

<HTML>

<BODY>

<H1>INTERNET
RESEARCH
METHODS</H1>

<P>SECOND
EDITION</P>

CLAIRE HEWSON,
CARL VOGEL AND
DIANNA LAURENT

</BODY>

</HTML>



INTERNET RESEARCH METHODS



SAGE was founded in 1965 by Sara Miller McCune to support the dissemination of usable knowledge by publishing innovative and high-quality research and teaching content. Today, we publish more than 850 journals, including those of more than 300 learned societies, more than 800 new books per year, and a growing range of library products including archives, data, case studies, reports, and video. SAGE remains majority-owned by our founder, and after Sara's lifetime will become owned by a charitable trust that secures our continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC

<HTML>

<BODY>

<H1>INTERNET
RESEARCH
METHODS</H1>

<P>SECOND
EDITION</P>

CLAIRE HEWSON,
CARL VOGEL AND
DIANNA LAURENT

</BODY>

</HTML>



Los Angeles | London | New Delhi
Singapore | Washington DC



Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y 1SP

SAGE Publications Inc.
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
B 1/1 1 Mohan Cooperative Industrial Area
Mathura Road
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Editor: Mila Steele
Assistant editor: James Piper
Production editor: Ian Antcliff
Copyeditor: Neville Hankins
Proofreader: Christine Bitten
Indexer: Martin Hargreaves
Marketing manager: Sally Ransom
Cover design: Lisa Harper-Wells
Typeset by: C&M Digital (P) Ltd, Chennai, India
Printed and bound by CPI Group (UK) Ltd,
Croydon, CR0 4YY



© Claire Hewson, Carl Vogel and Dianna Laurent 2016

First edition published 2002. Reprinted 2003

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

Library of Congress Control Number: 2015943076

British Library Cataloguing in Publication data

A catalogue record for this book is available from the British Library

ISBN 978-1-4462-0855-7

ISBN 978-1-4462-0856-4 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the PREPS grading system. We undertake an annual audit to monitor our sustainability.

<> CONTENTS </>

<i>About the Authors</i>	vii
1 Introduction	1
2 The Internet	15
3 Internet-Mediated Research: State of the Art	33
4 Sampling in Internet-Mediated Research	71
5 Ethics in Internet-Mediated Research	99
6 Tools and Design Strategies for Internet-Mediated Research	117
7 What Can Go Wrong?	167
<i>Afterword</i>	193
<i>References</i>	197
<i>Index</i>	217

<> ABOUT THE AUTHORS </>

Claire Hewson is Lecturer in Psychology at The Open University. She has a long-standing interest in using the Internet to carry out primary research, and has collected data using a range of IMR methods including surveys, psychometrics and experiments, to investigate issues in common-sense understanding, particularly folk psychology; online assessment methods in an educational context; and the validity of IMR methods. She was commissioned to lead a working party to produce the recent British Psychological Society (BPS) guidelines on ethics in Internet-Mediated Research (2013), has delivered a number of talks and training sessions on this and related topics, and has published more broadly in the area of IMR.

Carl Vogel, a Fellow of Trinity College Dublin, is Senior Lecturer in Computational Linguistics and Director of the Centre for Computing and Language Studies at Trinity College Dublin, the University of Dublin. His work in computational linguistics, from the perspective of cognitive science, frequently draws upon evidence abstracted from Internet accessed data, and accordingly he dwells on the accompanying research methodology issues. Vogel was on the Management Committee of the successful European COST Action IS1004: Web-based data-collection – methodological challenges, solutions and implementations (WEBDATANET).

Dianna Laurent teaches a variety of classes for both the English and the Languages and Communication departments at Southeastern Louisiana University and at St. Joseph's Seminary College. She publishes on a variety of subjects involving the Internet. Dr Laurent is the associate editor of the yearly ATTW bibliography and the business manager of *19th Century Studies* for the Nineteenth Century Studies Association.

