



Personal documents on the Internet: what's new and what's old

Laura Arosio ¹

Abstract

In this article some of the main methodological issues that arise in doing research using the Internet are presented. We firstly distinguish two main methodological areas: sociology with Internet and sociology on Internet. This work is mainly devoted to the sociology on Internet.

We suggest that research on Internet has features which are highly continuative with documentary analysis, a type of enquiry which is already part of the tool box of social researchers.

Some on line documents in particular may rightfully be considered as personal documents in traditional sense: e.g. blogs, e-mails, personal sites, forums, albums on line. A possible matching between traditional personal documents and web-mediated personal documents is proposed.

It is certainly true that personal documents on line reveal specific features. In particular, since they are published on the web, “personal” documents are permeated with a “public” dimension which makes their significance distinctive. Specific stratagems are also necessary in the collection, contextualisation and analysis stages.

In this paper some of the theoretic, ethic and technical issues that arise from the use of web mediated personal documents are discussed. This article contains no solutions, but rather a list, a sort of reminder, of the open questions. Nor is it possible to provide answers to many of the questions submitted: due to their complexity and relative newness, the scientific community has not yet given clear and shared answers to these questions.

Keywords

Documentary analysis, personal documents, social research using the Internet, web-mediated personal documents, blogs, forums, albums on line.

¹ University of Milano Bicocca – Italy, laura.arosio@unimib.it

1. Introduction

The Internet is a very rich source of materials (texts, photographs, videos, songs, audio tracks, etc.) dealing with a wide variety of subjects. Each of them narrates a fragment of the world of those who produced or commissioned them: their tastes, ideas, values, interpretations and points of view on reality. In their turn they reflect the values, practices, opinions and cultural meanings of the group, and at most of the entire society as a whole in which the individuals are placed.

Social researchers are realizing the great potential of the web and are preparing to use it in their work. We consider it increasingly likely that in the future this type of study will enable us to obtain an image of our society.

Before using the Web as a tool for social research, it however appears opportune to clarify the question of *meaning*. What does the use of the Internet signify? What contribution may it give to social research? In what relationship of continuity or discontinuity does it collocate with traditional methods of research? An important issue in research making use of the Web is to define, or rather set limits to the questions it may answer.

We may firstly distinguish two main methodological areas, according to the use made of the Web: 1) sociology *with* Internet, 2) sociology *on* Internet.

1.1 Sociology with Internet. Sociology with Internet makes use of the Web as a research tool to collect information. Some examples are *focus groups* on line, interviews (qualitative and quantitative) carried out on the web, observation in Internet². Making sociology *with* Internet means utilizing a technologically advanced variant of the conventional tools for data collection, or also, and this depends on the points of view, a “new” data collection tool although within conventional research methods. To make an example, a focus group conducted on line may be seen as a different variation of the traditional focus group, with which it will therefore share strong features of continuity. At the same time, it is evident that the mediation of Internet techniques requires a series of particular skills and precautions. Again on the subject of focus groups, when they are conducted on line interaction between participants is not direct but mediated by the Web and this has significant repercussions on the development of the discussion. According to some authors (e.g. Greenbauman 2008) online focus groups, especially if conducted through *chat lines* or *forums*, may reveal these and so many other features differing from the traditional version that they cannot be considered as real focus groups, but completely updated tools for data collection. Sociology with Internet however does have some advantages which must be assessed³. It firstly reduces the problems linked to geographical distances, limiting necessary journeys and moves to reach the subjects of the research, and makes it possible to easily include in studies subjects residing in all parts of the world (also in situations in which they would be otherwise difficult to recruit). Research costs are therefore reduced in terms of money

² For an introduction to these various techniques, see, among the others, Dillman *et al.* (1999 and 2001), Hine (2008), de Lillo (2010).

