction
 - 17.4 Main Body of the Report
 - 17.4.1 Literature Review
 - 17.4.2 Objectives
 - 17.4.3 Methodology
 - 17.4.4 Sampling and Sample
 - 17.4.5 Research Instruments
 - 17.4.6 Data Presentation
 - 17.4.7 Findings and Analysis
 - 17.5 The Final Section
 - 17.5.1 Summary
 - 17.5.2 Conclusions
 - 17.5.3 Recommendations
 - 17.5.4 Appendix
 - 17.5.5 References
 - 17.6 Effective Writing
 - 17.7 Let Us Sum Up
 - 17.8 Keywords
 - 17.9 Further Readings
 - 17.10 Check Your Progress: Possible Answers
-

17.0 INTRODUCTION

Report writing is both an art and a craft. In the previous Blocks and Units, you have learnt the whole process of communication research, research methods, and data analysis. This unit focuses on bringing it all together in the form of a report that reflects your work. It presents the report to the reader, detailing what it means and what you have found during the research process.

A report is a structured, formal document written to convey factual, research-based information. This information is often used as the basis for making decisions and acting. Research reports generally follow a particular standardised structure and format, which are described and discussed in this unit.

17.1 LEARNING OUTCOMES

After completing this unit, you should be able to:

- Outline major components of a research report;
- discuss various stages in research report writing and
- write good research report.

17.2 STAGES IN REPORT WRITING

The following stages are involved in writing a report:

- Planning the report
- Organising the information
- Structuring the information
- Writing the first draft
- Checking and revising
- Writing the final version

Planning your work: You will be writing a report for a reason. The best reports are the ones that are clear and purposeful, so you need to ask yourself:

- Who am I writing for?
- Why am I writing this?
- What am I trying to say?
- How should I be saying it?

A report usually falls into one of the following categories:

Informative - presents statements of fact and opinions.

Persuasive- tries to obtain approval from the reader or advocates a position.

Explanatory - answers given questions and gives a conclusion. Operational - records an event and gives evidence of work done. Which category does your report fall into?

Activity: 1

Answer the following four questions: Who am I writing this report for? Why am I writing this? What am I trying to say? How should I be saying it?

A report informs the readers and, if it is logically presented, can be extremely useful to both the reader and the writer. It is a more structured form of writing than, say, an essay and is designed so that it can be read quickly and accurately, though reports may not necessarily be read in a consecutive order. Structure and convention in written reports stress the processes by which the information was gathered as much as the information itself. Reports should be organised for the convenience of the intended reader. These are written on a wide range of

subjects for various reasons. A research report follows the general principles and standard structure of general report writing.

Format of the Report

Over the years, a standard format for reports has been worked out. Although there may be circumstances when it is advisable to change the format to fit a particular need, generally, the structure is common. The standard sections of a report are as follows:

The Beginning

The main body of the report

The Final section

Let us discuss each section in detail.

17.3 THE BEGINNING

The Beginning of the report contains the following: Title Page, Abstract or Executive Summary, Table of contents, List of tables, List of figures, and Introduction. Let us examine each section of a report in greater detail.

17.3.1 The Title

The report title is the first thing a reader sees. It tells them of the report's content and should thus be precise, clear, and concise; it should also give the reader ample information to differentiate it from other reports on similar subjects.

17.3.2 Abstract

An abstract, sometimes called the Executive Summary in a report to an organisation, is an independent text written in a brief, non-repetitive style containing the essential details of the research presented in the report. It offers a quick way for the reader to decide whether the report is relevant for him/her and whether it contributes important knowledge/key results/valuable scientific methods. An abstract is usually written for a wider readership than the report. You should not include extensive background, literature review, detailed methods, figures, or references. Also, generally, you should not add any material not in the report. For short reports, about 150 words are enough; for a longer report, the abstract should not exceed one page. An executive summary may be longer, but it is never two or three pages long. Typically, abstracts of research reports, including dissertations, are often made publicly available on the Internet, enabling anyone interested in the subject to review and see if the content is relevant to their research.

17.3.3 Acknowledgements

Acknowledgements are an important part of the research report, and you can thank those who have helped you in any way during the research work. In an academic report, they are generally your research guide, as well as the department's faculty, administrative, and support staff. In addition to people, you may thank organisations, institutions, or libraries you visited for material

collection. They can also be people who have extended you emotional support, such as your family, friends, and near and dear ones. You should not forget to thank the respondents if your study is based on field research, as there will be no study without them. You should be generous in your approach but not gush and make it sentimental. Using various words and expressions for different sets of people adds to the impact of your acknowledgement section.

17.3.4 List of Contents

The list or table of contents helps the reader find specific sections of the report at one glance. The contents are presented at the beginning of the report as a list of section and sub-section headings, along with the corresponding page numbers.

17.3.5 Introduction

This section broadly introduces the research by outlining its rationale, aims and objectives. The purpose of the introduction is to set the scene for your readers, so they know what to expect from the report. The introduction answers the following questions:

- Why is the report being written?
- What kinds of information does it contain?
- How is the problem being approached?
- For whom is the report being written?

It is often useful to try a three-paragraph approach to your initial report writing:

Paragraph 1 - What are you investigating, and why is it interesting?

Paragraph 2 - What information is already known on the topic?

Paragraph 3 - What were the aims of the study you have just done?

Ensure you connect the paragraphs with a linking sentence following the fundamentals of good writing.

Check Your Progress: 1

Note: 1) Use the space below for your answers

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

1. What are the three key sections of a research report?

.....

.....

2. Describe the sub-sections in the three key sections of the research report.

.....

.....

17.4 MAIN BODY OF THE REPORT

The body of the report includes several important sections - theoretical framework, literature review, research design, objectives, methodology, sampling and sample, research tools, data collection process, findings and analysis, and interpretation.

The body presents the actual content of the report and describes the subject matter. Being the most extensive section of a report, it answers several questions like, "What did I do?" "How did I do it?" "What did I do it with?" and "What did I find?" The section should have detailed information on what you have done so that the reader understands it clearly, can replicate the tests if required, and can check their validity.

Now, let us understand some of the important sections of the main body of the report in detail.

17.4.1 Literature Review

In this section of the report, one locates and reviews the articles, reports, and documents related to or about the research topic or question under consideration. This section also gives an insight into the theoretical framework that may have been used previously and how that could help you relate, generate, and base your research on a sound theoretical base. In Block 1, Unit 5, we discussed how the literature review is done. In this section of your report, you must present the findings of the survey of the literature, summarise previous research, and place the review in the context of your work. From this context, you can connect your research to previous work, clarify the gaps in the field, and justify your research. This section is not a listing of what research was done by author or title; rather, it is a compelling argument connecting your work to that of others. The literature survey is organised around ideas and themes, and you should conclude what you have learned from reviewing the literature so far. After reading your review, your reader should be convinced that your proposed research project will be necessary to further knowledge in your field.

