Issue 61: Spring 2011

social research Update

Internet Research and Unobtrusive Methods

Christine Hine

Christine Hine is Senior Lecturer in the Department of Sociology at the University of Surrey. Her main research centres on the sociology of science and technology, including ethnographic studies of scientific culture, information technology and the Internet. She has a long-standing interest in promoting discussion of methodologies for sociological understanding of the Internet.

<c.hine@surrey.ac.uk>

- **The Internet offers social researchers unrivalled access to the minutiae of daily life**
- **Using data from websites, forums and social networking sites continues a long tradition of unobtrusive methods in social research**
- **Unobtrusively acquired online data can make otherwise ephemeral aspects of everyday life amenable to research**
- **Unobtrusive Internet research can reduce the burden placed on those whose behaviour is being researched**
- **Cautionary notes include the need to establish ethical grounds for accessing online data, and the possibly of various forms of bias shaping the data**

Research methods which do not require active participation from those being researched have a respectable, if not always prominent, role in the history of social research. The most notable recent discussion of their potential is Lee's (2000) work on unobtrusive methods. Lee builds on Webb et al (1981) and Kellahear (1993) to explore advantages and ethical and epistemological challenges which these non-reactive approaches offer. In reactive methods such as interviews and surveys, respondents are aware of the researcher and may respond in socially desirable fashion or adapt their behaviour in consciousness of being under scrutiny. Unobtrusive use of "found" data can expose biases in data collected via reactive methods, or allow an otherwise hidden population or practice to be explored.

In his final chapter Lee (2000) explores the opportunity which the Internet offers social researchers interested in analysing "found" data. Since his book was written these opportunities have multiplied as the Internet has become an increasingly mainstream phenomenon (in 70% of UK households in 2009 (Dutton et al. 2010)). More people, and a more diverse range of people, are now online, doing a wider array of things, including participating in discussion forums and building web sites as they

were in 2000, but also using social networking sites, uploading their photographs and videos, leaving their opinions via tagging, commenting and reviewing and leaving electronic traces of their actions in logs of server activity, search engine usage and the like. As Savage and Burrows (2007) describe it, this wealth of data on social activity poses a significant alternative (and potential threat) to the traditional techniques social researchers use to collect their data. This paper reviews some recent studies which make use of found Internet data before moving on to summarise advantages and cautionary notes.

Recent unobtrusive studies on the Internet

Seale et al (2010) compare data from Internet discussion forums with face-to-face interviews in the area of health research. They suggest that whilst previously we often conducted interviews because direct observation was too difficult and time consuming, the Internet has now made acquiring observational data less laborious than interviews. They illustrate the research advantages that accrue from unobtrusive online research using the example of keyword analyses of the frank conversations which take place in online health advice forums under the cover of anonymity. Similarly, Harvey et al (2007) carry out a study on email messages sent to a health advice site aimed at young people, using approaches from corpus linguistics. They too find that the frankness of online discussions on sexual health contrasts strongly with data derived from non-obtrusive methods. Thelwall's (2008) research into contemporary swearing practices uses data derived from publicly available MySpace members' home pages, comparing users of different age, gender and nationality. Twitter messages have been used to analyse people's responses to terrorist incidents (Cheong and Lee 2011). In each case the researcher is able to access an aspect of social existence not readily uncovered in an interview setting.

Social network analysis and hyperlink analysis have deployed online data collected unobtrusively to interesting effect (Garton et al. 1997; Park and Thelwall 2003). Golder et al (2007) use data from the headers of messages exchanged by Facebook users to examine the weekly working rhythms of this largely student population. Beaulieu has discussed the challenges and opportunities of integrating various forms of online traces into ethnographic enquiries (Beaulieu 2005; Beaulieu and Simakova 2006). Dirksen et al (2010) discuss integration of social network analysis of log file data into an ethnographic approach. Rogers has proposed a programme of digital methods for researching culture, deploying a wide range of natively digital traces (Rogers 2010).