³ On the opportunities and limits of researching with Internet, see for example Cohen *et al.* (2000).

and time. From the point of view of research materials, sociology with Internet makes it easy to utilize supports such as images, videos and audio tracks. Moreover, it remains a faithful documentation, an actual transcription of what occurs on line. The *problems* of research with Internet include the absence of direct interaction between the researcher and those taking part in the research and between the participants themselves; it makes it impossible to control the identity of those taking part and the context in which the data collection is carried out, and it limits the sharing of significant in that it does not make it possible to access non-verbal language (Greenbauman 2008). In conclusion, sociology with Internet is still in its trial period, and its use in research needs to be assessed with caution but with interest, seeing the great potential it appears able to offer.

1.2 *Sociology on Internet*. Sociology on Internet makes use of material stored (archived) on the web and treats it as an object for study in order to track down significant and elements indicative of culture and social relations. In particular, the material studied is not solicited by the researcher (as instead happens with sociology with Internet), but spontaneously produced by web users and for different purposes from those of the research. Only subsequently, when the material has been collocated and spread in Internet, may researchers have access to the data entrusted to the web and make it the subject of their enquiries. It is clear that sociology on Internet, with which we intend to concern ourselves in this paper, may to all intents and purposes be considered a particular case of *documentary analysis*, a type of enquiry which is already part of the tool box of social researchers, and therefore has to re-start from this basis.

2. Documentary analysis

The material stored on the Internet by web users (with different purposes, but in any case independent of social research) may be considered as “documents” and their use in social research is part of the so-called *documentary analysis* (for an introduction see Scott 1990, Prior 2003, McCulloch 2004, Scott 2006, Arosio 2010).

By *documents* we refer to different kinds of material which individuals and groups produce spontaneously in carrying out their activities, therefore without any link with the research goal (unlike the material produced, for example, during an interview, which, whether it be qualitative or quantitative, individual or group material, is generated in response to the stimulus of a researcher and hence with goals closely linked to the research).

There are many types of documents in existence, produced on different occasions, by different authors and for different purposes. A first distinction may be made between institutional documents, personal documents, media documents, and cultural documents (Arosio 2010). Each of these documentary categories offers the researcher a different point of view on reality. Institutional documents (regulations, minutes, registers, balance sheets, rulings, laws, speeches...) present the official version which institutions and groups give of themselves and the situations involving them. Personal documents (letters, notes, memos, reminiscences, messages, postcards, private photo and videos,

confessions, diaries, autobiographies, spiritual testaments, writings for private ends...) contain the summary of a personal experience, the interpretation given by the authors of their own world, whose image they wish to communicate. Media documents (messages conveyed by the press, radio, cinema, television and the new media) utilize a repertoire of shared concepts, images and arguments and often anticipate norms, values and behaviours which will be spread in society. Cultural documents (literary works, visual, plastic or performing works; but also fairy stories, tales, nursery rhymes, songs, legends, metropolitan legends...) are interpretations of the world through the lens of belonging to a specific social context. Visual documents (e.g. photographs, paintings, films, advertising billboards, posters, murals) require specific consideration, as they do not constitute a category in themselves in that they closely overlap the previous ones. Photographs taken for private ends should be considered as personal documents; television programmes or advertising campaigns appearing in newspapers belong to mass communication, while paintings belong to artistic production or popular culture...

Documents were highly appreciated and used by the classical authors of sociology. Emile Durkheim, Max Weber and Karl Marx, for example, used them widely in their works⁴. Although its use has fluctuated over the years, documentary analysis has never been abandoned, also because it has been greatly employed in combination with other research techniques. In ethnographic studies, for example, together with observation and interviews, researchers very often utilise the study of documents produced in the contexts they are examining.

As pointed out above, an important feature shared by all these types of documents is that they have been produced for different purposes from social research and the researcher therefore has not taken part in the process of creating the material for study. Documentary analysis is therefore part of the “unobtrusive” research process (Webb *et al.* 1966). This term indicates research strategies which do not involve a *direct* demand for information on research subjects, and so the researcher is not involved in the data production phase⁵. Unobtrusive methods, also called “non reactive” methods, thus limit some of the problems characterizing the direct questioning and open observation (when a subject is aware of taking part in a research study and modifies his/her verbal and nonverbal behaviour⁶). It must however be recalled that by making use of these techniques, the researcher and his/her subjectivity also strongly influence the data collection, the analysis and more generally the numerous choices which social research demands.