<> ONE </>

<> INTRODUCTION </>

INTERNET RESEARCH METHODS: THE SECOND EDITION

It is now just over a decade since the publication of the first edition of *Internet Research Methods (IRM)*. There we reflected and speculated on the actual and imagined scope of the Internet for supporting and enhancing social and behavioural research. At that time, several early pioneers had been making headway in devising and testing out what we dubbed *Internet-Mediated Research (IMR)* methods; that is, procedures for collecting primary research data which made use of the Internet. We distinguished between *primary* and *secondary* research online. The former involves the acquisition and analysis of data to produce novel evidence and research findings;¹ the latter involves research which utilises secondary information sources (such as books and journal articles) to summarise existing findings and conclusions. We presented some early examples of IMR, as well as selected case studies from our own research, to illustrate some of the techniques, procedures and opportunities available. We outlined some of the caveats that can emerge, drawing upon our own experiences as researchers, and highlighted and offered our own perspective on issues and debates that were prevalent at the time (such as the biased nature of Internet-accessed samples).

Over the last 10 years IMR has expanded massively, increasing in both reach, across disciplinary boundaries, methodological orientations and research domains,

¹Our definition at the time was rather narrow in focus, referring to primary research as that which gathers data from participants. The vast array of existing traces of activity online which may now form data for primary research, and the emergence of methods which make use of these traces, require us to revise this emphasis. Unobtrusive approaches which do not directly recruit participants, but gather non-reactive data, form a key strategy in present-day IMR.

and volume. There now exists a large, diverse body of evidence from IMR studies upon which to draw in assessing its impact to date, and its scope and future potential. New methods, data sources and strategies have emerged. A major and significant revision to the present volume is a consideration of this new evidence, which has now informed many of the issues and debates we outlined in the first edition. We now find it necessary to expand the methodological scope of our discussion in order to reflect properly the present-day nature of IMR. Thus in this second edition we include a more extensive discussion of unobtrusive approaches in IMR; for example, those which take advantage of the wealth of potential data sources created by the emergence of social networking sites (SNSs). We also consider unobtrusive approaches which use advanced 'data mining' techniques to obtain and process very large volumes of information about the structures and processes of people's online activities and interactions ('big data').

Ethical issues in IMR are also given more attention in this second edition. Given the extensive discussions on a range of ethical issues that have occurred over the last decade, it now seems fitting to devote an entire chapter to this topic. Rapid technological developments have also occurred, some of which have confirmed the predictions we made previously. We highlight these developments. A major development has been the emergence of dedicated software tools for implementing IMR study designs. These tools have made the need for presentation of detailed programming techniques, such as those we offered in the first edition, no longer necessary, although we mention some of the main current technologies available for implementing more complex, bespoke designs. Another important development involves the shifting patterns of access to and usage of the Internet, creating changes in the way Internet users can be recruited to participate in IMR studies, as well as who is available. The range of recruitment procedures possible, and the research on the effectiveness of these, are now considered in a dedicated chapter on sampling in IMR. All the above developments, as well as the emergence of Web 2.0 (see below) which has arguably created a qualitative shift in the nature and societal impact of the Internet, have led to what may be called a 'new era' of IMR. It is the purpose of this second edition to provide an updated review and evaluation of IMR methods in this new era. As before, the text is a handbook which outlines and discusses key theoretical issues and debates, methodological principles, and detailed implementation procedures which will assist the readers in creating and launching their own IMR study.

As in the first edition, we draw extensively on our own experience in using the Internet as a research tool. We have gathered primary data to address questions on human reasoning (e.g. Hewson & Vogel, 1994), common-sense beliefs and understandings (e.g. Hewson, 1994), use of social signals, such as emoticons, in online communication (e.g. Janssen & Vogel, 2008; Vogel & Janssen, 2009; Vogel & Mamani Sanchez, 2012) and online assessment methods in an educational context (e.g. Hewson, 2012a; Hewson, Charlton, & Brosnan, 2007). We have also implemented studies to validate IMR instruments and procedures (e.g. Hewson & Charlton, 2005),

and been engaged in the development of software procedures and implementation of IMR systems (Buckley, 2004; Buckley & Vogel, 2003; Graham, 2006), and ethics guidelines (BPS, 2013). We offer showcase examples from research studies, including our own, throughout the book (replacing the previous edition's 'case studies' chapter).

