Shani Orgad

How can researchers make sense of the issues involved in collecting and interpreting online and offline data?

Book section

How Can Researchers Make Sense of the Issues Involved in Collecting and Interpreting Online and Offline Data?

Shani Orgad

In A. N. Markham & N. K. Baym (eds), *Internet Inquiry: Conversations About Method*. Thousand Oaks: Sage (pp. 33-53).

Before addressing the question that is the subject of this chapter, I want to introduce two working definitions: "qualitative Internet research" and "online and offline data", which, without being in any way prescriptive, reflect my personal understanding of these concepts. This understanding has been significantly influenced by my own research experience. Following a brief discussion of these concepts, I move on to explain why I think consideration of online and offline data is important in thinking through our research projects. Next, I explore critical junctures in the research process when these issues might arise and become a problem. I put forward possible justifications for doing research which combines online and offline data. I also discuss the implications of deciding not to obtain and analyze offline and online data, but to rely on one kind of data only. The paper concludes with some thoughts about the question of online and offline data in future qualitative Internet research, in light of current technological trends which are increasingly blurring the line between online and offline communication, and recent debates about the nature of the research field site in Internet studies.

Working definitions

Qualitative Internet research

Rather than being qualitative research that uses the Internet to facilitate data collection or data analysis, qualitative Internet research is a qualitative inquiry into Internet phenomena (Markham, 2003). More specifically, by the term "qualitative Internet research" I refer to the study of the various and complex meanings and experiences that emerge around the Internet in a particular context. These meanings and experiences can relate to contexts of use (by individuals, organizations, networks etc.), and/or to contexts of design and production processes. The task of a researcher

involved in a qualitative Internet research project is to inquire into those meanings and experiences and explore their significance.

The question underlying a qualitative Internet research project would be: what does "the Internet" stand for in a particular context, for particular agents? Clearly, "the Internet" is not a monolithic thing. Part of any qualitative exploration would have to be articulating what research arenas "the Internet" comprises, and how they shape, as well as being shaped by, participants' and producers' experiences of use. For example, my study *Storytelling Online: Talking Breast Cancer on the Internet* (Orgad, 2005) started by mapping 'the landscape of breast cancer patients' online communication', describing the kind online spaces and environments participants engage. This 'landscape' defined the arenas that the research focused on, which included cancer-related message boards, personal diaries, and e-mail.

Equally, 'qualitative' does not map onto one single thing. While I acknowledge the difficulty, if not impossibility, to reach a clear definition of 'online qualitative research' (see also Denzin, 2004: 4), my own perspective leans towards the interpretative as opposed to the more positivist and naturalistic conception of human experience and its analysis. More specifically, 'qualitative' to me implies a commitment for an interpretive understanding of people's experiences of the Internet, and of the texts (in the broad sense) they create online and offline. Crucially, as Denzin (2004: 7) usefully points out, "online interpretative work provide(s) the foundations for social criticism and social action". For instance, in my own study a qualitative approach meant not just documenting and describing patients' practices of telling their personal stories online – though this was a substantial part of the work – but fundamentally also thinking about these practices critically: whether, how and to what extent women's storytelling online transform their experiences and the cultural and social environments in which their experiences are embedded.

Online data and offline data

To investigate the above question and inquire into a specific Internet phenomenon, as in any research, the researcher has to obtain data. The data can be obtained from two main types of sources: online and offline. They can include *texts* such as online postings and textual elements such as threads or links, face-to-face interview accounts, or ethnographers' field notes, *images* such as pictures from websites or photos of spaces that are related to users' experience of the Internet, and *sound,* for

example online clips. In short, 'data' refers to all the information derived from employing qualitative research procedures. *Online data* are obtained using what have been often described as "virtual methodologies": methodologies implemented by and through the Internet. These include, for instance, participant observation in online spaces such as MUDs (see, for example, Baym's study of an online community of soap opera fans, 2000; Kendall's study of BlueSky, 2002; Schaap's online ethnography of a role-playing MUD called New Carthage, 2002). The ethnographic material that researchers reap from their online ethnography constitutes what I refer to here as "online data". Another example of online data is texts of interviews with research participants that are conducted online. Kivits (2005), for example, conducted interviews with Internet users via e-mail, to explore their use of the Internet for seeking health information. She analyzed the online data she obtained, namely the e-mail transcripts of the online interviews, to account for users' information seeking practices on the Internet, and their sense-making of this information.

The other kind of data is obtained using "traditional" methodologies in "traditional" offline settings. Here, in order to study Internet-related phenomena, the researcher employs methodological procedures in offline contexts, which generate *offline data*. For example, in studying the integration of the Internet in the everyday lives of users, researchers such as Bakardjieva and Smith (2001), and Mackay (2005) conducted ethnographic visits to and interviews in the domestic settings of Internet users. Influenced by studies of television audiences, studies of this kind are based on offline data that consist of users' accounts obtained through interviews, participant observation in users' households, and in the case of Mackay, users' diaries recording their media use.

Why is it important to consider the question of online/offline data in thinking through our research projects?

A distinction between online and offline has never been made in qualitative research concerning different communication media. For instance, researchers do not discuss the use of "television data" versus "offline data", or "telephone data" versus "everyday data". More generally, beyond the methodological context, we do not tend to talk about the "television world" versus the "offline world", or about "radio contexts" versus "offline contexts" in the same way as we refer to "online" and "offline" in relation to the Internet. This distinction between the online and the

offline, and consequently between online and offline data in the research context, is rooted in an interrelated distinction that has specifically characterized common thinking about the Internet. Hine (2000) usefully describes this distinction as that between the view of the Internet as a "cultural artifact" and as a "culture". On the one hand, the Internet, like other communication media, has been seen as a medium. Researchers working within this premise have explored how it is used as a means of communication within our social lives. As with studies of other communication technologies, they have studied the Internet within specific bounded social settings, for example, in the home. The focus therefore has usually been on offline contexts, and thus research has relied on *offline data*. On the other hand, the Internet has been commonly viewed as a communicative social space in its own right. Unlike other media such as the television or the telephone, Internet spaces have been often studied as self-contained social spaces which encompass relations and practices of their own (Slater, 2002). These spaces have commonly been seen as distinct and separate from "offline", "real" social life. In research terms, this view established cyberspace as a plausible research field site (Hine, 2000, p. 9) and advanced investigations of online social spaces independently of offline social relations (Slater, 2002, p. 535). Such studies, which are mainly versions of online ethnography, rested on the assumption that online sociality has an inherent cultural coherence which is internally meaningful and understandable in its own terms (Slater, 2002, p. 542). Consequently, study of these online contexts was predominantly conducted on online sites, relying mainly, and often exclusively, on *online data*.