An important issue in research studies making use of documents is defining the question they may answer. What can documents tell us? What knowledge do they bring us? When can we use them? Through documentary analysis the researcher may obtain information on a segment of social reality. Through documents we may discern the ideas, norms, values, interests, and ways of thinking of their producers or commissioners. In

⁴ See the reconstruction made by Macdonald & Tipton (1993).

⁵ On unobtrusive research methods, see Lee (2000).

⁶ The so called “Hawthorne effect”.

other words, documents help the researcher to perceive the way in which authors interpret and present their experience and the world in which they live⁷. Documents are in fact “situated products”, which are inevitably created in a certain context whose influence they feel (Prior 2003). From the elements contained in the material in question the researcher will seek to deduce the values, behavioural models, ways of thinking and arguing which bear traces of the context in which the documents took shape.

Using documents in social research makes several transitions mandatory: particular attention should be paid to the collection, contextualization and analysis phases. Collection is a central moment in documentary research. The documents must be accessed and questions then asked relating to “selection”: understanding exactly how much and which part of the existing documentation the researcher may manage to take possession of. In the contextualization phase researchers have to understand the specific features of the documents they are working on: what they are, in which contexts they were produced, by whom, for whom and for what ends. They also have to understand the cognitive contribution of each type of document: this means reflecting on which version of reality the various documents may give (the official version, the private version...). Various options are available for analysis and, to make but a few examples, they may range from discourse to narrative analysis, from the metaphors to that of conversation analysis. We may recall two strategies in particular which are fairly widespread and may be applied both to texts and to images: *content analysis (qualitative or quantitative)* and *structural analysis (semiotic)*. Qualitative content analysis, also called thematic analysis, traces a “list” of themes communicating the variety of documentary contents and is mainly aimed to identify the themes and analytic categories which express them⁸. Quantitative content analysis breaks down the documents into categories previously identified by the researcher and focuses mainly on text/image description, the identification of relations between its constituent parts, and on verifying the hypotheses formulated on the communicative facts it contains⁹. Structural analysis seeks definitions, presentation techniques, structuring rules of the document through which the author expresses his/her message. The analysis dwells on the people appearing in the stories (told in the form of texts or images), on the way in which they interact, and what they represent. It also takes into account elements factors such as settings, narrative sequences and the unfolding of the events¹⁰.

⁷ Not only what is contained in the documents, but also what they do not contain indicates a process of selection and choice mirroring the context in which the documents have been produced.

⁸ On qualitative content analysis see, among the others, Thomas & Znaniecki (1918-20), Kracauer (1952), Mayring (1988), Altheide (1996).

⁹ On quantitative content analysis see, among the others, Lazarsfeld *et al.* (1944), Berelson (1952), Krippendorf (1980), Weber (1990).

¹⁰ See the works of Ferdinand de Saussure (1916), Algirdas J. Greimas (1970, 1983) and Roland Barthes (1964, 1982), as a starting point of this perspective.

3. Personal documents

On opening an observation window on personal documents, they may be seen to be texts, images, videos or other types of material, produced personally by individuals for private ends. The content of these documents are personal experiences interpreted by the author, accounts of a personal nature (Plummer 1983). In personal documents events are narrated, facts and persons are cited and thoughts and ideas are reported. In a way, they outline the limits of their world: who it includes, how it is made, how it is experienced and narrated. Producing a document, whatever its format, is a way to impress order on and give a meaning to one's experience, and for this very reason personal documents enable the researcher to access the meanings attributed by the individual.

In order to understand the message they contain, personal documents should be contextualized (see above). It is particularly important to consider the figure of the recipient: one or more persons to whom the document is addressed. In some situations (such as letters, for example) the receiver is clearly identified. In other cases, when the divulgation of the document is not planned (as with private diaries or autobiographies), the figure of the addressee still exists, even if is not clearly defined. The author preserves his/her thoughts in images and words for them to be conserved and handed down (we may even state that the addressee of the document is the writer him/herself). The content of the document is therefore the outcome of the interaction between sender and receiver, a product arising from the author's reflections, which permeate it.