Google Trends offers possibilities of exploring changing health information-seeking behaviours (Ellery et al. 2008). Indeed, many Internet activities generate log files which can be used as proxies for social behaviour. Just as Webb et al (1981) suggest that noseprints on the glass of museum cabinets could act as proxies for the interests of visitors, so too can log files of website activity be used to stand in for what people are interested in on the Internet (McLaughlin et al. 1999). Such analysis is commonplace in commercial web activities, where marketers want to know how visitors spend their time on a site and what influences them to click on online advertisements. Whilst some analysis of logfiles and hyperlinks has to be a specialist activity using bespoke approaches, there are also ways of using ready-to-hand tools on the Internet to carry out sociological analysis of found data. Visualization tools such as Touchgraph (http://www.touchgraph.com) can be used to explore how users construct the web landscape (Hine 2007).

Unobtrusive methods can also be used to gather data for a qualitative exploration, using "found" data on the Internet to explore people's understanding of a topic or cultural phenomenon. In order to study responses to the television programme *Antiques Roadshow* I used Google and the YouTube search facility to find references to the show (Hine 2011). By qualitative thematic analysis of the resulting diverse array of mentions of the programme I was able to formulate an idea of how people used it as a cultural resource. I found that the everyday mentions of the show deviated from the "preferred readings" offered by a purely textual analysis of the show and also from the intense discussions of committed fans in dedicated forums. My Google and YouTube based search strategy uncovered a more casual form of engagement with the show.

Using the Internet for qualitative exploration of responses to cultural phenomena echoes some non-Internet unobtrusive approaches such as Dahlgren's (1988) study of casual mentions of news in everyday interactions, Press and Johnson-Yale's (2008) study of women talking about Oprah in a hair salon, or Levine's (2007) use of photographs and high school yearbooks to explore engagements with television in everyday life in the past. In each of these cases the researcher works creatively to find a way of rendering everyday talk available for study, in order to avoid the contrived situation of an interviewer asking people direct questions about their response to the object of the investigation.

Advantages

Since almost every conceivable aspect of daily life is reflected somewhere online, and since the cloak of anonymity can lead people to a frankness they rarely show in face-to-face encounters, the Internet

offers rich data for almost any social researcher and particularly those working in sensitive areas. The Internet can also offer, via topic-specific forums, a naturally occurring form of focus group which has the added advantage of being ready-transcribed.

Unobtrusive collection of Internet-derived data can be less labour intensive not only for the researcher but also for the researched. Where research focuses on sensitive areas, or particularly vulnerable groups, it can be a considerable imposition to ask respondents to recite their situation for the researchers' benefit. Unobtrusive methods using Internet-derived data make use of what people have already said and done. In addition, the Internet can make previously ephemeral and hard to record aspects of daily life into researchable phenomena.

Cautionary notes

Simply because we can access data does not mean it is ethically available for research purposes. Careful evaluation needs to be made of the extent to which particular research techniques make unwarranted intrusions or may have undesirable effects on those studied. Content analysis and other quantitative techniques which summarise insights across populations will often be of less ethical concern than in-depth qualitative studies focused more on the individual.

A caveat to bear in mind while analysing Internet-derived data as if it informs understanding of a broader social phenomenon is that not all of the population have access to the Internet, that access is segmented according to socio-demographic characteristics such as nationality, age, gender, education, ethnicity and income, and of those that do use the Internet, many will be passive consumers of Internet information rather than active participants in the "web 2.0" environment. The data set will therefore be subject to participation biases that may be very difficult to identify.

The ability to locate relevant Internet data may also be subject to significant but hard to define biases (Wouters and Gerbec 2003). Search engines do not index the entire Internet, and their ordering of results depends on proprietary algorithms. Using ready-to-hand tools such as search engines can therefore lead to a biased portrayal of whatever actually is out there on the Internet. Similarly, data made available in archives (whether archives of specific forums, or larger resources such as Google Trends) can be subject to choices over which the researcher has no control or may be withdrawn at short notice (Gaffney et al. 2011). Data which is publicly available on social networking sites may only be accessible if users have set their privacy controls to allow public access: this portion of the online population may be more naïve about privacy controls, or more interested in publicity, than the general online population.