17.4.2 Objectives

The section on objectives draws the reader's attention to the specific issue under study. Though each research study has its specific purpose, we may think of research objectives as falling into several broad groupings:

- To gain familiarity with a phenomenon or to achieve new insights into it. Studies with this object in view are termed exploratory or formative research studies.
- To accurately portray the characteristics of a particular individual, situation or group, studies with this object in mind are known as descriptive research studies.
- To determine the frequency with which something occurs or is associated with something else. Such studies are known as diagnostic research studies.

- To test a hypothesis of a causal relationship between variables. These types of research studies are called hypothesis-testing research studies.

The objectives of the given research must be stated in unambiguous terms. Following the statement of objectives, the hypotheses to be studied should also be stated clearly.

17.4.3 Methodology

The first section of the Methodology chapter provides a broad picture of the study's approach. Here, you need to state whether the study is exploratory, historical, descriptive, ethnographic, or experimental. Each approach carries with it a set of methods that are usually followed. Once you have defined the approach, you can describe what methods and tools you will be using.

17.4.4 Sampling and Sample

It would help to describe the study population as explained in Units 4 and 14. Then, describe the sample size, what kind of sampling procedure was followed, and how exactly the sample was drawn and why. At this stage, a flow chart would give a graphic description of the sampling process. You should also outline the sampling error or confidence interval (discussed in Block 1, Unit 4). Describe what problems were identified and how they were addressed in sampling from the population to help the reader understand the sampling context. For example, in an experimental design, you must describe how you selected the subjects and placed them in either the control or experimental group.

17.4.5 Research Instruments

Before describing your research instruments, the report must state the 'unit of study' and the 'unit of analysis'. Let us understand this with an example. If you are conducting a survey, the unit of study is the individual respondent, whether it is a questionnaire or an interview schedule. However, when you report your data, you will report results in broad categories and groups. These broad categories are your unit of analysis.

The next step is to describe your research instruments, such as the questionnaire, interview schedule, attitude scale, or knowledge test (in an experimental design). Describe the structure of the instrument and what each section contains. Place the research instruments in the Appendix.

You also need to describe how the instruments were pretested and administered by mail, telephone, online, as individual interviews, or in groups of people and/or when the knowledge test or attitude scale was administered. You may mention any limitations you faced when you collected data to place the research you have done within a context.

Mention any limitations you faced in the field during data collection. For instance, these could be smaller samples, low survey response rates, respondents' resistance to answering your questions, or even problems with human resources and logistics, which may have impacted the overall findings. Writing this here helps the reader understand why the research may

have deviated from the original design and helps to place the findings in context.

17.4.6 Data Presentation

Describe the data processing operations regarding sifting, checking for accuracy, classification, and categorisation.

- **Sifting** is a process of examining the collected raw data to detect errors and omissions and correct them when possible.
- **Classification-** the process you followed in arranging data in groups or classes based on common characteristics.
- **Categorisation according to attributes**—here, data is analysed based on common characteristics, which can either be descriptive, such as literacy, sex, religion, etc., or numerical, such as weight, height, income, etc. You can also include the data levels used (discussed in Unit 14) and whether the data were nominal, ordinal, interval, or ratio.

Such classification can be either:

- **Simple classification:** We consider only one attribute and divide the universe into two classes—one class consisting of items possessing the given attribute and the other class consisting of items that do not possess the given attribute.
- **Manifold classification:** Here, we consider two or more attributes simultaneously and divide the data into several classes.
- **Classification according to class intervals** is done with data relating to income, age, weight, tariff, production, occupancy, etc. Such quantitative data are known as the statistics of variables and are classified based on class intervals.

You also need to describe the data analysis methods used and the statistical procedures followed in the study. In this section, you would describe what statistics you used; for example, are they just frequency distributions describing the data, or did you test the data for association or significance? If you tested the data, describe the tests and how you calculated significance. In this section, you need to justify your choice of statistics and/or, in the case of qualitative data, how you analysed the data and presented your data.

17.4.7 Findings and Analysis

In this part of the report, you must present all the evidence collected for your research. You should ensure that the information is organised logically so that conclusions can be drawn naturally from the facts presented. The findings may be structured in the following ways.

Describe the real sample in the first part of the findings and analysis section. The final sample of your study is never the original sample drawn; the real sample is always lower than what you planned. If you have a 15 per cent response rate on your questionnaire-based survey, say so. The quality of your findings depends on this information.

- You might want to present the demographics in one or many tables. If possible, collapse the demographics table into one where a full sample picture emerges. Always present a table first, describe and explain the findings regarding the most important factor first, and proceed to the least important. For instance, "70 per cent of the sample agreed that. . ." and then proceeded with lesser percentages and numbers. Give both percentages and frequencies (n) here.
- The second section of the findings and analysis section focuses on the research questions and hypotheses being tested. Again, present your data in a sequence of each hypothesis listed, with the tables showing the results (measures of central tendency, dispersion, variation, association, and significance). Display the table first, then analyse the results to determine whether the hypothesis is upheld or disproved. Never place a table without an explanation.
- Tables, graphs, and diagrams are useful for presenting information in a manner that is easily understood. If the information being presented through these tools is necessary to prove the study's main objective, then it is prudent to place it within the main text or the body. If not, these can be placed at the end of the report in an appendix. Also, you should ensure that tables and diagrams are properly labelled, and the information is presented systematically to not confuse the reader.
- Do not report unnecessary data. Just because you collected vast amounts of information during your fieldwork and analysed it, you do not have to report all of it. Excess information will clutter your findings. You can use the additional data for other research papers.
- Do not include any opinions in this section—report only what you find. Your opinions do not belong here, and they will only distract the reader from your research study's main objectives and hypotheses.

As the writer, it is your job to present the facts and evidence related to the subject of your study to the readers. With that in view, it is not enough to just present a graph or table and expect your reader to pick out what is important. Therefore, you must include some discussion of the facts, but to repeat, there are no opinions or judgments about the findings.

In this section, discuss your findings in terms of their relationship to your study's objectives and previous literature on the subject. You can include the reasons for the similarities or differences between previous research studies. You must link your study to the body of theory and research already conducted because it brings continuity, and the cumulative nature of research adds value to your hard work.

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. List the sub-sections of the Main body of the report.

2. What is a literature review?

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

3. What are the different ways of data presentation?

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

17.5 THE FINAL SECTION

The final section of the report presents a summary and conclusions of the report. It also comprises recommendations, an appendix, references and/or a bibliography.