The distinction between the online and offline has been constitutive of the understanding of the Internet, from the earliest days of Internet research. Methodologically, the distinction has led to a large extent to a separation between the use of offline and online data. Large-scale surveys of Internet use such as those conducted by Pew Internet & American Life Project rely their analyses of online life mainly on offline data such as information elicited by phone surveys or tracking surveys of Internet activities (see for example Howard et. al., 2001). Other researchers, such as Bakardjieva and Smith (2001), though working from a different perspective to the study of the use of the Internet, also rely their analysis predominantly on offline data including interviews with domestic users, a tour of the computer and Internet-related spaces in respondents' homes, and a group interview with respondents' family members.[1] On the other side there have been numerous

studies, especially in the early days of Internet research, drawing exclusively on online data. For example, Donath's (1999) study of identity deception in an online community, Reid's (1999) exploration of social control in MUDs, and Danet. et. al's (1997) study of language use in computer media - all relied on analysis of online texts and interactions.

More recently, however, this separation is being increasingly deconstructed. It has become clear that the separation between the online and offline cannot be sustained. Researchers have consistently argued for the need to frame the online both in its own right, and in relation to other contexts and realities (Slater, 2002). That is, in Hine's terms, to take account of the Internet as both culture and cultural artifact. This recognition clearly undermines the assumption "that only things that happen on the Internet were relevant to understanding the Internet" (Haythornthwaite and Wellman, 2002, p. 5).

Recognition of the complex relationship between online and offline has profound methodological implications. In particular, two key questions arise at two critical junctures of the research, in relation to online and offline data. The first question arises at the stage of designing an empirical research: do we need offline data in order to make sense of online phenomena? If qualitative Internet research is the study of the complex meanings and experiences that emerge around the Internet in a particular context, do we necessarily need offline information to be able to adequately account for these meanings and experiences? Or can we produce high quality, persuasive and grounded qualitative research of an Internet-phenomenon that draws merely on online data? As Slater points out (2002, p. 543), these questions become even more of an issue in a context which is notorious for identity playing and deception on the part of participants with few negative consequences. What claims can a study relying only on data retrieved online make? The "opposite" question is as intriguing: if the Internet is treated as simply a means of communication that is used in an everyday social context, can it therefore be studied as such, that is, merely by using methodological procedures in offline contexts, without any online data?

The second key question arises at the stage of data analysis and concerns the actual use and interpretation of online and/or offline data. If it was decided to obtain both online and offline data, a researcher has to grapple with the triangulation between the sets of data: are the two kinds of data comparable? Can they be integrated, and if so, how? If a decision has been made to rely on only online *or*

offline data, at the stage of analysis researchers must confront the adequacy, validity and limitations of their analysis. This problem also arises, of course, for those researchers who use *both* online and offline data.

I now turn to the two questions that I presented in relation to the issue of online and offline data in conducting and evaluating qualitative Internet research. I discuss each question in the context of the particular research junctures at which it may arise, highlighting possible ways of tackling it. Crucially, my intention is not to provide prescriptive answers, but rather to demonstrate what I regard as useful, sensible, ethical and context-sensitive approaches to these questions.

Is obtaining offline and online data necessary? If so, when? If not, why?

As mentioned earlier, the question of, and thus the decision about, whether it is necessary to obtain online and offline data arises at the very early stages of designing the empirical research. The answer to the question seems simple: "it depends on the question you ask and on the context you study". However, in what follows I want to unpack this seemingly straightforward answer, by pointing to particular considerations that might be involved, and by grounding the discussion in specific examples from my own study and that of others.

In their study of young people's cultural life and social resistance, the question Wilson and Atkinson (2005) asked was: "what is the relationship between young people's online (activist) activities and offline social action?" The emphasis in the question is on an understanding of this group's culture, and the ways its online and offline contexts inform and enable each other. In terms of empirical design, it therefore seems clear that the research must draw on both online and offline data. In other words, the theoretical concern over the relationship between online and offline contexts, in terms of a specific group's activities and practices, informs a methodology that would aim at capturing the online, the offline and the connections between them. Wilson and Atkinson's study did indeed include both online and offline components. It was based on an analysis of the contents of twenty-eight webpages, in-depth face-to-face interviews with website producers and organizers, fieldwork that involved attending events that were organized by the groups involved, and a study of the media coverage of the groups' events.

Similarly, in my research into the online communication of women suffering from breast cancer (Orgad, 2005), obtaining both online and offline data regarding the

participants' lives and experiences was crucial for making sense of the Internet context that I studied. The aim of the research was to inquire into the meanings of online engagement for women with breast cancer. Patients' online participation and their use of the Internet are deeply embedded in their everyday experience of chronic illness. Therefore, if we are to understand patients' online contexts, we clearly have to have knowledge of their offline contexts, that is, of the everyday life aspects of their coping with breast cancer. By the same token, to make sense of patients' experience of breast cancer (offline), it is necessary to get to grips with their online engagement, which is a significant part of their experience of coping with their illness. In short, since at the heart of this inquiry stood the connections between participants' online and offline experiences, how patients' online participation affects and enables their life offline, and vice versa, it seemed crucial to access both the offline and online environments of participants. So as early as at the stage of designing an empirical study I made a decision about the need to obtain online and offline data, on the basis of which I would build my analysis.

As we can see, in both these examples the decision to obtain online and offline data is situated in the context of the specific research goals. It might seem that it would always be more sensible and context-sensitive to seek access to both online and offline data. However, this is not necessarily the case. Eichhorn's (2001) study demonstrates how the researcher's decision to rely primarily on online data, and deliberately avoid the study of participants in their offline environments was an informed context-sensitive decision, which stemmed from a careful understanding of the phenomenon being studied.