Who the Book is Aimed at

Active researchers and students alike may find the Internet useful for both primary and secondary research. Thus we have aimed to make the book accessible and of interest to both. Using the Internet to locate secondary resources can have great pedagogical value and, we argue, enhance (but not necessarily replace) more traditional library-based methods. For those involved in research and teaching, the wealth of information available online – from databases of journal articles to copies of lecture notes – is invaluable. These resources can help locate information quickly and cost-effectively. The important issue to bear in mind is establishing the quality and accuracy of the resources found, and in Chapter 2 the book provides guidelines for ensuring this. Primary research using the Internet may benefit students who are undertaking an undergraduate project, or conducting postgraduate research, due to the scope for obtaining large volumes of data in a short time and with minimal costs. Researchers at smaller institutions, where the resources available (time and funding) for supporting research may be more limited, may similarly benefit. However, the value of IMR is not limited to these contexts, and as we shall see the last decade has seen a dramatic increase in the volume and range of IMR methods being implemented.

While our own disciplinary commitments will inevitably lead to some degree of bias towards examples of IMR from psychology and cognitive science, we have taken care in this second edition to include a broad range of illustrations from other disciplines and areas which fall within the general category of the social and behavioural sciences.² The methodologies we discuss (which include surveys and questionnaires, interviews and focus groups, observational studies, document analysis and experiments) are certainly widely used across disciplines. Certain disciplines, or sub-areas within those disciplines, may of course be associated with particular methods (e.g. cognitive psychology draws heavily on experiments), but the information presented here should be broad enough in scope for researchers from different disciplines to pick out what is relevant to their own particular research domain. Our aim has been to select a good range of illustrations to demonstrate the diversity and scope of IMR methods. Earlier chapters explore the theoretical, methodological and ethical issues which IMR raises, and these are then taken up

²Which we consider to include (among others) psychology, linguistics, sociology, economics, political science, cognitive science, anthropology.

in more detail in later chapters, which outline how the issues interact with specific procedures and implementations. The scope of IMR is discussed early on, through consideration of the range of methodologies that can be and have been adapted, looking at the successes to date, and considering the advantages and disadvantages of using IMR as opposed to more traditional approaches. Thus the readers can get an impression at this point of the extent to which their own research might be supported by IMR methods, before reading more about implementation details and resources in later chapters. Potential problems are also highlighted so that the researcher may be forewarned of these, and take steps to avoid them. In the unfortunate event that problems do occur (even in a carefully designed study unforeseen problems can nevertheless arise), recovery strategies are suggested.

A further important aspect of the book is that it does not assume the reader has any prior computing expertise, at least not beyond some basic skills such as sending and receiving emails, reading and posting to discussion boards, and searching the World Wide Web.³ Thus the book is aimed at providing the practising researcher, or student, who has some minimal level of computer literacy, with the necessary information, tools and insights to be able to assess the extent to which the Internet can help support their research needs, and to be able to carry out Internet-based research.

We now conclude this introduction by giving a brief outline of the history of the Internet, followed by a description of the content of each chapter.

HISTORY OF THE INTERNET

The Internet grew out of the ARPANET, commissioned in 1969 by the US Department of Defense for research into computer networking (for a more detailed history of the Internet see Zakon, 2015). In 1971 there were 15 nodes on ARPANET connecting 23 host computers. Email was invented in 1972 by Ray Tomlinson of Bolt, Beranek and Newman (BBN). The first international connections did not come until 1973 when England and Norway each added nodes. BBN introduced the first commercial version of ARPANET in 1974. From then on a number of network systems emerged, including Usenet in 1979 and BITNET in 1981. The latter started as a cooperative email-based system between CUNY⁴ and Yale. The first MUD⁵ was produced in 1979 by Richard Bartle and Rob Trubshaw of the University of Essex.