As outlined previously, personal documents may take on several forms (and be expressed through various languages): letters, notes, memos, photographs, messages, memoirs, confessions, diaries, autobiographies, testaments, as long as they are drawn up personally by individuals for private ends (a business letter is not a personal document). Despite the variety of formats personal documents may have, social research has dwelt above all on letters, diaries, photographs and family videos.

3.1 *Private letters* are addressed to relatives, friends and acquaintances for personal ends¹¹ (often only the sender's letters are available; when we also have their answers, we may speak of "correspondence"). They contain the author's account of facts, events, situations, comments and thoughts. Of particular interest for the researcher are collections containing a large number of letters, covering a more or less lengthy period of time.

3.2 *Diaries* are daily accounts of the author's experience, drawn up on a regular basis for a long period of time (Girard 1963; Lejeune 1989; Alaszewski 2005). A characteristic of the diary is the *simultaneity* of the writing: since no time elapses between the event and its coverage, the diary captures its narration "in real time". Diarists do not know what will happen in the future - and in the next pages (Lejeune

¹¹ They may even be written to newspapers and public personages, although in this case the nature of the letter is not completely private.

2009)- and do not know how it will end the story (unlike what happens to those who write autobiographies).

3.3 *Autobiographies* are the account of an entire life experience or even a limited portion of a life, drawn up personally by the protagonist. The writing up of the account takes place in a limited period of time (and in this respect it differs from the diary) and for this reason is *retrospective* in nature. The question of memory is important: with the passing of time the author tends to forget some passages and emphasize others, so that the returned account is the outcome of re-invention and re-interpretation embedded in time (Ruspini 2002).

3.4 *Photographs* and “*family*” *films* are collections of photographs and videos produced by people in everyday life events (Chalfen 1997). They tell the story of a person, of a family, of a group of individuals, to be preserved for themselves and to be shown to others.

4. Personal documents on line

As mentioned in the first pages of this paper, the new forms of communication spreading in the last few years, and in particular communication via Internet, have led to the creation of a growing number of documents. Like those the researcher is accustomed to dealing with, there are also various types of documents present in the Internet. Some *on line* documents in particular may rightfully be considered as “personal documents”: e-mails, text messages, forums, blogs, personal sites, on-line photo albums and private videos.

These objects of study, often presented as an absolute novelty, in reality have many affinities with traditional personal documents. Both conventional personal documents and personal documents *mediated by the Web* offer the researcher the opportunity to access material which is very rich in information, revealing the representational world of the persons producing them (but only if the researchers are able to understand their characteristics, limits and potential).

It is certainly equally true that personal documents on line reveal specific features to be studied and verified. In particular, since they are published on the web to be divulged, “personal” documents are permeated with a “public” dimension which makes their significance distinctive.

We will dwell on *web-mediated* personal documents which, in our view, correspond to the principle personal documents *in a traditional sense*; we will be dealing above all with e-mail messages (*letters*), forums (*correspondence*), blogs (*diaries*), personal sites (*autobiographies* and albums on line (*photographs and family videos*)). See table 1.

Table 1: Synthesis view: A possible matching between traditional personal documents and web-mediated personal documents

"Traditional" personal documents	"Web-mediated" personal documents
Letters	E-mails
Correspondence	Forums
Diaries	Blogs
Autobiographies	Personal sites
Photographs, family videos	Albums on line (photos/videos)

4.1 *E-mails*, messages mailed through the web by means of apposite programmes, may be considered the electronic equivalent of traditional letters. We can summarize the main features shared by e-mails and traditional mail. Both in electronic and in traditional mails the relationship between sender and addressee is very important: this relationship is part of the process of construction of the communication's significant. In both cases the nature of the message is "asynchronous": the content is read by the addressee some time following the sending. It is a confidential communication, visible only to the addressee of the missive. Some important special features of the electronic mail compared with traditional mail are as follows. Communication via e-mail is extremely fast: the message is received instantly, may be read almost in real time by the addressee and may receive an equally swift reply. An e-mail may be addressed to several persons at the same time, and groups and mailing lists may be created. The e-mail message may easily be copied and sent to other persons, re-addressed and propagated even some time after the sending.