Eichhorn (2001) explains that in her study of girls' textual online community ("zines"), which she carried out primarily online, it was unlikely that the tactics and practices that she aimed to examine would have been rendered visible had she opted to carry out a study within an offline environment, such as a school or a classroom. Notwithstanding the fact that the practices of production of 'zines, on which Eichhorn's study focused, may sometimes be reproduced in school-based settings, for various contextual reasons, studying them in those settings would have been an obstacle rather than a facilitator (see p. 573-574).

More generally, Eichhorn (2001) challenges the assumption that ethnographic research of an online-based phenomenon has necessarily to be dependent on face-to-face relationships with the study's participants. She shows that in her work,

participating in informants' everyday lives did not necessarily mean accessing their offline environments. On the contrary, Eichhorn insists that "understanding people's lives, particularly in the technologically driven Western world, may sometimes require ethnographers to do what the people they seek to study do, even if it necessitates staying at home" (p. 566).

A related argument against the use of offline data is that in seeking to combine online and offline data, particularly when the data relate to participants' lives and activities, researchers run the risk of implying that online data are not as authentic as offline data. That being said, employing procedures to study participants in their offline environments could be a fruitful way of contextualizing and adding authenticity to the findings obtained online (Hine, 2000, p. 48). Turkle (1996), for example, in her infamous study *Life on the Screen* reflects on the significance of conducting face-to-face in-depth interviews with her online informants, as a way to further "explore an individual's life history and tease out the roles technology has played" (p. 324). She goes so far as only to include findings on those online informants whom she also met in person, a methodological decision she justifies with her concern with the relationship between users' experiences in online reality and real life.

In my study, the transition from e-mail correspondence with my informants, to face-to-face meetings, proved extremely significant for understanding the experience of breast cancer patients' online engagement. I consider this transition to have been a key turning point in my understanding of the relationship between patients' lives and their online experience. In hindsight, I realize that so long as I only had access to participants' construction of their online experience through their e-mail accounts, the relationship between patients' lives and their online experience seemed fairly palpable and straightforward. In most of the e-mail accounts I initially received from patients, "the Internet" is described in a fairly idealized way: either as a "dazzling" and "empowering" "miracle" (reproducing popular emancipatory constructions of the Internet) or in a reductive way, as being nothing but another source of information about cancer. Participants' e-mail accounts foregrounded the role of the Internet in their experience of coping with the illness. Alternatively, the offline data, which I obtained later through face-to-face interviews, enlightened much more complex connections between patients' online and offline experiences. These accounts were primarily personal narratives about how they coped with their illness. Respondents'

reflections on the Internet and the role it played in their experience of coping were far more subtle and implicit. Rather than being about the Internet, patients' face-to-face accounts were about their lives and selves. Rather than foregrounding the experience of using the Internet (as in the e-mail accounts), in the face-to-face accounts this experience was embedded in their stories. The face-to-face interviews also enabled respondents to move away from utopian or dystopian discourses and clichés about "the Internet". Instead, in their face-to-face accounts, "the Internet" was usually disaggregated into its different components, in the particular contexts where it played a role in their coping. While in the e-mail accounts "the Internet" appeared a pretty much singular "thing", the face-to-face interviewed revealed its various facets and situated contexts.

Even Eichhorn (2001), who persuasively explains why conducting research relationships online and offline is not always appropriate and does not necessarily fit the context and goals of the research, reflects on the invaluable significance of the only face-to-face meeting she had with one of the participants of the online community she studied. She describes this meeting as an important turning point in her understanding of this community, and as presenting an opportunity to ask questions she had previously failed to recognize as being relevant to her research (p. 571).

Crucially however, in all these examples, the researchers do not treat the offline data on participants' lives and experiences as more "truthful" or "authentic" than the data obtained online. In obtaining offline data their aim was not to introduce some external criterion for judging whether it is safe to believe what informants say (Hine, 2000, p. 49). Rather than validating the veracity of the data obtained online, their rationale for deciding to gather offline data was the need to add context, to enhance information, and to yield insights into aspects that would have otherwise remained invisible, but which were consequential to the research. More generally, rather than being led by some general rules of inquiry, what guided the researchers in the above examples were the particular research contexts and the demands of their research goals.

The question of whether there is a need to enhance online data with offline data can arise at later stages of the research project. Rutter and Smith (2005), for instance, undertook research on the "RumCom" newsgroup online. They sought to discover how sociability is discursively constructed in this text-based online

environment. The major component of data used in their study was the messages that were published on "RumCom". However, this kind of data did not seem on its own to be sufficient. "We also wanted to add some depth beyond what we could discover through the analysis of messages. We felt that our online ethnography had to do more than merely observe and collect textual data" (p. 87). They therefore complemented the online data they had initially obtained, by a series of phone and face-to-face interviews with some of their online informants. This offline data allowed them to inquire into the ways in which online participants became involved in the RumCom newsgroup and what they got out of it, information that had remained obscure as long as they obtained only online data.

I have so far discussed the relationship between online and offline data only in one direction, that is, as the research moves from online to offline. One can, however, picture a research situation that starts offline, and then moves to obtaining online data. The rationale for such a move might be similar to what I described in relation to the move in the opposite direction. That is, the need to add depth to the phenomenon being studied, contextualize and enhance the offline data.

Bakardjieva and Smith (2001) designed a quasi-ethnographic study that aimed to explore computer networking from the standpoint of the domestic user. They sought to devise a methodology that would allow them to investigate "both the real-life contexts and actions of our [their] subjects *and* their exploits in cyberspace" (p. 69). Influenced by studies of the domestication of media and technology in people's everyday lives, the researchers deliberately focused on the offline environment of users' homes, as the sites in which they studied domestic practices of Internet use. The offline data they obtained included interviews with domestic users, a tour of the computer and Internet-related spaces in respondents' homes, and a group interview with respondents' family members. These data were complemented by one component of online data, which they describe as a tour of users' "computer space", that is, the traces of Internet use that were saved in respondents' computers, or in their accounts on the provider's server (p. 70). Arguably, a more elaborate use of online data, such as, for example, the ethnography of the actual Internet spaces in which these domestic users participated, could have further augmented the researchers' understanding of the ways in which the Internet is integrated into users' everyday life situations, and tied in to specific social–biographical situations.