³These procedures are very easy to get to grips with, for readers who are not already familiar with them. Any introductory guide to the Internet should provide instructions (e.g. Buckley & Clark, 2009; Kennedy, 2001). See also <http://www.archives.gov/research/alic/reference/internet-users-guide.html> (accessed April 2015).

⁴The City University of New York.

⁵Multi-User Dungeon (cf. 'MUD, object oriented' or MOO).

Protocols, namely the Transmission Control Protocol (TCP) and Internet Protocol (IP), were introduced in 1982. Effectively, the 'Internet' began with that standardisation and denotes networked TCP/IP systems.

It was not until 1984 that the number of networked host computers exceeded 1,000, and in that same year the Japan Unix Network was put into place, as was the Joint Academic Network (JANET) in the United Kingdom. In 1986 the NSFNET was established by NASA and the US Department of Energy as a way to facilitate connections outside the ARPANET security and bureaucracy. By 1987 the number of Internet hosts exceeded 10,000, and BITNET hit the 1,000 point. Just a year later there were 60,000 Internet hosts, this number increasing to over 100,000 within the next year. By 1992 the number of Internet hosts stood at over 1 million, increasing to 3.2 million by July 1994 and reaching a figure of 56,218 million networked hosts by July 1999, five years later. In July 2014, the Internet Systems Consortium (ISC) counted over 1 billion IP addresses that have been claimed with domain names, which can be taken as a reasonable estimate of the number of Internet hosts (<http://ftp.isc.org/www/survey/reports/current/> [accessed October 2014]). A plot of the expansion in use of available domain names (roughly corresponding to the number of Internet hosts) is given in Figure 1.1; measurements were made in January and July in each year except 2003. Growth between the January and July measurements appears to have slowed. The Oxford Internet Survey report (Dutton, Blank, & Groselj, 2013), published every two years since 2003, indicates that 78% of the UK population over the age of 14 has Internet access. This growth in number of hosts is mindboggling, particularly when considering that the individual hosts can serve many more individual users. While writing the first edition of *IRM*, we noted that the terms 'World Wide Web (WWW)' and 'Internet' had become household words, and that every day more and more people were accessing the Internet through academic, private, military, government and commercial interests, primarily through networked computer systems. Now, a decade on, some significant developments in Internet technologies and patterns of usage have taken place.

At a most basic level, the sheer estimated size of the Internet-user population (IUP) has grown from around several hundred million in 2003 (Hewson, Yule, Laurent, & Vogel, 2003) to over 2 billion in 2013 (see Chapter 4 for further details and sources), dramatically expanding IMR sampling possibilities. A particularly noteworthy development has been the emergence of social networking technologies and services (e.g. Facebook⁶ and Twitter⁷). Another has been the

⁶Facebook (2011) reports more than 800 million active users, 50% of whom log on to Facebook in any given day (<http://www.facebook.com/press/info.php?statistic>; accessed November 2015). Recent estimates put the number of active users at close to 1.5 billion (see Chapter 4). However, this figure can only be an upper bound, since it is not difficult to find people who profess to have more than one Facebook account and access the system from more than one IP address.

⁷For a (journalistic) comparison of Facebook and Twitter usage, see <http://www.guardian.co.uk/media/2011/nov/07/twitter-facebook> (accessed April 2015).