17.5.1 Summary

In this section, you should summarise your report in terms of what you had set out to do, what you did, how you went about your research, and what your findings were - be brief here; no need to elaborate. Readers often look at this section before proceeding to the main body of your work, so all necessary information should be repeated here.

17.5.2 Conclusions

The conclusions should arise naturally from the evidence presented in the previous sections. You may include, for example, statements on what has happened, what the situation is and what might happen. You may also consider further action that could be taken, as well as an analysis of the advantages and disadvantages of various courses of action. If appropriate, you should give your recommendation, considering your experience and the evidence you have presented, of a preferred course of action. Suppose any gaps in your knowledge prevent you from concluding. In that case, you should clarify and perhaps outline what further investigations would be necessary to provide fuller information. You should not include any new information that does not appear in the main body of the report and should not make statements that you cannot support with the evidence.

17.5.3 Recommendations

It is not always necessary to present recommendations in a report, but if they are included, they should flow logically from the study's findings and conclusions. Usually, the recommendations are set out as brief statements and suggestions for further research.

17.5.4 Appendix

The appendix is the section of the report that contains all the secondary or marginal information which does not directly influence the central objective of your study. This information is presented separately so as not to overload

your main argument. Texts of questionnaires, interview questions and transcripts, specification sheets, etc., can be presented in the appendix if needed. Some of the other materials that may be included are as follows:

Glossary: It contains the meanings and explanations of the technical terms used throughout the study and is best presented in an Appendix. A glossary of terms is especially useful when presenting a technical study to people who may not be subject experts. The terms and their concise definitions are listed alphabetically in this section. You may also include acronyms, i.e. short forms for names of organisations in this section.

Tables: When presenting a study with a lot of statistical data, it is not always possible to present all of it in the main body. In such a scenario, the data is placed together in an Appendix. When presenting data through tables in an Appendix, it is important to give them clear numbers and titles and present them in the order of their reference in the text.

Diagrams and Illustrations: Like tables, diagrams can also be presented within the main text. But, if the writer feels that this disrupts the text too much, the diagrams and illustrations may be presented in the Appendix.

17.5.5 References

All references should be given in full and in alphabetical order. For detailed guidelines on punctuation, capitalisation, and abbreviations in reference citations, see APA, MLA, Harvard, Chicago, CSE, or other relevant manuals. The organisation for which the report is being prepared will have its style manual. Ask for it and follow it carefully.

Book: The author's surname is written first, followed by the initials or the full first name, the date of publication, the complete title of the book, the edition, the place of publication, and the publisher.

Journal article: the author's surname with initials or the complete first name; date of the journal's publication; full title of the article (in quotation marks); journal name (in italics); volume and issue number of the journal; page numbers on which the article can be found.

Electronic or Website reference: The general principle in APA style is to include the same information as you would for a print source, in the same order, and add just enough information about finding the source electronically to help your reader find that reference. If an online source has a digital object identifier (DOI), a permanent identifier for an electronic publication, you should use that instead of a URL. Also, cite the date on which you accessed the document.

A Note about Plagiarism

Research-based writing in Indian or international institutions, both educational and corporate, is filled with rules that writers, particularly beginners, are not aware of or don't know how to follow. Many of these rules have to do with research and proper citation. Gaining familiarity with these rules, however, is critically important, as unintentional mistakes can lead to charges of plagiarism - the unaccredited use (intentional and unintentional) of

somebody else's words or ideas. You can avoid plagiarism by knowing what must be documented when using information from various sources. As a rule of thumb, always provide references and citations.

Specific words and phrases: When using an author's words verbatim, they must be placed in double quotation marks, and due credit must be given to the source.

Information and Ideas: Even when paraphrasing or presenting someone else's ideas in your own words, you must document the source of the information.

Information: If a piece of information is not common knowledge, you must provide a source, even if it is an unpublished document.

Ideas: As a writer, you must acknowledge the source/ author if it provided - points made and conclusions drawn, specific methods or theory, the arrangement of material, or a list of steps in a process or characteristics of a medical condition, etc.

Check Your Progress: 3

Note: 1. Use the space below for your answers

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

- 1) List the different sub-sections of the final section of the research report.

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

- 2) What are the different ways of documentation styles?

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

- 3) Differentiate between conclusions and recommendations.

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

17.6 EFFECTIVE WRITING

Several aspects make writing effective. We have identified six features of effective writing.

Clear Objective: A clear objective is the most important feature of any writing. The purpose of all the sections and paragraphs in the report must be completely clear.

Good Organisation: A well-organised document always presents information in the Order of Importance. All sections should start with the conclusion and then elucidate it with supporting information.

Brief and Concise Writing: The reader should never be overwhelmed by information. It is prudent to be clear and concise and avoid unnecessary or peripheral information wherever possible. As a writer, you should carefully use shorter sentences and familiar words.

Appropriate Language: You should avoid using too much jargon or technical words and use simple words that readers can easily understand. If you must use a technical term, define it the first time you use it.

Relaxed Style: Style is something personal. It's 'how' you express your ideas rather than what you say. Your style should aim to be persuasive and interesting.

Correct Spelling, Grammar, and Punctuation: The content's accuracy is paramount. Although a good report may not directly depend on your language capabilities, it does depend on your ideas and ability to persuade others. Therefore, writers are advised to be accurate and to do a spell and grammar check.

Numbering: In this Unit, the point numbering system is used. Main sections are numbered 1, 2, 3, etc., with each sub-section given a decimal subdivision. This system of numbering helps readers keep track of where they are within the report.

However, this is only one numbering style, and it is unnecessary to use it always. Some documents, like government reports, may number all paragraphs consecutively, and others may completely skip numbering. As a writer, you should follow the technique that best suits your work.

Other Conventions

A writer needs to remember some other rules for presenting a report. The documents should ideally be typed, double-spaced on one side of the paper, and have adequate margins, as an overcrowded page is difficult to read. Remember to correctly number your pages and refer them to the table of contents.

Most organisations have their own rules for presenting reports and other documents, and these may vary from others. You should follow the rules laid out by the organisation you work for.

Style, Presentation and Level

Everyone has their writing style, which should show through even when your material must be presented in a conventional format. Convey to your reader your enthusiasm for your investigation without using colloquial phrases ("Branded cumulative experiment on the roof ') or unnecessary scientific jargon ("most interesting investigation into the psychology of a member of the common Indian Diaspora that occurs on the geographic stratification of our local coastal Australian cities"). Write clearly and simply. Try reading it

aloud; if your sentences are so mouthful that you stumble over them, redo them.