In this context, Sanders' (2005) research is quite illuminating, as it uses a multi-layered research design, consisting of both online and offline components. In studying the sex work community in Britain, Sanders started her ethnography offline, observing indoor sex markets and street prostitution for ten months. She later found it necessary to explore the impact of CMC on the organization of sex economies, to which end she turned to the Internet to collect online data, mainly through instances of lurking. The main focus of her online observation was PunterNet, a popular site for male clients and female sex workers in the UK. Observing forums such as message boards and live chat sessions where sex workers and clients interact (textually) gave Sanders insights into how commercial sex was advertised, discussed, selected and negotiated online between clients, sex workers and owners of establishments. The researcher then realized that in order to fully understand the role of the Internet in sustaining the identities of sex workers, she needed to move back offline: to recruit online participants for face-to-face interviews. "In the same way that sex workers and clients inevitably transfer their relationship from online to real encounters", she reflects, "questions relating risks and management strategies led me to move beyond the screen to face-to-face relationships" (p.71).

Sanders' study reveals other important considerations that need to be taken into account when making decisions about the use of online and offline data, particularly when the latter involves moving relationship with participants from online to offline, and even more particularly when sensitive or high-risk groups are concerned. The nature of these considerations can be ethical, involving questions of the researcher's trustworthiness and rapport with her informants. One of the lessons Sanders and other researchers (e.g. Kendall, 2002; Mann and Stewart, 2000; Orgad, 2005) learned, is that it is highly problematic, if not impossible, to move from online to offline with informants without establishing *bona fide* status, trust and rapport. Obtaining offline information on online participants could also simply prove infeasible, particularly where hard-to-reach populations, such as sex workers, are concerned.

Another aspect that the researcher has to consider is the sample of respondents. As long as we rely on online methodologies our access is limited only to those who actively participate (e.g. who post messages), and therefore are visible. However, there are many online participants who are only lurkers, but whose participation and practices can be extremely significant and highly consequential for

understanding an Internet-related context. Yet from a discursive point of view, the "silent" are difficult to incorporate into the analysis, as they leave no observable traces (Hine, 2000, p. 25).

Let me give an example from my study of breast cancer patients' online communication. A breast cancer patient writes to her fellow-sufferers' mailing list:

> [I]f you want to...post as much as you want...even a lot in one day. If you want to, stay silent and get support without posting. If you want to, stay away for a while and come back. We have some members who come and go. AND wow, some of our members "graduate" and feel they are not in need of support ...if those members want to come back...they are always welcome.

Staying in the background--only reading messages, as the patient cited above describes it, can play a significant role in how patients' cope with their illness. Lurking enables patients to learn about others' experiences, to relate their own situation to that of others, without having to necessarily expose themselves and their feelings. Similarly, another patient reflects on the valuable therapeutic effect of putting her experience into text by typing it – before actually interacting online:

> Probably the best part of the internet is that you need to type your question or feeling before you can share it and sometimes just writing it down is a therapy of its own.[2]

To be able to inquire into these highly meaningful practices, such as lurking, or simply typing out one's experience (without necessarily publishing it online), I had to go beyond the screen, that is, I had to gain access to those participants and their activities, which would have otherwise remained invisible. Relying only on the observable representational level of online activities (i.e. texts) was not sufficient on its own to explain the significance and capture the complexity of these activities.

But how do you do this? How do you access the invisible? To tackle this, I recruited some interviewees by snowballing offline. I exploited initial contacts with women whom I met online in order to recruit their acquaintances as interviewees who participated online, but not necessarily actively and visibly. This enabled access to an appropriate range of participants engaged in different levels of involvement, in different kinds of online activities in relation to their illness.

Researchers may also be interested in studying those who are not online, those who "fail" or refuse to engage online. This can be an interesting phenomenon to study in itself. It can also shed light on the study of Internet use and online participation. For example, in my study one of the interviewees was a patient who initially visited breast cancer patients' forums, however after a short time became

very critical of these sites and stopped participating in them. Nor did she reply to my online request for women willing to participate in the study. "You would have never found *me* online", she told me in our face-to-face interview. Indeed, I recruited her through snowballing (another patient referred me to her), rather than online, as I did with the majority of the interviewees. Though my research focused on the experience of women who participated online in the context of their illness, rather than non-participants, the experience of this woman and of a couple of other non-users I met proved invaluable. They illuminated some of the significant constraints of the spaces in which patients were actively participating and helped me think critically about the phenomenon I studied: to what extent are these online spaces inclusive, allowing 'people from all walks of life' (as one forum describes its mission) to share their experience?

Clearly, if we wish to study those who are not online, relying on online data is not sufficient. We need to gain access to informants' offline contexts and retrieve offline data. Indeed, driven primarily by the digital divide agenda, researchers have recently recognized that studying those who are not online can be a significant aspect of understanding Internet phenomena (e.g. Lenhart, 2001, based on a telephone survey). I suggest that exploring participants who are excluded from certain CMC contexts, or have "failed" to engage in CMC, can be very fruitful for qualitative studies of Internet phenomena, and not just in relation to the digital divide. In my study, for example, although my primary concern was with those patients who were "successfully" engaging online, exploring cases of patients who for different reasons "failed" to participate in breast cancer online spaces proved highly enlightening. For instance, two of my interviewees, despite having the technical capacity and competence for engaging in CMC, found the breast cancer Internet sites they encountered inappropriate and unsatisfactory. Their experience of rejecting the Internet as a communicative space in coping with their illness was extremely telling – not only in terms of the specificity of their experience, but also in terms of the light it threw on the majority of the "successful" cases. These two patients were looking for a forum that would allow a critical and rational discussion on breast cancer whereas the majority of the forums they found online focused on patients' experiential, subjective and personal stories. This distinction helped me understand the centrality of the subjective, experiential and confessional discourse that governs many patients' Internet spaces, and in particular the significance of storytelling as a key social

activity in which breast cancer patients engage online (see Orgad, 2005). Clearly, if I had not recruited interviewees offline (rather than through the Internet), and interviewed them face-to-face, I would not have gained the insights yielded by these valuable accounts, nor recognised the complexity and specificity of the phenomenon I was studying.

