Online Qualitative Research in the Age of E-Commerce: Data Sources and Approaches

Nikhilesh Dholakia & Dong Zhang

Abstract: With the boom in E-commerce, practitioners and researchers are increasingly generating marketing and strategic insights by employing the Internet as an effective new tool for conducting well-established forms of qualitative research (TISCHLER 2004). The potential of Internet as a rich data source and an attractive arena for qualitative research in e-commerce settings—in other words cyberspace as a "field," in the ethnographic sense—has not received adequate attention. This paper explores qualitative research prospects in e-commerce arenas.

Table of Contents

1. Introduction

As a medium for commercial transactions—that is, as a vehicle for e-commerce and e-business—the Internet has attracted enormous attention (NEGROPONTE 1995). In these commercial transactions, Internet based technologies also often collect large amounts of qualitative data. The potential of Internet as a rich qualitative data source for studying e-commerce settings—in other words the commercial cyberspace as a "field," in the ethnographic sense—has not received adequate attention (KOZINETS 2002). This paper explores the nature of qualitative data in e-commerce settings and possible ways of mining such data using qualitative methods. [1]

The paper has three main parts. First, it establishes a *typology of online ecommerce-related data sources* that present potential arenas for qualitative research. Dimensions of the proposed typology include technological forms, sociological functions, purposes, and sponsors of web sites. Second, this article identifies *seven characteristics of the online data sources in contrast to conventional data sources*. Through this process of contradistinction, many of the important properties of cyberspace as a "field"—the computer-mediated arena where e-commerce happens in reality—are mapped out. Finally, specific research techniques that could be applied to each data type are discussed, in light of the general and specific properties of various types of cyber-data. This is done by developing a table that juxtaposes the characteristics of qualitative data with three qualitative methods relevant to online settings: interviews, focus groups, and netnography. The cells of this table provide our assessment of how the data characteristic relates to the methods. In conclusion, the methodological implications of the data-versus-technique characteristics are discussed to draw practical as well as theoretical lessons of conducting qualitative research relevant to e-commerce. [2]

2. Typology of Qualitative E-Commerce Data Sources

Internet is much more than a communication medium—it is also an evolving nexus of cross-referenced databases (POSTER 1990, 1995). In the commercial part of the Internet, such databases are regarded as central to the strategic interests of business firms (ZWICK & DHOLAKIA 2004a, 2004b). Qualitative data accumulates in the commercial cyberspace in a number of ways:

- Bulletin Board Systems (BBS): A bulletin board is a medium that displays all messages that have been posted on it and their respective replies. Bulletin board postings usually remain in place for some time, and archived versions are frequently available as well. For example, in February 2004, the "Home Audio" BBS at http://www.ecoustics.com/ had accumulated nearly 27,000 postings relating to audio products such as Amplifiers, Cassettes, CDs, MP3, MD, DAC, DAT, Receivers, Speakers, and Tuners.
- Newsgroups: The function of newsgroups is to distribute all messages posted in its specific topic area to all users who have requested to receive them. Newsgroups are categorized and organized on the basis of a variety of interests. Newsgroups follow the global naming structure of USENET based on subject domains, the threading of messages, and the fact that the messages are moved around from server to server. Users need some technical competence in configuring access to newsgroups. For example, the newsgroup comp.sys.mac caters to miscellaneous issues relating to Apple Macintosh computers.
- Chat Rooms: When two or more users are signed on to the Internet at the same time and they wish to communicate on a real-time basis, they can meet in an Internet site called a chat room. Each can type a message that will be transmitted immediately to the chat room where it can be read by anyone

present. For example, one can chat about fast food brands of pizza and burgers at http://www.kenkuhl.com/fastfood/.

- Server Log Files: The web site server will automatically generate a log file for each visitor. The log file records the IP address (an ID for each computer logged on Internet) of the visitor, how long the visitor has viewed each web page, from where did the visitor link to this web site, etc. For example, the NetTracker web analytics tool from http://www.sane.com/ can help companies answer questions such as "Which paths are being used by visitors to reach my e-commerce web site?"

- Web sites: These include commercial web sites or web sites sponsored by individual or noncommercial organizations. Some of the commercial web sites, such as those of Amazon.com, have become rich repositories of customer comments and reviews of products featured on those sites. Other sites, such as Epinions.com, are designed explicitly to review and comment on products and services. Yet other sites could have the explicit aim of attacking or parodying a brand.