So, who do you pitch the report to? How much information is necessary? Your friends, parents and a professional environmental scientist may read your report. The best idea is to write it for an intelligent person (for example, a colleague who is also doing communication research) with a general knowledge of the subject like you. You do not need to go into detail, but you do explain new information or unusual methods that you alone know about.

Do whatever is necessary to include the readability. Do not hesitate to use the semicolon, short sentences, 15-20 words long, with one idea per sentence. Run-on sentences can be very annoying; it is a poor style to repeat a complex subject, just to be exact. In some cases, putting paragraph indents may also be clearer, especially for separate mathematical formulae or quotations. Use active voice rather than passive voice.

Typing a report produces a document that looks good, but typing is not essential. Whatever method you use, there are several things to check (that are more visible on a typed report). It is vitally important you do this. What sounded perfectly reasonable to you at 4 am often does not sound reasonable to anyone else on the planet the next day. By proofreading, you also ensure that you are not making silly mistakes. Below is a checklist you should follow:

1. Typographical and spelling errors in the text
2. Mathematical or typographical errors in tables and graphs
3. Omissions of headings and captions on tables and figures
4. Errors in the references
5. Try reading the work carefully (and slowly) sometime after you have finished.

Writing well is not easy. Even great authors find it difficult. The best way to write is to write, edit, rewrite, edit, and revise several times until the report is clear, concise, and readable. One way to do this is to have someone else read and/or edit your report.

Check Your Progress: 4

Note: 1. Use the space below for your answers
2. Compare your answers with those given at the end of the unit

- 1) What are the formats of reference styles?

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

- 1) What are the elements of editing and rewriting?

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

- 3) What does 'Style' of writing mean?

Activity: 2

You can download some research reports from the net or access them from the library and analyse the typical report format structures and variations in them.

17.7 LET US SUM UP

In this unit, we have taken you through writing a research report simply and sequentially. We discussed that a research report is a presentation of facts and findings, usually as a basis for recommendations, written for a specific readership, and probably intended to be kept as a record. A research report follows a typical format - the beginning, the main body of the report, and the final section. The unit's structure, given at the beginning, can serve as a template for what should be included and how a report should be structured.

Writing a report is perhaps one of the most difficult parts of the research process because there is so much material that it is hard to include everything. You are also so deeply involved that you sometimes do not see the errors. Hence, due care needs to be taken while writing a research report. In the final analysis, this report is presented for evaluation by an examiner, for a panel of experts, or even for other researchers to read.

With this unit, we have come to the end of the course on Communication Research Methods. The purpose has been to provide you with enough input and knowledge to enable you to undertake systematic research in your area of interest and write the report. Remember, the more research you do, the better you will become and the more you will contribute to generating scholarship in communication research. Practice makes perfect.

17.8 KEYWORDS

APA Format: APA refers to the American Psychological Association format for documentation. Typically, APA style is used for social science, biological science, and nursing research papers.

Bibliography Card: A bibliography card contains the information needed when writing the Reference or Works Cited page. Use a different bibliography card for each reference source or keep the information in a file on your computer.

Bibliographic Information: The last page of your research paper or

documented essay will list all the sources you used to write your paper. Each source must contain bibliographic information, including author, title, and publication information.

Citation: Each entry on the Works Cited or Reference page is called a citation, and it contains the bibliographic information needed for a reader to find the source information.

Journal: Journals are periodicals written by and for professionals in specific fields.

MLA Format: MLA refers to the Modern Language Association format for documentation. Typically, MLA style is used for language and literature research papers.

Paraphrase: One kind of citation to use is the paraphrase. When paraphrasing, you express another person's ideas in your own words. A paraphrase includes all the information from the source, but the wording is your own. All paraphrases must be documented with a parenthetical citation, and the source must be listed on the paper's Works Cited or Reference Page.

Parenthetical Citation: Parenthetical citation is used to identify reference information in the paper. A brief notation is placed at the end of the last sentence of each reference. It contains information needed to identify the source.

Periodical: Periodicals are publications that are printed at regular intervals, such as journals, magazines, newspapers, and newsletters.

Plagiarism occurs when other people's words or ideas are used without crediting the source.

Primary Sources: Primary sources are the source of information. They include original writings of an author, autobiographies, speeches, and laboratory experiments.

Quotation: A quotation incorporates an author's exact wording in your writing. Quotations should be used sparingly in your paper and must always be cited.

Reference Page: The reference page, using the APA format, lists all the sources used in compiling the paper. Each source must be fully documented with complete bibliographic information.

Secondary Sources: Secondary sources contain information about the primary source. They include newspapers, periodicals, and journal articles written about a topic. Biographies, critiques, and analyses are also examples of secondary sources.

Summary: A summary is one kind of citation used in research papers. It captures the most important ideas from the original source. It is written entirely in your own words and is shorter than the original material. All summaries must be documented with a parenthetical citation, and the source must be listed on the paper's Works Cited or Reference Page.

Works Cited Page: The Works Cited page, using the MLA format, shows all the sources used in compiling the paper or report. Each source is fully documented with complete bibliographic information.

17.9 FURTHER READINGS

1. Evans, D., Gruba, P., & Zobel, J. (2011). *How to write a better thesis*. Melbourne Univ. Publishing.
2. John Bowden, Writing a Report: How to prepare, write and present effective reports, Publisher: How To Books, 2004
3. Livingston, S. A. (2012). How to write an effective research report. *Research Memorandum ETS RM-12-05search*.
4. The Reporting Skills and Professional Writing Handbook, ELD Publications, 2008

17.10 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. The three key sections of the report are the beginning, the main body, and the final section.
2. The three subsections explain why it is interesting, what information is already known on the topic, and what the aims of your study were.

Check Your Progress: 2

1. The main body of the report contains many sections. It includes the survey of the literature, the statement of the problem, the research design, the methodology, the data analysis, the findings, and the conclusion.
2. A literature review summarises what other researchers have researched and written about the subject.
3. Data can be presented in many ways. Quantitative data are presented in tables and graphs accompanied by descriptions and explanations. Qualitative data can be organised and presented by theme, with case studies and quotations from respondents.

Check Your Progress: 3

1. Summary, conclusions, recommendations, appendix, references.
2. Both are ways in which citations are given in written reports.
3. Conclusions state what the research has found regarding previous theory or research; recommendations suggest a future course of action.

Check Your Progress: 4

1. Formats of reference style are APA, MLA, and Harvard.
2. Clear objectives, good organisation, appropriate academic style, correct language, and spelling.
3. How an individual writes to convey thoughts to other people.