Whereas I started online, and then moved offline with my research participants, Eichhorn (2001) decided to locate her research almost exclusively in an online site, relying primarily on online data. Situating her research online, rather than in an offline setting such as school or classroom, enabled her to examine a group of young women not always visible in the school system. As she reflects:

> Significantly, many of these young women wrote about feeling either invisible or even at risk in the school environment…In contrast with the lack of visibility many of these young women experienced in their schools, the textual community of 'zines was a space in which these young women, many not 'out' in their local communities, could have their identities and experiences recognized and validated" (p. 574).

So Eichhorn's decision to locate her research in an online site, relying primarily on online data, opened up the possibility of studying "this often unaccounted for group of young people" (Eichhorn, 2001, p. 574). Whatever decision is made, the crucial point is that it should be sensitive to the context being studied, and situated within the demands of the research question.

How to use and analyze online and offline data?
The other critical juncture at which the issue of online and online data arises is the stage of analysis and interpretation of the data. If the researcher has decided to obtain online and offline data, she may find herself puzzled at this stage in the analysis about how to integrate the two sets of data. Do they correspond? Are the two sets of data comparable, and if so, how?

Such questions become an issue particularly if the rationale for obtaining online and offline data was to break down the online/offline distinction conceptually. In regarding the data obtained as "online" versus "offline", i.e. as two sets of distinctively different and separate data, we run the risk of reproducing the very idea that we aimed to challenge, that is, that the online and the offline are two separate distinguished realms.

Thus, a rule of thumb for analysis of the data is that an attempt to break down the distinction between online and offline cannot be pursued only in theory; it is a project substantially implicated in methodology, and in this context, particularly in the way the data are treated. In what follows I will try to demonstrate some of the implications of adopting this rule of thumb, in reflecting upon the data analysis in my research on breast cancer patients' online communication.

My data analysis involved three different types of texts: (1) e-mail accounts; (2) online texts from breast cancer websites; and (3) face-to-face interviews. The analysis attempted to combine these different texts rather than treating them separately based on "online" (1 and 2) and "offline" (3). Crucially, no hierarchy was implied among the different texts; the three types were treated equally in terms of their contribution to the data analysis. In addition, rather than organizing the analytical discussion by kinds of data, and the information elicited from each, I organized it by three thematic dimensions that characterized what I described as participants' "storytelling online". The different kinds of data (twelve face-to-face interviews, twenty-eight e-mail accounts and one letter, and various texts from breast cancer websites) were coded according to the three dimensions. When analyzing the data, the aim was to identify participants' understandings of their online experience in relation to each of the three thematic categories. I looked for the different manifestations, as well as absences of each of the three aspects in patients' accounts (e-mail and face-to-face) and in texts on breast cancer websites (e.g. a website's instructions for how to post a message). In reading the various texts, I asked myself: what do they say is significant about the exchanges? What do they emphasize and what do they omit or understate? What is surprising about what they say about their online interactions? What is problematic? In light of these questions, I examined differences and similarities between the different sets of data, and tried to make sense of them.

I used discourse analysis of the website texts to contextualize patients' accounts (both e-mail and face-to-face) of their illness and online experience, and vice versa: patients' accounts of their illness experience and Internet use were used to make sense of breast cancer websites' texts. For example, a common feature of the face-to-face interviews was that participants understated or even denied their participation in exchanging personal stories online, whereas examination of their e-mail accounts and observation of the websites they visited showed that, often they

were quite actively engaged. Also, the face-to-face and e-mail accounts were produced for me, the researcher: they were the stories of these patients' online experience in relation to their illness. The online texts taken from breast cancer websites, on the other hand, were stories about the experience of illness, and coping with it, produced by patients and posted for their online fellow-sufferers. The significantly different audiences had crucial implications for the content and form of these texts, an issue I took up in analyzing the data. Another difference between the data obtained online and that gleaned from face-to-face interviews derived from their timing: the online accounts women posted on websites were often created when they were going through the illness and undergoing treatment. The e-mail accounts they wrote for me were often still temporally close to their actual experience (since I recruited interviewees from the websites where they posted their stories, usually close to the time of posting). The face-to-face interviews, however, were mostly conducted at least a year later than this. Naturally, women often had a very different perspective of their experience of illness and, inextricably, of Internet use. For all those reasons, it appeared crucial to integrate the different kinds of accounts and perspectives from the different sets of data into an understanding of the communicative context that I studied.

Fundamentally, in reading and analyzing women's accounts, my aim was not to evaluate whether they were 'truthful' or not. Rather, the aim was to get enhanced understanding of women's experiences of using the Internet in relation to their illness. So, for example, a woman told me in a face-to-face interview that the Internet played a very limited role, if any, in her experience of coping with breast cancer. However, this was contradicted by the online data I gatherer, which included her various postings and revealed her rather active participation. There was also an online account she wrote me two years earlier in reply to the recruitment message I posted on one of the breast cancer boards, in which she recounted her use of the Internet and its significance as a tool for information seeking and a space for support. How do you reconcile such differences between the same person's accounts? The principle that guided me is rooted in the interpretative approach to life stories (Plummer, 2001): all autobiographical memory is true. When people talk about their lives they inevitably forget, select, exaggerate, become confused and lie sometimes. It is the interpreter's task to identify these gaps and discern their meaning. My interpretation of the case I cited above was that the face-to-face interview, which took place a year after that

women was already cured, was part of her attempt to construct herself as a healthy person. She associated her online participation in breast cancer forums with her illness, a chapter she wanted to forget and put behind. She therefore tended to marginalize and almost dismiss the significance of this chapter, and the Internet's part in it, in her life. Marginalizing the role the Internet played in coping with the illness was a recurring phenomenon in women's accounts. When asked to reflect on the place of the Internet in the experience of their illness, interviewees often depicted their online experience as insignificant. I found that it was part of patients' (often unconscious) attempts to construct themselves as the key exclusive agent who 'made it', reducing the role of other possible factors (such as the Internet) that might have helped them in their battle against cancer.