- Blogs: A blog—the shortened slang for "weblog," and now a word in itself—is akin to an online diary where one or more individuals record their observations about topics of interest. Many blogs focus on products, services, and technologies that go into products. For example, at My iPod Blog, one can find various reports, tips, and musings about prices, problems, and uses of Apple's iPod music device. [3]

3. Characteristics of E-Commerce Data Sources

Data generated in the commercial cyberspace usually have characteristics that are different from other types of primary or secondary data (MANN & STEWART 2000). The various characteristics of Internet-derived qualitative e-commerce data, in turn, influence the methodologies that can be employed with such data. [4]

3.1 Text-based

HOLGE-HAZELTON (2002) contends that an oral culture is different from a written culture, and the culture of the Internet is different from both. Essentially, these cultural differences originate from the specific language characters of computer-mediated communications (CMC). Written communications differ from oral communications in that people are able to carefully select what they want to say, how much they want to say, and how they want to say it—without fear of being interrupted before they have fully made their points. When people have to write things down to communicate, they often are more articulate. They work harder at saying what they mean. On the other hand, the language of computer-mediated communication is more informal than ordinary written texts. It has therefore been characterized as *written speech* (PACCAGNELLA 1997)—a text-based communication halfway between written communication on paper and speech. [5]

Netnography, a term coined by KOZINETS (2002) to represent online ethnography, is based primarily on the observation of textual discourse observable on the Internet. Netnography thus differs from in-person ethnography in that in netnography there is no balancing of discourse and observed behavior that occurs during in-person ethnography (KOZINETS 2002). Compared to the data sources of traditional ethnography, the text-based CMC sources lack aural and visual cues. In the field, the aural cues may include pausing and reflection, emotions discernible in voice, loudness and pitch of speech, and speaker characteristics such as age, gender, national origin, ethnic group, and class accents. Visual cues include characteristics of the observed person such as appearance, height and weight, attire, gender, age, race, ethnic group, facial expressions, eye contact, body language and gestures, and emotive responses. The drawback of this lack of sound and visual cues in netnography is apparent— important information is lost, making it hard for the researcher to be sure about the real meanings and intentions of the "online informants." The lack of sound and visual cues among the informants also has some advantages—the online environment may lead to more democratic and open communications, and researcher interpretations that are more grounded in content than in stylistic and inherited elements. [6]

3.2 Publicly available

Accumulated online data can be downloaded easily and economically. Such data can be stored and analyzed anytime, anywhere, by anyone who has Internet access to the data files. The critical challenge is not getting the data, but getting the right data (MUNIZ & O'GUINN 2001). Given the tremendous quantity of information online, researchers may face serious information overload. It is well known how, in social sciences as well as in other fields, the phenomena being studied are modified by the very act of observing them. Even in the case of soft, qualitative techniques such as participant observation, problems arise because of the presence of the researcher in the field (PACCAGNELLA 1997). In some cases, unobtrusive observation can be carried out without informing the people being studied. As KOZINETS (2002) observed, in a sharp break from traditional ethnography, an online researcher can conduct rigorous netnography using only observation and downloads and without writing a single field note. Non-participant observation of the naturally occurred communication could fundamentally reduce the dangers of distorting data and behavior by the presence of the researcher. [7]

3.3 Anonymous

The anonymous character of online data provides researchers with an unprecedented chance to observe tremendous amounts of extensive social communications, connections, and cultures. Anonymity offers several advantages. First, anonymity helps create great quantity of online data that, compared to the data collected in traditional fieldwork, is more likely to reflect the real thoughts of the informants. The online anonymous communication environment makes it possible for informants to be more open to discuss their own real thoughts. They may have a chance to get away from the social classes,

rituals, obligations, and taboos, and say whatever they want to say. Well-known laboratory experiments comparing face-to-face communication with electronic mail have shown that computer networks have a *status equalization* effect (DUBROVSKY, KIESLER & SETHNA 1991), and participants are more inclined to have a democratic conversation. [8]

Second, anonymity makes it possible for the researcher to lurk in the online communication environment, conducting a really unobtrusive participant observation. The hovering presence of the research can impose some coercive requirements on the research subjects and this is a serious shortcoming of traditional research methods (LOEWENSTEIN 2001). The unobtrusive nature of Internet-based qualitative research removes such "demand artifacts" and perceived pressures on research subjects. [9]

Anonymity, however, raises the issue of the identity of the informant (ZWICK & DHOLAKIA 2004a). Two questions may be asked: how could we be sure that people are who they claim to be? Secondly, does it matter? [10]