4.2 *Forums*, web sites hosting discussions between web users, may be considered the equivalent of correspondence between several people. Forums and traditional correspondence share the following features. The writing process continues over time and the nature of the relationship between sender and addressee is reciprocal. There can be many authors who are revolving around a central nucleus of writers and discussion themes. The various authors are often linked by a network of relationships (in forums the users are often habitual users and this may create a sense of virtual community¹²). In both cases the interaction takes place in an asynchronic way (see above)¹³. We list some important features of the forums compared with traditional correspondence. In forums each author may read all the contributions of participants, even when they are not directly addressed. The contents are public and visible to all surfers, so it is possible to merely read the contents of other users' messages (although forums generally require users to register before sending messages and in some cases even just to see them).

4.3 *Personal sites* are comparable with an autobiography, in that they are a set of web pages constructed as "shop windows" for the writers themselves and their story¹⁴.

¹² On "virtual communities", see Rheingold (1993).

¹³ Forums differ in this aspect from *chat lines*, in which the exchange of messages takes place in real time (synchronic communication tool). For this reason chat lines are more like an actual chat session.

¹⁴ Although generally much more synthetic and schematic, the personal profiles on social networks are in a sense analogous to personal sites.

The features common to personal sites and autobiographies are as follows. Both contain information which is mainly autobiographical or focused on personal interests and are organised directly by the protagonists. The contents are static, written once or updated with a very low frequency. Both in personal sites and in autobiographies there is no interaction between the author and the reader, at most an address is given to send messages to. In fact, the receiver is not clearly identified. Some particular features: personal web sites are accessible to all surfers at any moment, and the author may modify, delete or add material fairly easily.

4.4 *Blogs*, web sites personally managed by the authors (*bloggers*) where thoughts, opinions, information and news may be aired, may be described as web diaries. We list the main features shared by blogs and conventional diaries. In both cases the contents are constantly updated (“writing in real time”), the point of view of the authors is not retrospective and writers do not know what will be the development of their history. Both traditional diaries and blogs are written over a fairly long period and do not have a precise addressee. Some major features of blogs compared to traditional diaries are as follows. The contents of blogs are visible to all Web users; comments are invited from readers and may be followed up by the author’s replies: there is therefore a high level of interactivity. Visitors are often regular and create small virtual groups.

4.5 *On-line albums*, sites which offer users the chance to insert photographs and personal videos to preserve and share their own images, correspond to family albums and videos. Both on-line albums and conventional collections are produced directly by the protagonists and both offer images with the most varied contents, taken on various occasions in everyday life (travel, weddings, children, pets, etc.). Some important differences of on-line albums compared with traditional albums are as follows. On-line albums are generally visible to all surfers (but access keys may also be provided) and they have a low level of interactivity (they are not looked at “together” as in the case of traditional albums but are viewed without physical co-presence).

In summing up, we may attempt to by single out the main point of contact between traditional personal documents and web-mediated personal documents, and the particular characteristic which most differentiates them. We will start with points of contact. Like traditional ones, web-mediated personal documents are places where the author’s subjectivity is expressed, and there is ample room for personal reflections. We may find thoughts, points of view, definitions and interpretations of reality, which may therefore be of great use to social researchers.

On the subject of discontinuity, compared with traditional personal documents, web-mediated personal documents place a great deal of importance on the relational dimension of the communicative act. In letters, diaries, autobiographies and family videos an addressee is also present and the message in the documents is read in the light of the relationship set up between author and addressee. But in web-mediated personal documents the relational dimension is even more marked. Firstly, the messages are made accessible to a potentially vast audience. Moreover, users may often interact with the author and the very contents of the document are the outcome of the work of several persons. As soon as they are made public on the Web for divulgation, “personal”

documents become permeated with a “public” dimension, which makes their significance very special (see also Lejeune 2000; Sorapure 2008).

5. Continuity and unprecedented challenges

It should now be clear that the use of documents on Internet, and in particular of personal documents on line, reveals features showing a strong continuity with the past and with “traditional” styles of research, and at the same time confronts the researcher with unprecedented challenges.