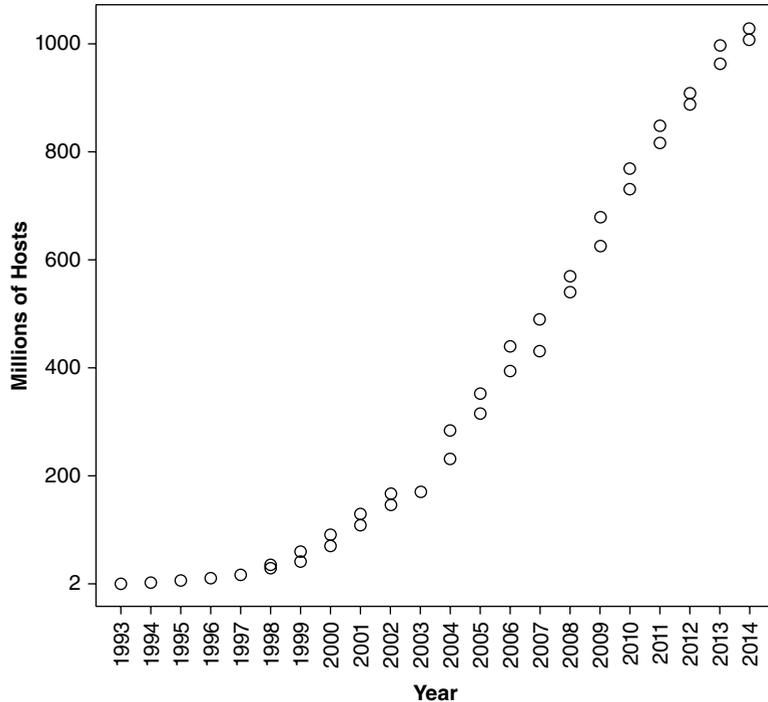


Figure 1.1 Yearly growth in Internet domain name use

Note: Measurements taken in January and July of each year surveyed.

Data Source: Internet Domain Survey of the Internet Systems Consortium.

emergence and proliferation of mobile technologies (e.g. smart phones, tablets) and ‘apps’ (short for ‘applications’) that reside on them. These technologies and services are described and discussed in more detail in later chapters, where their relevance to particular IMR methods is considered. Hand in hand with these developments, the penetration of the Internet and its services into daily lives has expanded on a vast scale; we discuss these developments and their impact upon IMR methods in later chapters (particularly Chapters 3 and 4). *Web 2.0* is a term (O’Reilly, 2005)⁸ which has emerged to describe some of these developments, particularly the shift from the Internet as a relatively static space for the dissemination of published final documents to a massively connected, fluid, interactive, participatory, collaborative space where content is constantly

⁸Though the term was first coined by Darcy DiNucci back in January 1999 (DiNucci, 1999). She describes the notion as follows: ‘The Web we know now, which loads into a browser window in essentially static screenfuls, is only an embryo of the Web to come. The first glimmerings of Web 2.0 are beginning to appear, and we are just starting to see how that embryo might develop. The Web will be understood not as screenfuls of text and graphics but as a transport mechanism, the ether through which interactivity happens.’

in flux and under revision. Wikipedia – itself an example of a Web 2.0 service⁹ – offered the following description in December 2011:

The term Web 2.0 is associated with web applications that facilitate participatory information sharing, interoperability, user-centred design, and collaboration on the World Wide Web. A Web 2.0 site allows users to interact and collaborate with each other in a social media dialogue as creators (prosumers) of user-generated content in a virtual community, in contrast to websites where users (consumers) are limited to the passive viewing of content that was created for them. Examples of Web 2.0 include social networking sites, blogs, wikis, video sharing sites, hosted services, web applications, mashups and folksonomies. (http://en.wikipedia.org/wiki/Web_2.0 [accessed December 2011])

In March 2015, this Wikipedia entry started as follows:

Web 2.0 describes World Wide Web sites that emphasize user-generated content, usability, and interoperability. The term was popularized by Tim O'Reilly and Dale Dougherty at the O'Reilly Media Web 2.0 Conference in late 2004, though it was first coined by Darcy DiNucci in 1999. Although Web 2.0 suggests a new version of the World Wide Web, it does not refer to an update to any technical specification, but rather to cumulative changes in the way Web pages are made and used. (http://en.wikipedia.org/wiki/Web_2.0 [accessed March 2015])