UNIT 18 ETHICS OF INTERNET RESEARCH

Structure

- 18.0 Introduction
 - 18.1 Learning Outcomes
 - 18.2 Internet Research Issues and Concerns
 - 18.3 Internet as a Research Tool
 - 18.4 History and Development of Internet Research
 - 18.5 Ethical Research Considerations
 - 18.6 Let Us Sum Up
 - 18.7 Further Readings
 - 18.8 Check Your Progress: Possible Answers
-

18.0 INTRODUCTION

In research, we are familiar with primary and secondary research methods. In the primary research method, we gather data from participants, and in the secondary, data is collected from secondary information sources. Primary and secondary research can be conducted using the Internet in the digital era. Like we use the Internet to access secondary sources of information available from library databases and online journals, we can also use the Internet to recruit participants, administer materials, and collect responses as primary sources of information. Ethical concerns are involved in all activities related to research. Ethical concerns bother Internet research not only as a tool but also as a field. Since the Internet plays an important role in all areas and subareas, i.e. social sciences, arts and humanities, medicine, and biomedicine, from commerce to natural science, Internet ethics has become part of every discipline.

Using the Internet to locate secondary resources can have great pedagogical value as we get access to innumerable data published in areas related to our area of interest. When we compare the effort in terms of time and cost regarding reviewing the information available, searching for information on the Internet is quick and cost-effective compared to library-based research. However, though the work towards gathering the information appears simpler in the era of Internet-mediated Research, the challenges that surface while undertaking research using the Internet as a primary source cannot be sidelined.

18.1 LEARNING OUTCOMES

After completing this Unit, you should be able to:

- Understand the basic tenets of Internet research ethics.
- Explain the suitable course of action during any ethical dilemma situation.
- Undertake Internet related research with the suitable understanding on ethical issues involved in it.

18.2 INTERNET RESEARCH ISSUES AND CONCERNS

Internet-mediated Research has presented many challenges to researchers. If the participating respondents fear the misuse of their private information, the researchers are worried about the credibility of the participants. Since social media platforms provide scope for opening anonymous and fake accounts, the credibility of the participant in research work has doubled the challenges of a researcher. Identifying the participants, making them agree to participate and obtaining informed consent from the participants are additional problems of Internet Research. The easy availability of information through Big Data and the security of the data – Cloud Computing are other research challenges concerning primary data. Easy access to information has also led to plagiarism and copyright violations of the original author. The information downloaded can be slanted and adjusted to avoid legal concerns, Ex, Hacking. Artificial Intelligence support is another concern of IMR. Despite the challenges, undoubtedly, the exponential growth of the Internet has provided a new dimension to research. Today's researchers are exposed to unprecedented opportunities in the area of research.

Following ethics is very important for a person interested in taking up research work. Knowledge of ethics becomes more important if the research undertaken is based on the Internet. Knowledge of Research Ethics facilitates a student not only to understand the challenges but also regarding the methods of preparation to overcome the challenges. Research ethics is also now an essential part of the planning and management of big data research projects for the corporate sector. With social networking sites turning into enormous "living laboratories," e.g., dating apps, Instagram, or Facebook becoming a platform for # movements, understanding ethics is becoming important. The challenges leading to Internet research demand strict adherence to ethical guidelines, and researchers who wish to contribute to the knowledge related to their area of interest should adhere to established codes of conduct and ethical frameworks. Staying informed about evolving ethical standards and technological developments will enable the researchers to responsibly address emerging issues and challenges related to the Internet.

Research is a very broad term. It is an activity of gathering information to further an understanding of the topic of interest. Some post-collection analysis, such as a concern for quality or synthesis, might be included. Internet research is like any other systematic investigation and gathering of information using online resources. The activity related to research involves exploring websites, databases, and other online platforms to collect not only information but also data, facts, and necessary insights on a particular topic under study.

IRE is "the analysis of ethical issues and application of research ethics principles pertaining to research conducted on and on the Internet. Internet-based research, broadly defined, is research that utilises the Internet to collect information through an online tool, such as an online survey; studies about how people use the Internet, e.g., through collecting data and/or examining

activities in or on any online environments; and/or, uses of online datasets, databases, or repositories."

Internet research can be understood as the systematic process of gathering, analysing, and interpreting information from online sources to acquire knowledge or insights on a specific subject. Using the Internet to complete research makes longitudinal study practically feasible. Here, the researcher can establish connections with numerous people from various places who share an interest in the same subject to obtain and share information.

In a simple sense, Internet research is the practice of using the Internet to gather information available on different pages of the World Wide Web. In most cases, the search is to gather information from pages/ sites that are available for free. It is important to know that since Internet research provides immediate and worldwide access to information most of the time for free, the results may be affected by unrecognised bias and difficulties in verifying the source's credentials. Additionally, if the searcher lacks sufficient skill to draw meaningful results from the material available, the challenge of compiling the information will add to the credibility of the research. In addition, the challenges get further complicated if the site, once referred, cannot open again for various reasons, including reconstruction of sites. Thus, it is not advisable to work on the Internet without establishing guidelines to check the reliability of web pages under search.

In this pretext, where Internet has developed into a platform and tool for social communication and has turned present-day activities into data-driven from human-centred activity; in this digital era, there has been ongoing discussion over the development of the Internet as to whether new ethical conundrums are arising or whether those that already exist are comparable to those found in more conventional research methodologies. Since Internet-based research involves ethical issues that include challenges like participant knowledge and consent, data privacy, security, anonymity and confidentiality, the integrity of data, issues related to intellectual property, and norms of community and professionals, it is important to understand the ethics that need to be followed while conducting Internet-based research. Personal identifiability, reputational risk and harm, ideas of public space and public text, ownership, and data longevity are some ethical and methodologically complex issues surrounding Internet research.

Additional ethical concerns are related to issues of anonymity and confidentiality of data and challenges related to data integrity, as research projects can be outsourced or crowdsourced to online labour marketplaces. For a researcher, it is important to remember that with technological advancements, data is processed, stored, and disseminated via cloud computing or in remote server locales, adding to legal complexities. Further, the dominance of big data, which has enabled tools for Artificial Intelligence and Machine Learning for researchers, has its complexities.

Advances in technology have raised many more obligations. Since research activities include interventions in "Public" Internet spaces in terms of surveillance and data tracking, confidentiality or anonymity of the source, methods of obtaining informed consent, methods of researching minors

(children), chances of deception, and withholding identifiable information are some of the common challenges faced by researchers.