Mediation of Internet techniques calls for a series of particular skills and precautions, which are reflected in all the research steps: not only (not so much) in the analysis as in the preceding access, selection and contextualisation phases (see above).

Net-mediated personal documents pose numerous theoretical and technical questions which, due to their relative novelty, have often not yet received answers from the scientific community. We will set out some of these questions, without seeking to find a solution to each one, but rather to place them on researchers’ agendas and to underline the need for them to be suitably dealt with.

5.1 *Data collection.* From the point of view of research operations, the questions of access and selection must be considered.

Different situations exist regarding access. On the one hand, private communications exchanged via Internet are confidential correspondence, and may not be easily accessed by the researcher. It will therefore be necessary to establish suitable strategies to succeed in obtaining this kind of material. The guidelines usually suggested for access to “traditional” personal documents may hold good here (see for example Bichi 2004). The collection of personal documents may be part of a broader research design: researchers may decide to select and contact some important persons for their research, interview them and when a situation of reciprocal acquaintance and trust has been developed, may attempt to ask them whether they have personal documents, for example an e-mail archive, and whether they are willing to make them available. Another strategy is to set up awards, competitions or other initiatives, perhaps supported by an institution able to create confidence, in which people are asked to send collections and personal documents. A third possibility is to access already existing archives. In several European countries there are important examples of archives of “traditional” personal documents (in Italy see for example, the Archivio Diaristico Nazionale (National Diary Archive) of Pieve Santo Stefano and the Archivio della scrittura popolare (Archive of Popular Writing) at the Trento History Museum). In the future the material collected in these archives may be enriched (or new *ad hoc* archives set up), also inserting letters and correspondence on line.

Blogs, personal web sites and on-line albums, on the other hand, generally provide for public access, and the material in these cases is easily accessible. Forums have a “half way” status between public and private, above all in cases where it is necessary to register in order to access discussion. It is therefore necessary to weigh up whether to present oneself as a researcher in the virtual community or to create a fictitious profile to

conceal one's identity. In any case, the use of web-mediated personal documents obliges the researcher to deal with some important *ethical questions* (in reality common in the use of "traditional" personal documents): to what extent is it permissible to use this material for research ends? Must the authors be acquainted with the use of their work? Must their permission be sought? Do information and sensitive issues exist which must be protected? (on ethical issues regarding internet based research see also Hine 2008).

As mentioned above, another issue concerns selection. The researcher must firstly be sure that the entire documentation on which he/she intends to work is available. A problem would arise if a part of it were systematically excluded from analyses, with consequent distortions of observations and the results obtained. When the volume of material is such that it reasonably enables it to be analysed in its entirety, the researcher will be able to work on the whole documentary corpus. When the material available is sizeable (e.g. personal blogs) a choice must be made. How to carry out a selection? Much depends on the objectives of the analysis, the research questions and the techniques which one intends to use. One must also take into account that the characteristics of the Web also create technical difficulties which are not easy to solve, and make sampling procedures complex.

A first possibility is to carry out a non-probabilistic strategy and here the use of a purposive type of sampling may be particularly useful. In this case the choice of material to be included in analyses is entrusted to the researcher and to his/her knowledge regarding the phenomenon to be researched. The most interesting cases for analyses are chosen, above all including typical cases, and differentiating positions (taking care to include "extremes")¹⁵. The breadth of the sample is also decided by the researchers, who may opt for the criterion of *saturation* (Glaser & Strauss 1967) in which the process of including new material is suspended when subsequent additions would not lead to an increase in knowledge.

In desiring to construct a probabilistic sampling, the main difficulty when working on material stored on line lies in the unavailability of a sampling list, from which to choose the cases to be included in analysis. To overcome this problem, two paths appear feasible: 1) the use of a search engine, 2) the use of a "snowball" based strategy of sampling. Both methods may be used to create a sampling list from which a sample may subsequently be extracted, and also to identify the sample itself¹⁶. A specific feature of the search engine is the fact that it produces a list whose results, in ultimate analysis, are ordered according to the importance/popularity of the material (from the most to the least-widely read). By the term snowball sampling we are here referring to a wide set of recruiting techniques which exploit the links between the research cases. As far as Internet research is concerned, some attempts have been made in this direction, using both *crawler* programmes (identifying the hypertextual links present in a document and

¹⁵ For a research example, see Paldacci 2003.