These two descriptions demonstrate the way Wikipedia entries evolve and change as various users edit, update and add to them (which anyone can do, having first created an account; clicking on the 'View history' tab offers a detailed record of the edits that have been made to an entry, searchable by year and month). There has been some debate about whether the emergence of 'Web 2.0' represents a *qualitative* shift in the nature, form and use of the Internet, as opposed to merely an expansion in size and diversity.¹⁰ Either way, the technological developments we have seen in more recent years, as embodied in concepts such as Web 2.0, have no doubt been a significant instrumental factor in the expansion and increasingly widespread use of the Internet for conducting primary research across a very large number of disciplines and research perspectives. As noted by Lee, Fielding, and Blank (2008, p. 4) the terrain occupied by IMR methods is by now large and variegated, as well as complex and rapidly changing, with arguably particular importance to the social sciences. Coverage of some of these newly emerging approaches and

⁹As opposed to, for example, 'Encyclopaedia Britannica Online' which is a Web 1.0 model of knowledge management (Wakeford & Cohen, 2008).

¹⁰For a challenge to the idea that Web 2.0 is something fundamentally distinct from earlier notions and structures of the WWW, see the developerWorks interview with Tim Berners-Lee, on 22 August 2006, available from <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html> (accessed April 2015).

OVERVIEW OF CHAPTERS

Chapter 2: The Internet

The focus of this chapter is firmly on secondary research, that is the gathering of information that is already constructed, though some of the Internet resources mentioned may also be useful in primary research. This chapter outlines how secondary research can be carried out using the Internet.¹¹ We present an updated description of the resources available to conduct a literature search of secondary resources for research purposes (e.g. discussion groups, library catalogues, newspapers, indices to periodical literature, art archives, government official databases, electronic texts, online newspapers, Google Scholar, social media and networking sites) and explain how best to locate and use these resources. We emphasise both the time that can be saved, as compared with traditional (library-based) methods, and the additional information that can be accessed. However, the chapter also stresses the importance of assessing the relevance, quality and reliability of these resources. The amount of information available is limitless, and the researcher (and student) should be especially careful in ensuring that sources drawn upon are of high quality. Thus emphasis in this chapter is on providing guidelines for how to use the Internet effectively, and how to recognise bogus or limited sites.

The secondary information sources described provide a wealth of information relevant to many disciplines within the human sciences. The chapter explains how to use these resources, without assuming any prior familiarity. Yet the chapter also provides valuable information for the user who may already have a basic familiarity with these resources, by providing recommendations for good practice, which can enhance efficiency and quality of materials gathered. Many readers will be familiar with using search engines to seek specific information, but will also be aware of the frustrations involved in having to trawl through volumes of useless information to find what is required. There are a number of alternative and extremely useful access points to information available, with which many researchers may be unfamiliar. These are described, with an emphasis on selecting those sites which are unlikely to go out of date soon. In summary, Chapter 2 gives a comprehensive introduction to the range and depth of information available on the Internet (with a focus on that relevant to social and behavioural research), and how to access and test the quality of this information.

Chapter 3: Internet-Mediated Research: State of the Art

This chapter presents a review and evaluation of the range and scope of IMR methods, considering what has been tried and tested, and outlining and assessing the successes as well as less successful applications and techniques to date. From early

¹¹The reader is referred also to Ó Dochartaigh (2012) which is complementary to the current book, though at a more introductory (undergraduate) level.

pioneering examples, through to recent emerging methods and techniques, illustrations selected to demonstrate the range and penetration of IMR methods are presented and discussed. Thus the chapter aims to provide an up-to-date account of the state of the art in IMR, highlighting key issues, debates and theoretical underpinnings of the various methods and strategies which have emerged. Both obtrusive and unobtrusive approaches are considered, and a framework for classifying different IMR methods along the *obtrusive–unobtrusive* dimension is offered. The blurred boundaries which can sometimes emerge between different methodological approaches in IMR, due to the idiosyncratic nature of some of the supporting technologies and Internet structures, are explained and illustrated. The scope for adapting traditional offline methods to an IMR medium, as well as the potential for facilitating new, creative approaches and techniques, are considered.