A further key concern about the unobtrusive use of Internet-derived data is that we lack access to how Internet users might interpret and make use of online information or what, ultimately, their browsing, hyperlinking and social networking mean to them. This lack of information about the consumption of online interactions, leaves the researcher, as Lee (2000) suggests, often relying on unobtrusive methods as part of a combination of strategies rather than able to use them as a stand-alone method for understanding a phenomenon.

Conclusion

In summary, with due methodological caution a wide range of unobtrusive studies using Internet-derived data are possible. Indeed, such is the richness of the portrayal of social life that we find online that it would seem perverse to ignore it. Many existing qualitative and quantitative approaches can be adapted to take advantage of data from the Internet. Often these unobtrusive uses of Internet-derived data allow researchers to access something much closer to the experience of everyday life than we ever encounter in interview settings.

Further reading

In addition to the specific studies mentioned above, social researchers starting out on use of Internet-derived data may find Thelwall's (2009) introduction to quantitative analysis using data from the web helpful. Internet research ethics are discussed from a variety of perspectives by Johns et al. (2004) and by McKee and Porter (2009). For a recent overview of the social implications of digital communication see Baym's (2010) *Personal Connections in the Digital Age*.

References

Baym, N. K. (2010) *Personal Connections in the Digital Age.* Cambridge: Polity.

Beaulieu, A. (2005) Sociable hyperlinks: an ethnographic approach to connectivity. In *Virtual Methods: issues in social research on the Internet*. Ed. C. Hine. Oxford: Berg pp. 183-198.

Beaulieu, A. and E. Simakova (2006) Textured connectivity: an ethnographic approach to understanding the timescape of hyperlinks. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics* **10** (1). http://www.cindoc.csic.es/cybermetrics/articles/v10i1c5.html

Cheong, M. and V. Lee (2011) A microblogging-based approach to terrorism informatics: exploration and chronicling civilian sentiment and response to terrorism events via Twitter. Information Systems Frontiers **13**(1): 45-59.

Dahlgren, P. (1988) What's the meaning of this? Viewers' plural sense-making of TV news. Media, *Culture and Society* **10** (3): 285-301

Dirksen, V., A. Huizing and B. Smit (2010) 'Piling on layers of understanding': the use of connective ethnography for the

study of (online) work practices. *New Media & Society* **12** (7): 1045-1063.

Dutton, W. H., E. J. Helsper and M. M. Gerber (2010) *The Internet in Britain 2009* Oxford: Oxford Internet Institute http://www.oii.ox.ac.uk/research/oxis/OxIS2009_Report.pdf

Ellery, P., W. Vaughn, J. Ellery, J. Bott, K. Ritchey and L. Byers (2008) Understanding internet health search patterns: an early exploration into the usefulness of Google Trends. *Journal of Communications In Healthcare* **1** (4): 441-456

Gaffney, D., S. Gilbert and I. Pearce. (2011) 140kit: Regarding Twitter's API Rules and Data Export. Retrieved 25 February, 2011, http://140kit.com/documents/Regarding_API_Change.pdf.

Garton, L., C. Haythornthwaite and B. Wellman (1997) Studying online social networks. *Journal of Computer Mediated Communication* **3** (1).http://jcmc.indiana.edu/vol3/issue1/garton.html

Golder, S. A., D. Wilkinson and B. A. Huberman (2007) Rhythms of Social Interaction: Messaging within a Massive Online Network. *3rd International Conference on Communities and Technologies* (CT2007). June 28-30, 2007, East Lansing, MI.http://www.hpl.hp.com/research/idl/papers/facebook/