The more general point I wish to draw from this example is that in their analysis, researchers should try to use the different kinds of data as mutually contextualizing each other. There is a tendency, as Slater (2002) rightly observes, to interpret the relationship between online and offline as the relationship between phenomenon and context. Hence "the offline is treated as that which makes sense of, or explains, the online" (p. 544). As far as the analysis is concerned, such an interpretation implies that one should use offline data in order to make sense of and contextualize the online data. The danger here is that "Putting the online into the offline reifies both: it assumes a thing called society, or community, or social relations, and at best investigates how one affects the other" (ibid.).

The offline does not explain the online, nor does the online explain the offline. Rather, the aim should be to look at the ways in which each configures the other. As I have shown, in order to understand patients' online activities in my study I had to get to grips with their offline contexts, that is, their coping with breast cancer. The opposite, however, was also fundamental: to make sense of patients' experience of breast cancer, and of breast cancer culture in contemporary society more generally, I had to understand the significance of their engagement in related online spaces. Therefore, the online data served to contextualize the offline data and vice versa.

Yet it can be argued that a qualitative research project could aim at comparing online and offline manifestations of a certain context. In this case, to fit the method and the analysis to the research question, it appears most sensible to treat the online and offline data in a comparative fashion, analyzing one against the other. Early CMC research focusing on the 'cues-filtered-out' approach employed experimental studies

of small groups to compare face-to-face and computer-mediated group behavior (for a review of these see Baym, 2002: 63-64). Discourse and linguistic analysis have often been used to compare CMC discourses and offline discourses, oral or written (e.g. Baron, 2003). The assumption underlying these comparisons has often been that CMC is a constrained version of face-to-face embodied interaction. However, this is highly problematic view. Theoretically, it fails to recognize CMC's unique and varied qualities, understanding how users draw on their existing communicative capabilities to construct social meaning within the challenges and the opportunities posed by the online medium Baym (2002: 66). Consequently, an analysis which takes the face-to-face as its starting point is unable to explain the specificity of the online phenomenon it aims to study; it can explain what is going online only in terms of face-to-face qualities. In my study, if I were to treat the face-to-face as a conceptual and analytical point of departure for understanding the online communication of breast cancer patients, I would have been unable to acknowledge qualities that are central to this communicative context, such as anonymity and disembodiment, without implying that they represent an 'inferior' version of patients' face-to-face communication.

Normatively, regarding online communication as a constrained version of face-to-face communication implies that online communication is 'less' than face-to-face communication: less authentic, less 'real', less close, and less truthful. Methodologically, treating the online as a constrained version of the offline limits the tools and practices that researchers use to those that they can apply to the offline. It does not allow researchers to develop methods that are sensitive and specific to what happens online. In my study, for example, if I were guided by a need to compare the online to the offline I would have probably been unable to analyze and account for the significance of discursive forms such as threads – which do not have straightforward face-to-face parallels.

This is not to say that comparing online and offline data cannot yield interesting and important observations about the qualities of CMC. However, one needs to carefully account for the underlying theoretical and conceptual framework that invites such comparative treatment of the data in the first place.

Whether online and offline data are used in the analysis in an integrative fashion, or in a comparative way, a key thing to be wary of is making judgments about the authenticity of the data. There is often a tendency to imply, explicitly or implicitly, that the information the researcher garnered from online sources (e.g.

websites, CMC interactions) is not as authentic as that generated from offline ones. In treating online and offline data, we should be informed by recognition of the distinct character of online and offline contexts and interactions, and of their consequent texts, while at the same time accounting for the inextricable connections, similarities and continuities between the two.

How to present our interpretations?

Lastly, an important issue of concern in the construction of our analysis is the presentation of the data. Do we differentiate online from offline data, or do we present it as a coherent set of data? This may seem a technical and rather marginal aspect, but it constitutes a significant feature of the treatment of data. In reporting my study, I used different fonts to reflect the different sources from which I was quoting: (1) academic or any other published text which is not a direct part of the ethnographic material; (2) extracts from face-to-face interviews used to build my analysis; and (3) extracts from online texts, whether e-mail accounts from participants or texts posted in public online forums. I was challenged by one reader whose criticism was that in identifying the different online and offline sources by different fonts I was not acting in line with what I was advocating; i.e. that the online and the offline should be seen and treated as significantly separate rather than interwoven. While I do not think that either point of view is right or wrong, I do think that whatever decision researchers may make about differentiation, they should be reflexive of its possible implications. In my case, my decision to use different fonts was made to help the reader identify the different sources of the quotes (especially given the prevalence of quotes in my analysis), but in so doing, my intention was certainly not to imply that the online and the offline should be or were being treated as two separate or isolated realms (see also Markham, 2004 for further discussion of issues of presentation).

Our responsibility to reflexively interrogate our methods carries through all stages of research design, analysis, interpretation and presentation of findings. This applies whether researchers are relying on offline data, online data, or both. In the same way that researchers using both online and offline data are encouraged to reflect on their choices, researchers should carefully reflect on the merits and limitations of building an account that draws only on one kind of data. They should consider the implications of their use of data for the claims they make about the Internet-context that they are studying.

Pitts (2004), for example, demonstrates a level of reflexivity in her study of personal web pages of women with breast cancer. Unlike my study, Pitts analyzed only online data, namely the texts of fifty personal web sites of individual breast cancer survivors. In presenting her analysis of these data she reflects upon their limits:

> I can make no claims about the off-line identities of the authors who wrote the web sites, and I do not assume that cybersubjects' on-line identities are necessarily identical to their off-line identities…
>
> I operate under the assumption that the web pages are in some sense 'truthful', in that their authors do indeed have breast cancer or know someone with breast cancer…
>
> That this assumption is not empirically verified must be considered a limitation of this research (p. 40).

While researchers should be encouraged to reflexively interrogate their methods and analyses, I think that Pitts actually falls into the trap that I discussed earlier, that is, of treating online data as less authentic or truthful than offline data. Pitts seems to work with some absolute notion of offline data as inherently more "truthful" or "verified" than online data, and hence to judge the online data upon which she bases her analysis as limited and probably less authentic than their offline counterparts. Rather, standards of authenticity should be seen as situationally negotiated and sustained (Hine, 2000, p. 49). In this sense, Pitts's later reflection on her decision not to look for offline data on her informants seems more context-sensitive and sensible. It demonstrates an understanding of the perceptions of her informants and the judgments *they* make about the online spaces in which they participate:

> I believe that this would go against the spirit of personal web pages, which are intended to be public but also to afford varying levels of anonymity and a choice about making personal disclosures, such as one's real name, location, appearance and so on, to readers (p. 41).