Throughout Chapter 3, a focus on the extent to which Internet methodologies are able to produce valid and reliable data is maintained. Advantages and disadvantages of Internet-based implementations are considered, compared with traditional methods, as well as with each other. Some of the features of Internet-based primary research are access to larger and more diverse samples, reduced costs and reduced timescale. In addition, the Internet allows entirely anonymous communication, but with higher levels of interactivity than has previously been possible. Further, IMR opens up new possibilities for unobtrusive observational research. These and other features of IMR are discussed. Caveats are also addressed. A major concern is the reduced level of researcher control (over participants, materials and procedures) which IMR procedures often entail. Issues relating to current technological limitations, as well as demands for technological expertise of both researcher and participant, are also discussed. The ways in which all these issues can impact upon IMR are considered. Solutions are offered, though these await more detailed elaboration (in terms of the tools and procedures available) in subsequent chapters. Overall, the chapter provides a basic introduction to Internet-mediated primary research, outlining the scope, possibilities and issues which arise.

Chapter 4: Sampling in Internet-Mediated Research

In this chapter we consider the range of sampling possibilities in IMR. Sampling bias is an issue given some consideration in this chapter, though this is less of a concern than in the early days of IMR, for reasons outlined there. The chapter spends some time exploring the types of sampling strategies which are available in IMR, and highlights how these interact with the types of samples which can be obtained. Research is reviewed which compares Internet-accessed and traditional samples in terms of sample characteristics (e.g. representativeness of a broader population) and the quality of the data they produce. Recommendations for procedures which maximise data quality are given. Advice is offered on some of the tools and resources available, which can help to assist the researcher in accessing Internet samples. Overall, the chapter concludes that there is scope for obtaining high-quality data from Internet-accessed samples, including emerging

possibilities for obtaining data which are more broadly generalisable than has previously been considered possible (e.g. using large-scale online probability panels).

Chapter 5: Ethics in Internet-Mediated Research

This chapter considers the key ethical issues and dilemmas which emerge when porting research methodologies to an IMR context. The chapter covers guidelines for ethical behaviour by researchers from a wide range of disciplines. We draw upon a range of examples to highlight and demonstrate the nature of the ethical dilemmas that the Internet introduces as compared with traditional methods of data gathering. Key issues discussed are: informed consent, withdrawal and debrief; participant confidentiality; data anonymity; data security; the distinction between public and private spaces online. A number of caveats are outlined, and solutions offered, in adhering to key ethics principles in IMR. The chapter emphasises that ethics guidelines for IMR need to be flexible, and adaptable, as new technologies and research findings emerge. Ultimately, a researcher will need to make context-specific decisions and choices, taking into account a range of relevant factors in any particular research setting.

Chapter 6: Tools and Design Strategies for Internet-Mediated Research

This chapter outlines some of the key Internet technologies and tools which are available to support implementing IMR studies (some of which have been mentioned in Chapters 2 and 3). The key methodological categories outlined in Chapter 3 are revisited in order to offer a more detailed consideration of implementation details. Discussion of design issues, choices and good-practice principles is incorporated. The aim is to provide the moderately computer-literate researcher with knowledge of the resources and tools available to enable IMR study design and implementation without detailed programming knowledge or expertise; thus emphasis is on reviewing some of the software tools available to assist in this process.