18.3 INTERNET AS A TOOL OF RESEARCH

Research activity on the Internet begins like any traditional research method. In the pre-Internet era of research, after spending countless hours in the library's reference section, combing through card catalogues, a researcher would identify the journals and indexes of books to be referred to. Going by the publication standard, the sources' reliability was judged, and the bibliography was prepared, listing all the information sources. Contrary to this time-consuming process, the Internet facility in the university or organisation provides access to thousands of major library catalogues in addition to their own in no time. The facility allows the researcher to search the data by subject, author, or title. A working bibliography can be compiled with software in a lot less time. There is also a searchable online database on the Internet that offers access to thousands of journals and tables of contents. With the prevailing facilities, if the researcher is educated about methods of using the Internet and tools of searching and identifying authenticated reliable sources, then research work in the present day can be undertaken without wasting much time.

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. Define Internet Research Ethics.

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

2. What are the issues and concerns of Internet Research?

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

18.4 HISTORY AND DEVELOPMENT OF IRE

The emphasis on adding ethical dimensions to Internet research has developed due to the extensive dependence on the Internet for research purposes. In the mid-1990s, there was a surge in Internet research and a clear focus on ethical issues simultaneously. It was a time when the Internet was used for email, messages, and chats. The specifics of the ethical issues involved are unclear when the field of study is young, but as it develops, data are gathered. As the usage of the Internet moved from the human element to their presence in virtual space, the 'field' of study changed the dynamics of researchers and research.

The reasons for this can be attributed to the persistent change in technology in the field of information and communication; secondly, the Internet completely transforms how information is gathered, accessed, and shared across all domains. Today, graphical user interfaces allow global access to factual and bibliographic databases. It is important to note that not only is information available in abundance, but the facility of having an interactive source has made the Internet a popular source of educational applications. Thirdly, the provision of interaction has facilitated the formation of a "Network Society" where a researcher connects with another beyond geographical boundaries.

Internet research ethics are conceptualised differently depending on the discipline, even though many researchers have stuck to the conventional research ethics guidelines. Certain academic fields, particularly those in the humanities and arts, assert that research conducted on the Internet focuses more on context and representation than on "human subjects," implying that researching real people has no purpose and, therefore, causes little to no harm.

However, ethical issues were raised when data was gathered online. Especially when it comes to using recorded interviews to gather historical significance, personal commentary, and recollections; these were gathered during personal interactions and transcribed later. However, if such documents are available online, then the confidentiality of the source is violated, particularly in the case of medical history. In response to the argument made by the American Historical Association (A. Jones, 2008) that research of this kind should be "explicitly exempted" from ethical review board oversight, the use of the Internet could make this position more difficult to maintain if the data were to become publicly available or "downstream," where there could be unanticipated risks to one's reputation, financial standing, or psychological well-being should identification occur.

The same was discussed about blogger content; generally speaking, text from blogs is regarded as publicly available, published material, so a researcher does not need permission to use it.

A substantial body of scholarly literature detailing IRE across disciplines and methodologies had emerged by the mid-2000s, thanks to the publication of three significant anthologies and an expanding body of literature. As a result, anecdotal evidence began to surface from the review boards assessing this research. To gather empirical information about the actual review board procedures of Internet research from the standpoint of human subjects, Buchanan and Ess polled over 700 ethics review boards in the United States. They discovered that the boards' main concerns were privacy, data security and confidentiality, and ensuring appropriate informed consent and recruitment procedures (Buchanan & Ess, 2009; Buchanan & Hvizdak, 2009). Post 2008, many organisations across the disciplines have predominantly included IRE as part of the research theme. Researchers from fields not typically associated with studying human subjects have started investigating IRE independently. For instance, computer security researchers have actively investigated the principles of research ethics in CS and ICT.

18.5 ETHICAL RESEARCH CONSIDERATIONS

Ethics in research involve principles of integrity, transparency, and respect for the participants. Researchers should ensure informed consent, minimise harm, and maintain confidentiality. Ethical boards should be established in institutions to ensure studies meet the desired standard of research.

Research ethics offers standards for the ethical and responsible conduct of research. In addition, the set of guidelines educates and monitors scientists' research to ensure a high ethical standard. The same can be applied to Ethics for Internet-based Research. Some of the common ethical elements that are applied are:

Principles of Integrity: Principles of integrity include honesty, accuracy, and reliability. Honesty is desired while collecting data, analysing results, determining appropriate methods and procedures, and determining publication status. Researchers should attempt to be truthful with their data and not fabricate, falsify, or misrepresent it. It is equally important to disclose any conflict of interest. Integrity in research not only builds trust but also increases credibility within the scientific community.

Objectivity: Aim to prevent bias in the design of the experiment, data analysis, interpretation, peer review, hiring decisions, and conclusion.

Carefulness: This is an essential part of good research. carefully and critically examine for plagiarism while presenting the data.

Transparency: Publication and presentation of the obtained data are essential parts of research. Openly sharing methods, data, and results fosters reproducibility. Transparency also allows scrutiny and promotes trust, helping to establish the credibility and reliability of scientific findings.

Respect for Intellectual Property: Honour for patents and copyrights, as well as intellectual property in the form of diagrams, figures, etc. The prime responsibility of the researcher is to attribute.

Confidentiality: Protecting the confidential details of the respondents is very essential. More care should be given to protect the personnel records of the respondents. Care should also be given to protect any other details that the respondent wishes to protect, or that should be protected legally. The researchers need to prioritise the protection of the privacy of individuals who are involved in the studies. Safeguarding the rights by respecting privacy is as important as obtaining informed consent from the participant.

Responsible Publication: Publication of the research is an important aspect of research. When submitting the research paper for publication, care should be taken that the same research paper is not submitted to another journal without withdrawing from the first.

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. Write briefly about the arrival of Internet-based Research.

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

2. List the Ethical Considerations in Research.

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

18.5.1 Challenges of Internet Research

In this technology-driven era, we will probably have discovered new technology by the time we finish reading, leading to a new set of thoughts on ethics. Some of the ethical concerns that bother the Internet Research Ethics are:

Privacy – a major concern: Privacy is crucial to Internet research ethics. Researchers delve into online spaces and often encounter vast amounts of personal information. A researcher needs to respect individuals' privacy rights and safeguard the confidentiality of participants with utmost care. Principles of ethics stress that the primary role of the investigator is to safeguard the privacy and confidentiality of the source. The degree of disclosure of private or sensitive information may bring an embarrassing situation to the respondent, and being high in Internet-mediated research, care must be considered. Ethical rules of research suggest that the investigators should take sufficient measures to safeguard the information gathered. Disclosure of information may include data that may be legally protected.

Social Media Platforms are considered an abundant source of information for researchers. For example, posts on Facebook and Instagram by its account holders can be accessed by the public, thus bringing a challenge to privacy. Since the posts made by the account holders are accessible to the public, the published content on these platforms by the account holders does not fit into the legal purview of "private information". However, a researcher gathering information related to the topic of his/her interest instead of casually taking the content as granted obtains permission to use the published material, the research data receives authenticity.