In short, whether the analysis is based on both online and offline data, or only one kind of data, the question of the authenticity, validity and adequacy of the analysis is one that the researcher has to face, critically and reflexively.

Conclusions: Revisiting the online/offline distinction

The key argument in this chapter is that in thinking through their own research projects or evaluating others', researchers need to critically consider the data that they obtain and interpret. It is not enough to recognize the complex nature of the relationship between the online and the offline at a conceptual level, while ignoring its

methodological implications. I find it striking that researchers make claims about the immersion of users' experiences and practices in their everyday lives, while the data they rely upon provide them with very limited grounds to adequately understand the relationship between their participants' online and offline worlds. This does not, as I have stressed, mean that it is only through offline data that researchers can access respondents' everyday lives. Online data such as, for example, online self-narratives, can be rich, detailed, and illuminating about information that could have not been gained using offline methodologies. The key point is that the data on which researchers build their analyses, whether these are online, offline, or both, should be high quality. That is, the data should be collected and generated after solid preparation based on a clear rationale; should fit with the question and the context; should convincingly support the claims being made; should be used reflexively and be context sensitive, and finally, should be ethically grounded.

Throughout the course of the research project, researchers must ask themselves such questions as: does obtaining online and offline data fit the questions I'm asking and the context I'm studying? Would offline data reveal something significant about the context being studied that could not be obtained from online data? In what ways might the offline data enhance the interpretation of the online data? It must also be borne in mind, as I have argued, that combining online and offline data is not always an appropriate decision. It might be insensitive to the context being studied, it might involve problematic ethical consequences, or it might simply be impractical. Thus, an equally important set of questions includes: can I make a persuasive case with only one of the two kinds of data? What might possibly be lost or risked in obtaining the two kinds of data? It must be remembered that moving research relationships from online to offline and, more generally pursuing offline data to complement online data, can certainly open up research paths, but could equally be counterproductive and close off research routes (Eichhorn, 2001, p. 571).

Perhaps we should revisit the distinction between online and offline data, and reconsider its usefulness. The increasing trend toward the convergence of technology, particularly the mobility of wireless technologies such as SMS, is already blurring the line between online and offline communication (Herring, 2004, p. 33; Slater, 2002, p. 544). The term "online" itself does not map consistently into a single media technology. The mediascape becomes more hybrid and multi-layered, and "virtuality"

is not restricted to being online, but can embrace and link several media (Slater, 2002, p. 544), including what we would once have considered as "offline media", for instance the telephone. Furthermore, traditional modes of CMC are increasingly being used in order to establish face-to-face contacts (Herring, 2004, p. 33).

These changes challenge the methodological distinction between online and offline data, with which I opened this chapter, in significant ways. How, for instance, do we define away messages in Instant Messaging that enable their users to be physically mobile in their physical absence from the computer? [see Baron et al.'s (2005) recent study]. And what are the implications for our research and our analysis, as the data become even messier and less stable? This line of questioning connects to a recent discussion on the need to move beyond the concept of place-bounded ethnography, and redefine the field and its boundaries (Eichhorn, 2001; Hine, 2000; Leander and McKim, 2003). Influenced by ideas such as Marcus's (1995) "multi-sited ethnography" and Olwig and Hastrup's (1997) view of the field as being a "field of relations", qualitative Internet researchers are looking for ways to move beyond bounded sites, to follow connections made meaningful from a specific setting (Hine, 2000, p. 60-61). For example, in their discussion of methodological approaches to the analysis of adolescents' Internet literacy practices, Leander and McKim (2003) propose replacing the notion of users' everyday "sites" by that of "sitings".[3] They emphasize the need to develop methodologies that follow participants' practices of moving and traveling between online and offline, and within a far wider and hybrid mediascape.

However, even if the line between "online data" and "offline data" is blurring, the issues discussed in this chapter still have relevance for any researcher who is thinking through a qualitative Internet project, or evaluating that of another. Take, for example, Baron et al.'s (2005) recent study of how away messages in instant messaging are used by American college students to help manage their social spheres. One set of data the researchers collected consists of 190 away messages. This might not simply map onto what we have been used to think of as "online data". It has similarities with the kind of textual data that was studied in traditional CMC environments, but at the same time has distinctive "non-online" features. The other kind of data the researchers used is traditional "offline data", derived from interviews and a focus group with users. Although the distinction between online and offline data is not fully relevant to this research, some of the key issues that I discussed in this

chapter in relation to this distinction, may still arise and be relevant. For instance, the question of triangulating different sets of data; using the face-to-face interviews with participants to contextualize their instant messaging practices; and vice versa: using the data of the instant messaging to make sense of what respondents said in their interviews.

"Online social worlds are accessible to researchers in ways that few other worlds are. If we want to understand them, we need to look with rigor and detail" (Baym, 2000, p. 198). Looking with rigor and detail may mean adopting very different methodological strategies, and taking very different decisions in the course of the research project. In this chapter I have sought to discuss some of the questions, dilemmas, strategies and decisions that may be involved in grappling with aspects of online and offline data in qualitative Internet research. While there are no right or wrong answers to any of the issues and the questions discussed, what is important is that the decisions made should be grounded in the particular context being studied, and the specific questions being asked.

List of resources

For a collection of case studies and reviews that explore methodological solutions to understanding the social interactions mediated by information and communications technologies, see Hine's (2005) edited book *Virtual Methods: Issues in Social Research on the Internet*. For particular discussions on the question of online and offline data see chapters by Mackay, Sanders, Orgad, and Rutter and Smith.

For a critical review of key epistemological, conceptual and methodological aspects related to the relationship between online and offline see Slater's chapter on Social Relationships and Identity Online and Offline in L. Lievrouw and S. Livingstone (eds) (2002), *The Handbook of New Media*. London: Sage. pp. 534-547

For an ethnographic study of the Internet, that offers a sophisticated analysis of the online/offline relationship in a situated context (Trinidad), and draws on rich ethnographic online and offline material, see Miller and Slater's (2000) *The Internet: An Ethnographic approach* (London: Berg).