Web-based approaches are particularly important for many IMR methods, and we consider some of the available software packages which can help construct and disseminate web-based experiments and surveys. Web-based surveys in particular are given a more detailed consideration, in the final section, since they are in particular a widely used, attractive option, relevant to many research domains and disciplines. That section outlines software tools and design principles for web-based surveys, with illustrated implementation examples. This shows the reader how to get a web-based survey up and running. Other tools and resources considered include: email and mailing lists; discussion groups; online chat; social networking sites; multimedia sharing sites; online documents. Despite the more recently emerging large range of off-the-shelf software packages to support web-based surveys (and some to support experiments), more complex and bespoke implementations will sometimes be necessary and in this chapter we highlight some of the main current technologies available for implementing more complex, bespoke designs. Such techniques can

create very flexible and powerful systems with functionalities which may not be supported by many existing software solutions (though the more advanced and open-source solutions often allow bespoke editing options, also). Various issues are covered which are essential to producing well-controlled and robust designs which can lead to the generation of reliable, valid data (e.g. tracking participants, maintaining consistency over display parameters, and so on).

The chapter emphasises the need for simple, robust designs if an implementation is to be widely accessible. By following the guidelines in this chapter the reader should (presuming access to the necessary tools) be able to produce a well-designed survey which can be placed on the Web, and which incorporates some of the principles of good practice emphasised throughout this book.

Chapter 7: What Can Go Wrong?

This chapter revisits many of the issues from earlier chapters, particularly issues of sampling from Chapter 4, ethics from Chapter 5, and design and implementation from Chapter 6. Ways that things can go wrong in implementing Internet studies, including issues relating to methodology, etiquette and vandalism, are all considered. Hardware requirements and software tools and procedures are considered. One of the main issues at stake involves selecting the right level of technology for the research question at hand. The more technically involved the study, the greater will be the demands on a researcher's level of technological expertise in order to be able actually to implement it in a properly controlled fashion (particularly with more complex experimental designs). Also, as technological demands increase, the population of potential participants equipped to take part in the research may be restricted. We point out some common pitfalls that can occur, and suggest solutions. We consider the issue of hackers, whether random hackers or those moving under the mask of bona fide participants. We discuss issues of data protection, and make some suggestions about file location, password protection and user accounts generally. We stress throughout maximal researcher control over materials and data. We do not suggest that researchers should become competent programmers, however. We rather promote the alternative strategy as discussed in Chapter 6, involving using third-party software and servers provided for non-programmers to construct and conduct studies over the Internet. Nonetheless, some researchers will be sufficiently competent at programming to develop their own systems or adapt existing open-source systems to their purposes, and we hope the guidelines that we suggest will help them avoid risks as well. Even those who do not program their own systems will do well to ask questions about how the third-party software that they contemplate using addresses the issues that we raise.

Afterword

Here we summarise the key points raised throughout the book, and offer a final comment on the current state of the art in IMR, and our projection for likely future developments.

<> TWO </>

<> THE INTERNET </>

We are buried beneath the weight of information, which is being confused with knowledge: quantity is being confused with abundance and wealth with happiness. We are monkeys with money and guns. (Tom Waits, in Maher, 2011: 444)

INTRODUCTION

Research on the Internet begins like any other research. Possible research areas are explored until a final topic is settled; the reliability of sources is judged; solid information is gathered; a working bibliography is framed; and databases at research centres are accessed. Using traditional research methods, it is possible to spend countless hours in the library, interviewing sources, and combing through library catalogues, journals and indices. In fact, most student research starts on the Internet with a Google search, a check of the university library webpage, and in some cases the perusal of a wiki, blog or social network. Using the Internet to access information is by far the easier and more efficient method.

Thousands of library catalogues, including all major university research libraries, are easily accessible online. Searchable online databases of thousands of journals and tables of content are available. Online catalogues allow a search by author, title, subject or keyword to facilitate the compilation of a working bibliography and, indeed, even offer automatically downloadable citations for sources, in much less time than is required for traditional methods.

With all research, it is important to learn how to find information without wasting time. The best way to achieve this goal is to learn how to use the Internet search and discovery tools and learn to evaluate the findings in terms of reliability, quality and relevance.