Recruitment of Participants: In any research, sampling techniques are used to identify the participants from the study population. The selection of the sample should be based on impartial and objective rules. Applying the same rules while drawing samples in Internet research, the researcher should select respondents methodically and not those whose information is easily accessible. It is always advisable to crosscheck for the authenticity of the account holder to rule out the chances of including fake accounts in the study.

It is important to remember that the selected participants could be pseudonyms online or have multiple or alternative identities. The researcher should be careful with fact-checking. After verifying a subject's age and demographic information, only suitable respondents should be included in the study.

While conducting research using the Internet, recruitment takes place, including push technologies, sending a message using text format, ex, message, tweet, and direct mail. In terms of ethics, the researcher needs to disclose the method of recruiting the participant. The research adds value if the researcher seeks permission from the website owners and group moderators before posting recruitment announcements, along with the permission of the moderators and the opening and closing date for receiving details.

In the end, when recruiting participants online, researchers must ensure that those interested receive information from the recruitment materials and any forms. Researchers also need to ensure that hiring practices don't result in someone being identified without that person's consent.

Gathering Informed Consent: Given the diverse and dynamic nature of online environments, obtaining explicit consent from participants is challenging. To minimise this challenge, researchers must transparently communicate the purpose of their study, the data collected, and how it will be used. The researchers should also provide participants with the autonomy to opt in or out of the study. Fostering a sense of trust between researchers and the online community will enhance the quality of research.

The consent of a participant in the research process is largely gathered in writing format. Informed consent indicates that participants are voluntarily participating in the research with their knowledge of relevant risks and benefits. Here, the researcher explains the purpose and details of the research. The researcher also provides sufficient time for the recipient to clearly understand and inquire about the research process. The participant provides consent by signing in a printed format, by mail, voice recording, and video recording.

In Internet-based research, the researcher may create a portal or site to complete the consent form. The participant can read the documentation and click on an "I agree" for submission; the data will also be password protected. The researchers can also use electronic signatures to obtain consent. In the case of research related to minors, parental consent is obtained. Conducting research related to minors indeed is a challenge in the case of online research. In addition to obtaining the consent letter, conducting demographic verifications related to the minor is a challenge. Obtaining required data from the minor without parental influence is another concern.

Data security in Cloud Computing: Internet users are becoming familiar with cloud computing, which enables data storage, networking, etc., since maintaining physical data can be time-consuming. Cloud computing allows researchers to access resources when a cloud service provider requires them. The special features of cloud computing include providing data on demand

and quickly, easy networking, and pooling of resources, which has made it popular. For researchers, gathering information is easy and cost-effective due to its features. Data is gathered from different online platforms like Google Docs or Microsoft Office 365, as well as online file storage and sharing services like Dropbox or Box.net. Data is also pulled from social media platforms like Facebook Developers Platform. Cloud computing services are accessed by researchers in all activities related to research, ranging from finding a suitable topic to research and data collection to processing and analysing data. This conjecture presents the ethical implications regarding data security with special reference to security.

The researchers need to be careful regarding the data that contains sensitive or personal information that is not accessible to others. It is imperative to ensure that the users providing labour through crowdsourcing cannot store copies of the research dataset. Acknowledging the contributors is a challenge among the other challenges involved in crowdsourcing. Particularly if they are anonymous, research must accurately and fairly describe authorship to be ethical.

The primary concern of the digital age is data security. Employing encryption, secure servers, and adhering to established data protection standards are imperative to maintain the integrity of Internet research. Since the vast repositories of information are available on the Internet, it is important to prevent unauthorised usage of potential data. In addition to data privacy, data stewardship appears to be a big challenge. Stewardship is responsible for the management of data or cloud services. The demand is for the proper use of data, its security, optimisation, and compliance with the data. The primary objective is to minimise the risk and maximise the benefits. Stewardship demands arriving at proper governance before cloud computing.

The researchers, particularly those who share datasheets on 'cloud' with their partners working from different geographical locations nationally or internationally for their collaborative project, face data privacy challenges in "cloud computing". The researchers who take up research projects should arrive at policies and procedures beforehand. The situation becomes more complex if the cloud provider has spread the data across several locations. The team should identify and assess the risks that surface. The best method to protect shared data is to encrypt the same.

Crowdsourcing is another area that raises ethical questions for researchers. It is the practice of obtaining desired services or content from a large group of people via the Internet. Crowdsourcing can assist researchers in all aspects of research in generating the desired results at a lower cost than traditional methods. Researchers need to develop new procedures to safeguard their data by combating the challenges of labour management and authorship.

Potential Impact on Participants: The research work undertaken may sometimes impact participants in one way or another. For example, research work on an online community. Care should be taken while disseminating information that may directly or indirectly harm a community. It is the responsibility of the researcher to minimise or nullify the harm to any community or participants.

Challenges posed by Big Data: ‘Big data-based research, which emerged in 2012, has been growing among researchers. In the context of big data research, privacy's definition and nature have become muddled, making it difficult to ensure privacy is respected and protected in this new field. Big data research has led to algorithmic processing; from "algorithmic harms" to "predictive analytics," the power of modern algorithms surpasses long-standing privacy norms and beliefs. Hence, big data-based research can face ethical complications. For instance, it becomes challenging to determine what qualifies as "private information" in the context of big data research, raising specific privacy concerns. For example, when users of social media platforms take no visible steps to restrict access, big data researchers may conclude that respondents are not particular about privacy.

Here, users frequently don't realise that their social media activity can be routinely tracked, collected, and shared with outside parties. The privacy policies are not fully understood by the members. It is also difficult to decode the user's intention when sharing the information on a social media platform. These uncertainties can also lead to conceptual muddles when a researcher attempts to reach conclusions.

Machine learning and Artificial Intelligence also add to the ethical concerns of Internet research. For example, Deep fake videos and dating apps. In such and other similar cases, ethical concerns have been raised about the very purpose of creating algorithms. A video or photo published in the public domain is invariably 'public' and will be accessed by people without consent. The challenges of technology-driven digital media have already led to concerns about AI ethics.

Check Your Progress: 3

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

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

1. Privacy is a major concern of ethics in Internet research. Justify

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

2. “Negligence of Social Media Subscribers has posed ethical challenges to a researcher”. Justify your answer.