For the first question, even in the real world, we are not always sure about whether people are really who they claim to be. The way we authenticate the identity of our informants is mainly based on the social context. This same rubric can be applied to online informants. If we observe or interact with the informants long enough and extensively enough, we can be pretty sure as to who they are. After 18 months of interacting with his informants (diabetic children) via email, HOLGE-HAZELTON (2002) concluded that: "I would consider that the chance of someone constructing a false identity as a diabetic, with all the details and inside knowledge about the illness, would be very small." [11]

The second question may be more relevant in terms of leveraging the power of the new research arena. The reason we want to authenticate the identity of our informants is mainly because we use the Internet as a tool to answer research questions about the real brick-and-mortar world, where anonymity is rarely an issue. As the Internet has influenced almost all human social activities, there is an urgent need to take the anonymous character of online communication itself as a research focus, rather than treating Internet as an instrument to study "real" life. From this perspective, identification of the informants may not even be necessary in terms of pursuing research about the culture and symbolic implications of online communications. The personae inhabiting cyberspace indeed are the authentic "subjects" that populate such virtual spaces. [12]

3.4 No material incentive

Typically there are no material incentives offered to informants who provide data in online settings. In offline research, small monetary incentives, a complimentary sandwich, or extra-credit in a course may seriously skew the data, because the informants may participate only for the incentive, or they may feel some responsibility or a sense of reciprocity to provide whatever information the researchers would like to have. BRADFORD (1999) reported that some recruited

panelists were "professional" survey takers, and would supply whatever information they thought would make them more attractive for studies. Personal web sites and newsletters even promoted such panels and surveys as ways to earn easy money. For most qualitative data that is automatically generated in e-commerce settings, such incentives do not come into play. There may, however, be some real or perceived rewards for those Net participants who begin to emerge as "experts" in a commercial area that is of interest to a lot of people. [13]

3.5 Unsolicited

One of the dramatic cultural elements that we can readily observe in the online bulletin board systems (BBS) or newsgroups is that so many people spontaneously and voluntarily contribute to the collective efforts of these virtual communities. The only unsolicited data source that possibly parallels this is a person's private diary. [14]

Some relevant questions in this regard are: Why do people provide so much unsolicited information online? How do community members make sense of what they are doing in virtual communities? What are the methodological implications of such unsolicited data, a data warehouse that never existed before or was extremely difficult to access in the physical world? By and large, the spontaneously proffered comments in e-commerce settings appear to be genuine rather than contrived expressions of attitudes, opinions, and beliefs. [15]

3.6 Temporal flexibility

Compared to experimental design, survey, or real time interview, most netnography is based on the data generated through asynchronous communication. An apparent benefit is that such data usually do not lapse or get lost. Rather, the data are stored in cyberspace, waiting for downloading by researchers at times convenient to the researchers. Another benefit is that the researchers can conduct longitudinal research to discover and understand the dynamics and the evolutionary patterns of the specific culture and social climate of a particular piece of cyberspace. HEMETSBERGER (2003) argued that in order to fully understand the development of any online community culture, it is important to look at the historical context of that group's development. [16]

Empirical reality suggests that consumer enthusiasm or grievances often exist for prolonged periods before significant events and developments and/or social actors convert those energies into action. Therefore, for understanding the very core of a consumer community's culture, sampling of online and offline data that date back to the historical roots of consumer collective action is a must. [17]

3.7 Spatial flexibility

The traditional anthropological concept of "field" usually entails a specific geographic context, often in far away exotic locales, where the fieldworker is a sojourner. In online settings, these have been replaced with much more complex

and critical notions of both the field and the fieldworker (HEMETSBERGER 2003). With online data sources, it is possible for the first time to conduct cross-regional, cross-cultural "fieldwork." For example, STEWART, ECKERMANN and ZHOU (1998) conducted online focus groups including one with 16 teenage girls located in Australia and China. To study online discussions pertaining to these girls' knowledge of smoking related health risks, the researchers used chat rooms in English and Chinese (with translations). [18]

Several types of spatial flexibility are possible in online qualitative research. In most cases, there is pre-existing online data—the online ethnographic (or "netnographic") field has already been defined by the users of particular BBS, chat rooms, web sites or blogs (see MALLON 2004 for example of such a study). Researchers can, however, take the initiatives to create netnographic "fields" that are researcher-defined, or jointly defined by the researcher and online users. [19]

4. Matching Qualitative Techniques to E-Commerce Data

The field of online qualitative research in e-commerce settings is very new and on a fast-evolutionary trajectory. It would therefore be presumptuous of us to offer a comprehensive roadmap for qualitative research methods suitable for all types of data in e-commerce settings. It is, however, feasible to start building the components of such a roadmap. [20]

Table 1 provides our assessment of how the different types of qualitative research data generated in e-commerce settings fare on the characteristics of online qualitative data discussed in the preceding section.