For a collection of reflexive reports and short essays on researchers' experiences of doing qualitative Internet research, including some discussions of issues of online and offline data, see Johns, Chen and Hall's (eds) (2004) book *Online Social Research: Methods, Issues & Ethics* (New York: Peter Lang).

References

Bakardjieva, M., & Smith, R. (2001). The Internet in everyday life: Computer networking from the standpoint of the domestic user. *New Media & Society*, 3 (1), 67-83.

Baron, N. (2003). Why email looks like speech: Proofreading, pedagogy, and public face. In J. Aitchison and D. Lewis, (eds), *New Media Language*. London: Routledge. pp. 102-113

Baron, N. S., Squires, L., Tench, S., and Thompson, M. (2005). Tethered or Mobile? Use of Away Messages in Instant Messaging by American College Students. In R. Ling and P. Pedersen (eds). *Front Stage – Back Stage: Mobile Communication and the Renegotiation of the Social Sphere*. Springer-Verlag.
Available at: http://www.american.edu/tesol/Grimstad-Baron.pdf (last accessed 28 January 2005).

Baym, N. K. (2000). *Tune in, Log on: Soap, Fandom, and Online Community*. London: Sage.

Baym, N. K. (2002). Interpersonal life online. In L. A. Lievrouw & S. Livingstone (eds), *Handbook of New Media*. London: Sage. pp. 62-76.

Danet, B., Ruedenberg-Wright, L. and Rosenbaum-Tamari, Y. (1997). "HMMM…WHERE'S THAT SMOKE COMING FROM?": writing, play and performance on Internet relay chat. *Journal of Computer-Mediated Communication*, 2 (4). Available at: http://jcmc.indiana.edu/vol2/issue4/danet.html

Donath, J. S. (1999). Identity and Deception in the virtual community. In A., M. Smith and P. Kollock (eds), *Communities in Cyberspace*. London: Routledge. pp. 29-59.

Eichhorn, K. (2001). Sites unseen: Ethnographic research in a textual community. *Qualitative Studies in Education*, 14 (4): 565-578.

Haythornthwaite, C., and Wellman, B. (2002). The Internet in everyday life: An introduction. In B. Wellman & C. Haythornwaite (Eds), *The Internet in Everyday Life*. London: Blackwells. pp. 3-41

Herring, S. C. (2004). Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society*, 6 (1): 26-36.

Hine, C. (2000). *Virtual Ethnography*. London: Sage.

Howard, P. E. N., Rainie, L. and Jones, S. (2001). Days and nights on the Internet. *American Behavioral Scientist*, 45 (3): 382-404.

Kendall, L. (2002). *Hanging Out in the Virtual Pub: Masculinities and Relationships Online*. California: University of California Press.

Kivits, J. (2005). Online interviewing and the research relationship. In C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet*. Oxford: Berg. pp. 35-50

Leander, K. M., and McKim, K. K. (2003). Tracing the everyday 'sitings' of adolescents on the Internet: A strategic adaptation of ethnography across online and offline spaces. *Education, Communication & Information*, 3 (2): 211-240.

Lenhart, A. (2001, 21 September). Who's not online: 57% of those without Internet access say they do not plan to log on.
Available at: http://www.pewinternet.org/pdfs/Pew_Those_Not_Online_Report.pdf (last accessed 28 January 2005).

Mackay, H. (2005). New connections, familiar settings: Issues in the ethnographic study of new media use at home. In C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet*. Oxford: Berg. pp. 129-140.

Mann, C. and Stewart, F. (2000). I*nternet Communication and Qualitative Research: A Handbook for Researching Online*. London: Sage.

Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology*, 24: 95-117.

Markham, A. (2003, 19 October). *Broadening Options and Raising Standards for Qualitative Internet Research: A Dialogue Among Scholars*. Roundtable conducted at the Association of Internet Researchers 4th annual convention, Toronto, Canada.

Markham, A. (2004). Internet communication as a tool for qualitative research. In D. Silverman (ed.), *Qualitative Research: Theory, Method and Practice*. London: Sage. pp. 95-124.

Olwig, K. F., and Hastrup, K. (eds) (1997). *Siting Culture: The shifting anthropological object*. London: Routledge.

Orgad, S. (2005). Moving from online to offline relationships with research participants. In C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet*. Oxford: Berg. pp. 51-65

Orgad, S. (2005). *Storytelling Online: Talking Breast Cancer on the Internet*. New York: Peter Lang.

Pitts, V. (2004). Illness and Internet empowerment: Writing and reading breast cancer in cyberspace. *Health*, 8 (1), 33-59.

Plummer, K. (2001). *Documents of Life 2*. London: Sage.

Reid, E. (1999). Hierarchy and power: social control in cyberspace. In A., M. Smith and P. Kollock (eds), *Communities in Cyberspace.* London: Routledge. pp. 107-133.

Rutter, J., and Smith, G. W. H. (2005). Ethnographic presence in a nebulous setting. In C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet.* Oxford: Berg. pp. 81-92.

Sanders, T. (2005). Researching the online sex work community. In C. Hine (ed.), *Virtual Methods: Issues in Social Research on the Internet*. Oxford: Berg. pp. 67-79

Schaap, F. (2002). *The Words that Took Us There: Ethnography in a Virtual Reality*. NJ: Transaction Publishers.

Slater, D. (2002). Social relationships and identity online and offline. In: L. Lievrouw & S. Livingstone (Eds), *The Handbook of New Media*. London: Sage. pp. 534-547.

Turkle, S. (1996). *Life on the Screen: Identity in the Age of the Internet*. London: Weidenfeld & Nicolson.

Wilson, B., and Atkinson, M. (2005). Rave and Straightedge, the Virtual and the Real: Exploring On-line and Off-line Experiences in Canadian Youth Subcultures. *Youth & Society*, 36 (3): 276-311.

---

[1] As I mention later in the chapter, Bakardjieva and Smith's (2001) study included only one component of online data, which they describe as a tour of users' "computer space".

[2] An extract from an e-mail account of one of my research participants.

[3] In making this proposal, Leander and McKim (2003) are particularly inspired by the work of Olwig and Hastrup (1997).