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

3. True or False

- i. Online participants can be pseudonyms. (T/F)
- ii. Social Networking Sites have made it difficult to handle privacy in Internet research. (T/F)
- iii. Since the data is freely available a researcher need not gather informed consent in Internet Research. (T/F)

- iv. The best way to gather informed consent from a participant is to create a feedback portal. (T/F)
- v. Crowdsourcing does not involve ethical challenges of labour management and authorship. (T/F)

18.6 LET US SUM UP

In conclusion, the ethics of Internet research demand careful consideration and adherence to principles that safeguard participants' rights and well-being. Privacy, informed consent, data security, and mitigating potential harm should be at the forefront of researchers' minds as they navigate online research's expensive and complex world. Upholding these ethical standards ensures the credibility of research and promotes the responsible and constructive use of the Internet as a powerful tool for advancing knowledge.

We use the Internet extensively in research by visiting various sites for information. We also consider the usage or dependence of Internet human subjects as a topic for research, such as the gratification of social media and behavioural research. For undertaking research, we gather data using various techniques (such as data collection, mining, profiling, scraping, recording, monitoring, or observing already-existing data sets, chat room exchanges, blogs, social media posts, etc.). Search engines, data aggregators, digital archives, application programming interfaces (APIs), online survey platforms, and crowdsourcing platforms are some of the tools that make Internet research possible. Online research venues include online research venues for online research, online research venues for online multiplayer games, blogs, interactive websites, social networking platforms, instant messaging and discussion forums, and other conversation platforms.

Researchers looking to examine the dynamics of human interactions and their effects in this virtual medium find the Internet a prime target due to the abundance of social and behavioural data that can be found there. It may be possible for researchers to gather data from widely separated populations faster and at a lower cost than they could if they were to make comparable efforts in the real world. The number of Internet studies, ranging from surveys to naturalistic observation, has increased. This tendency also indicates that the desired information can be gathered without direct interaction with human subjects. Research works, either industry-based or academic, in the era when almost every individual has their presence on one or another social media platform, also provide the threat of intrusion into someone else's privacy.

The provision to process personal data available in the public domain due to ignorance or negligence of people who have the habit of sharing every minute detail of their lives may benefit some new acting talents, new creations, hobbies, etc. However, a majority of the public may feel surprised, outraged, or even harmed due to the intrusion of some stranger into their private life, such as deep fake videos/morphed photographs, etc. In all such instances, from the researcher's point of view, the account holder could be anonymous or hold a fake account, and content may lead to plagiarism challenges. To bring authenticity and credibility to the research work

undertaken, following, and practising the ethics of Internet Research is essential.

18.7 FURTHER READINGS

1. Buchanan, E. A., & Zimmer, M. (2012). Internet research ethics. <https://plato.stanford.edu/ENTRIES/ethics-internet-research/>
2. Convery, I., & Cox, D. (2012). A review of research ethics in Internet-based research. *Practitioner Research in Higher Education*, 6(1), 50-57. <https://insight.cumbria.ac.uk/id/eprint/1334/2/100-467-1-PB.pdf>
3. Cilliers, L., & Viljoen, K. (2021). A framework of ethical issues to consider when conducting Internet-based research. *South African Journal of Information Management*, 23(1), 1-9.
4. Ess, C. (2007). Internet research ethics. *The Oxford handbook of Internet psychology*, 487-502.
5. Ess, C. M. (2020). Internet research ethics and social media. *Handbook of research ethics and scientific integrity*, 283-303.
6. Hanganu, B., Manoilescu, I. S., & Ioan, B. G. (2021). Ethical challenges in online research. *Journal of Intercultural Management and Ethics*, 4(3), 17-24. <https://jime.cssem.org/index.php/JIME/article/download/192/155>
7. Tiidenberg, K. (2020). Research ethics, vulnerability, and trust on the Internet. *Second international handbook of Internet research*, 569-583.

18.8 CHECK YOUR PROGRESS: POSSIBLE ANSWERS

Check Your Progress: 1

1. Internet Research Ethics encompass principles and guidelines governing the conduct of research involving online data. It includes respecting privacy, obtaining informed consent, ensuring data accuracy, and avoiding harm to participants. Transparency in methods, protection of vulnerable populations, and acknowledgment of sources are integral. Adhering to ethical standards upholds integrity and trust in research outcomes.
2. Issues and concerns in Internet Research span privacy breaches, data security, and consent challenges. Ethical dilemmas arise regarding anonymity, data ownership, and the use of public versus private information. Digital divides, bias in algorithms, and the impact on marginalized communities raise ethical questions. Balancing research goals with ethical obligations amidst rapidly evolving technologies poses ongoing challenges for researchers. Addressing these concerns fosters responsible and impactful internet research.

Check Your progress: 2

1. The advent of Internet-based research revolutionized scholarly inquiry, facilitating unprecedented access to vast data repositories, and enabling global collaboration. Researchers leverage online platforms for surveys,

data collection, and analysis, enhancing efficiency and scope. However, challenges such as data security, validity, and representativeness require meticulous consideration to ensure robust findings. Additionally, navigating issues of digital ethics, including consent, privacy, and data ownership, is imperative to uphold research integrity and protect participants' rights.

2. Ethical considerations are paramount in research endeavours, guiding ethical conduct and safeguarding the welfare of participants. Key considerations include obtaining informed consent, ensuring confidentiality and anonymity, and minimizing harm or distress. Researchers must also uphold integrity in data collection, analysis, and reporting, avoiding plagiarism, falsification, or manipulation. Furthermore, equitable treatment of participants, respect for cultural diversity, and adherence to professional standards are essential ethical principles. Upholding these considerations fosters trust, credibility, and ethical integrity in research practice, contributing to the advancement of knowledge and societal well-being.

Check Your progress. III

1. Privacy is paramount in Internet research due to the potential for data exploitation and breaches. Individuals share vast amounts of personal information online, often unaware of the consequences. Research involving this data must uphold ethical standards to protect participants from harm, ensure informed consent, and maintain confidentiality. Failure to prioritize privacy can lead to breaches of trust, legal ramifications, and damage to individuals' reputation and well-being. Therefore, ethical considerations regarding privacy are crucial to safeguarding the rights and dignity of internet users, maintaining research integrity, and fostering trust between researchers and participants.
2. The negligence of social media subscribers introduces ethical dilemmas for researchers. Social media platforms are repositories of personal information, often shared willingly but sometimes without full awareness of the implications. Researchers must navigate complex ethical terrain when accessing and analysing this data, ensuring respect for user privacy, consent, and confidentiality. Neglecting these considerations can result in ethical breaches, such as unauthorized data use, exploitation, or harm to participants. Consequently, researchers face the challenge of balancing the potential insights from social media data with their responsibility to uphold ethical standards and safeguard the rights and well-being of subscribers.
3. True or False
 - i. True, ii. True, iii. False, iv. True, v. False