Data Characteristics	Online Qualitative Data Type					
	BBS	**News-group**	**Chat**	**Server Log**	**Web site**	**Blog**
Text-based?	Yes	Yes	Yes	Mostly No	Mostly Yes	Mostly Yes
Publicly available?	Yes	Yes	Yes	Mostly No	Depends	Yes
Anonymous?	Yes	Yes	Yes	Partly Yes	Partly Yes	Mostly No
Material incentive?	No	No	No	Mostly No	Mostly No	No
Unsolicited?	Yes	Yes	Yes	Mostly Yes	Partly Yes	Mostly No
Time aspects (LT: Long-Term, ST: Short-Term)	LT Archive	LT Archive	ST Archive	Depends	Mostly LT Archive	LT Archive
Space aspects/ The "field" **(UD: User defined, CD: Creator defined, JD: Jointly defined)**	JD	JD	UD	CD	CD tending to JD	CD

Table 1: Qualitative Data Types and Data Characteristics [21]

It is clear from Table 1 that not all online data types share all the online data characteristics equally. In general, the more control and direction that a central authority exercises on the structure of data accumulation, the less that data accumulation method correlates with the generic characteristics of online qualitative data. Thus, for example, log files offer mostly quantitative data in highly structured formats, with only some qualitative snippets such as "search terms employed by the user" to reach the web site that is being monitored by the log file analytic software. [22]

Let us turn to three possible qualitative research methods—interviews, focus groups, and netnography—that can be employed in e-commerce settings. Table 2 provides our assessment of how these three methods stack up in terms of the general online qualitative data characteristics that we have discussed.

Data Characteristics	Online Qualitative Data Collection and Analysis Method		
	Interviews	**Focus Groups**	**Netnography**
Text-based?	Yes	Yes	Yes
Publicly available Data?	Mostly No	Mostly No	Yes
Anonymous?	Depends on Research Design, Often No	Depends on Research Design, Often No	Yes
Material Incentive?	Depends on Research Design, Often Yes (Prizes/Contests)	Depends on Research Design, Often Yes (Prizes/Contests)	No
Unsolicited?	Mostly No	No	Mostly Yes
Time aspects (LT: Long-Term, ST: Short-Term)	Private Archive of Data, ST or LT	Private Archive of Data, ST or LT	Usually Public Archive of Data, typically LT
Space aspects/ The "field" **(SD: Subject defined , RD: Researcher defined, JD: Jointly defined, UD: User defined)**	RD, could tend to SD in Ethnographic Interviews	RD, could tend to JD with light moderation	Mostly UD, but possible to create RD and JD netnographic fields

Table 2: Online Qualitative Research Methods and Data Characteristics [23]

It is evident that "online ethnography" (WITTEL 2000) or "netnography" (KOZINETS 2002) is the least intrusive online method—far less intrusive than its offline predecessor. The degree of intrusiveness for the other two methods—online interviews and online focus groups—is less than their offline counterparts, but the gap on the "intrusiveness" dimension is not as dramatic as the offline-online ethnography gap. [24]

Therefore, we believe that commercially oriented qualitative research in e-commerce settings will turn increasingly to netnography in a fairly big way when the objectives of the research are truly qualitative: to interpret behavior and to understand the users of e-commerce. Online interviews and focus groups (HECKMAN 2000) will continue being favored in situations where the commercial objectives are specific and precise and analytical answers are needed for the research questions. [25]

5. Summary and Conclusions

E-commerce went through a wrenching period of structural competitive adjustment in the late 1990s. Contemporary e-commerce is hardened and tempered by the trial-by-fire that the "dotcom crash" represented. The e-commerce practitioners and researchers therefore are experienced in terms of their market research requirements. Many are increasingly generating marketing and strategic insights by employing the Internet as an effective new tool for conducting well-established forms of qualitative research. This paper explored the potential of Internet as a rich data source and an attractive arena for qualitative research in e-commerce settings. [26]

While cyberspace is still evolving as a "field," in the ethnographic sense, our review and assessment in this paper point to a growing role for qualitative research approaches in e-commerce arenas. In particular, when facing demands for deep interpretation of consumer behavior and understanding of market and competitive dynamics, commercially oriented qualitative research in e-commerce settings will turn increasingly to netnography. Other online qualitative methods, such as online interviews and focus groups, would be popular in situations where the commercial and research objectives are analytically narrower and less interpretive. [27]

Acknowledgments

References

Bradford, David P. (1999). Recruiting sources for on-line studies. *Quirk's Marketing Research Review*, July. Available at: http://www.quirks.com/.