Harvey, K. J., B. Brown, P. Crawford, A. Macfarlane and A. McPherson (2007) 'Am I normal?' Teenagers, sexual health and the internet. *Social Science & Medicine* **65** (4): 771-781

Hine, C. (2007) Connective ethnography for the exploration of e-science. *Journal of Computer Mediated Communication* **12** (2).http://jcmc.indiana.edu/vol12/issue2/hine.html

Hine, C. (2011) Towards ethnography of television on the internet: A mobile strategy for exploring mundane interpretive activities. *Media Culture and Society* **33** (4): 581-596

Johns, M. D., S. L. Chen and G. J. Hall (2004) *Online Social Research: methods, issues & ethics*. New York: Peter Lang.

Kellehear, A. (1993) *The Unobtrusive Researcher: a guide to methods* St Leonards, NSW: Allen & Unwin.

Lee, R. M. (2000) *Unobtrusive Methods in Social Research.* Buckingham: Open University Press.

Levine, E. (2007) Television, sexual difference and everyday life in the 1970s: American youth as historical audience. *Participations* **4** (1).http://www.participations.org/Volume%204/Issue%201/4_01_levine.htm

McKee, H. A. and J. E. Porter (2009) *The Ethics of Internet research: a rhetorical, case-based process*. New York: Peter Lang.

McLaughlin, M., S. B. Goldberg, N. Ellison and J. Lucas (1999) Measuring Internet audiences: patrons of an on-line art museum. In *Doing Internet Research:critical issues and methods for examining the net*. Ed. S. Jones. Thousand Oaks, CA: Sage pp. 163-178.

Park, H. W. and M. Thelwall (2003) Hyperlink analyses of the World Wide Web: A review. *Journal of Computer-Mediated Communication* **8** (4) http://jcmc.indiana.edu/vol8/issue4/park.html

Press, A. and C. Johnson-Yale (2008) Political talk and the flow of ambient television. Women watching Oprah in an African American hair salon. In *New Directions in American Reception Study*. Ed. P. Goldstein and J. L. Machor. New York: Oxford University Press pp. 307-325.

Rogers, R. (2010) Internet Research: The Question of Method—A Keynote Address from the YouTube and the 2008 Election Cycle in the United States Conference. *Journal of Information Technology and Politics* 7(2/3): 241-260

Savage, M. and R. Burrows (2007) The coming crisis of empirical sociology. *Sociology* **41** (5): 885-899.

Seale, C., J. Charteris-Black, A. MacFarlane and A. McPherson (2010) Interviews and Internet forums: a comparison of two sources of qualitative data. *Qualitative Health Research* **20** (5): 595-606

Thelwall, M. (2008) Fk yea I swear: cursing and gender in MySpace. *Corpora* **3** (1): 83-107.http://www.euppublishing.com/doi/abs/10.3366/E1749503208000087

Thelwall, M. (2009) *Introduction to Webometrics: quantitative web research for the social sciences*. San Rafael, CA: Morgan and Claypool.

Webb, E. J., D. T. Campbell, R. D. Schwartz, L. Sechrest and J. B. Grove (1981) *Nonreactive Measures in the Social Sciences*. Dallas, TX: Houghton Mifflin. 2nd edn.

Wouters, P. and D. Gerbec (2003) Interactive Internet? Studying mediated interaction with publicly available search engines. *Journal of Computer-Mediated Communication* **8** (4): 0-0. http://jcmc.indiana.edu/vol8/issue4/wouters.html

social research UPDATE is distributed without charge on request to social researchers in the United Kingdom by the Department of Sociology at the University of Surrey as part of its commitment to supporting social research training and development.

Contributions to *social research UPDATE* that review current issues in social research and methodology in about 2,500 words are welcome. All UPDATE articles are peer-reviewed.

social research UPDATE (ISSN: 1360-7898) is published by the Department of Sociology, University of Surrey
Guildford GU2 7XH, United Kingdom. tel: +44 (0)1483689450
Edited by Nigel Gilbert (n.gilbert@surrey.ac.uk)