Dubrovsky, Vitaly; Kiesler, Sandra & Sethna, Beheruz (1991). The equalization phenomenon: Status effects in computer-mediated and face-to-face decision making groups. *Human Computer Interaction, 6*, 119-46.

Heckman, James (2000). Turning the focus online. *Marketing New, 34*(5), 15.

Hemetsberger, Andrea (2003). *Understanding consumers' collective action on the Internet—a definition and discussion of relevant concepts for research*. Working paper presented at the annual ACR (Association for Consumer Research) conference 2002 in Atlanta, Georgia.

Holge-Hazelton, Bibi (2002). The Internet: A new field for qualitative inquiry? [36 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], *3*(2), Art. 15. Available at: http://www.qualitative-research.net/fqs-texte/2-02/2-02holgehazelton-e.htm.

Kozinets, Robert V. (2002). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research, 39*(1), 61-72.

Loewenstein, George (2001). The creative destruction of decision research. *Journal of Consumer Research, 28*(3), 499-505.

Mallon, Kathleen (2004). *The birth of bizdames: A Rupture in the conversation, engaging the body in decision-making*. Unpublished Ph.D. Dissertation, Kingston, RI: University of Rhode Island.

Mann, Chris & Stewart, Fiona (2000). *Internet communication and qualitative research: A handbook for researching online*. London: Sage.

Muniz, Albert M. & O'Guinn, Thomas C. (2001). Brand community. *Journal of Consumer Research, 27*(4), 412-432.

Negroponte, Nicholas (1995). *Being digital* (1st ed.). New York: Knopf.

Paccagnella, Luciano (1997). Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *Journal of Computer Mediated Communication, 3*(1). Available at: http://ascusc.org/jcmc/vol3/issue1/paccagnella.html [Broken link, FQS, August 2005].

Poster, Mark (1990). *The mode of Information*. London: Blackwell.

Poster, Mark (1995). *The second media age*. London: Blackwell.

Stewart, Fiona; Eckermann, Elizabeth & Zhou, Kai (1998). Using the Internet in qualitative public health research: A comparison of Chinese and Australian young women's perceptions of tobacco use. *Internet Journal of Health Promotion, 12*. Available at: http://www.monash.edu.au/health/IJHP/1998/12.

Tischler, Linda (2004). Every move you make. *Fast Company. April*, 72-75.

Wittel, Andreas (2000). Ethnography on the move: From field to net to Internet [23 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], *1*(1), Art. 21. Available at: http://www.qualitative-research.net/fqs-texte/1-00/1-00wittel-e.htm.

Zwick, Detlev & Dholakia, Nikhilesh (2004a). Whose identity is it anyway? Consumer representation in the age of database marketing. *Journal of Macromarketing* (forthcoming).

Zwick, Detlev & Dholakia, Nikhilesh (2004b). Consumer subjectivity in the age of Internet: The radical concept of marketing control through customer relationship management. *Information & Organization* (forthcoming).

Authors

Nikhilesh DHOLAKIA is Professor in the Marketing, E-Commerce, and Management Information Systems areas in the College of Business Administration at the University of Rhode Island. His research deals with problems at the intersection of technology, globalization, and consumer culture.

Contact:

Dr. Nik Dholakia

Professor
Marketing, E-Commerce & Information Systems Areas
College of Business Administration
7 Lippitt Road
University of Rhode Island
Kingston, RI 02881 (USA)

E-mail: nik@uri.edu
URL: http://www.cba.uri.edu/

Dong ZHANG is in the Doctoral Program in the College of Business Administration at the University of Rhode Island, specializing in Marketing and E-Commerce. His research focus is on the role of online virtual communities in marketing and consumption related processes.

Contact:

Dong Zhang

College of Business Administration
7 Lippitt Road
University of Rhode Island
Kingston, RI 02881 (USA)

E-mail: dgzhg@yahoo.com

Citation