¹⁶ In both these situations the information obtained may be used to create probabilistic samples and to construct non-probabilistic samples. For some research examples see, among the others, Jung *et al.* 2005; Arosio & Ruspini 2010.

adding the addresses to the list it is creating; the process continues until it reaches a determinate number of links or according to different instructions given¹⁷), and using specific techniques such as RDS (Respondent Driven Sampling) (see for example Henzinger *et al.* 2000).

5.2 *Contextualization.* In the contextualization phase, the use of web-mediated personal documents demands specific attention.

A first observation concerns authors. The issue of different access capacity to the new technologies (*digital divide*¹⁸) must firstly be assessed. Although the use and management of programmes consenting the production of this type of document demand a low level of computer literacy, and the necessary programmes and services are generally provided free of charge, we must recall that while some population groups are able to easily utilize these tools, others are excluded (elderly people and those with a lower level of literacy are especially left out). Both the authors of personal Internet documents and their potential addressees will therefore be “selected” from the start, and will not be considered in the same way as those who have access.

Further reflection is necessary concerning those who create web contents. In many cases their identity is not known, since the authors are anonymous or use pseudonyms. It is also possible to create fictitious profiles providing an identity which is partly or wholly different from the “real” one. To deal with this snare, which must in any case be taken into account since it is difficult to eliminate, the researcher may set up research questions and strategies for analysis which do not dwell on the authors’ characteristics (for example, the idea of relating the bloggers’ socio-demographic declared characteristics and contents of their diaries seems fairly impractical, unless we consider the identities proposed on sites as not “real” but as “virtual” images of themselves). Researchers ought rather to look at the variety of the contents, their evolution in time, etc., without dwelling on the single authors.

An important issue also concerns recipients. We have seen that the distinctive feature of personal documents on the Web compared with traditional ones is that they are open to the public. When they are placed on the Web, the materials may be seen by a very large number of users, who are also unknown to the author (e-mails are an exception, since they maintain a strong link between sender and addressee, as are materials whose access is limited to a small number of users, see above). We must also consider the fact that “virtual groups” tend to be created around forums and blogs, linking the authors and users of the message. Their significance must be seen in the light of this relationship.

We must also be careful in assessing to what extent documents are truly published on line “for personal ends”. Some materials in fact happen to be apparently divulged by single individuals for the purpose of mere personal expression, while a more attentive

¹⁷ http://en.wikipedia.org/wiki/Web_crawler [Accessed 24th April 2010]

¹⁸ The term *digital divide* indicates the division between those who have the possibility/ability to access new technologies (Internet in particular) and those who are excluded (see for example Norris 2001).

examination reveal that they are a means to reach other aims (typically for commercial promotion).

Researchers must lastly understand the *language* of the documents to be analysed. They may happen to come across documents in a foreign language, but this is not the only hurdle. A shared language may also take on varying nuances in the use which particular segments of the population make of it. For example, the blog of an adolescent will make use of a specific language which the researcher has to decode to fully understand the author's message (on this subject we may also think of what has just been said above on the subject of virtual communities, which might develop their "own" code of expression). The variety of formats Internet documents may take (above all iconic language) must also prompt the researcher to be ready to interpret specific, particular codes of expression.

5.3 *Analysis*. Since web-mediated personal documents have very varied formats (letters, diaries, biographical accounts, photos or videos, etc.) and use different languages in their expression (written texts, images, audio tracks, etc.), in the analysis phase the researcher may make use of very different analysis techniques, to be chosen accordingly with specific attention (we may also recall that the use of documentary material may be part of both a qualitative research approach and of a quantitative analysis).

In order to analyse the significance of a document, various strategies for analysis are therefore available, ranging, to make but a few examples, from discourse analysis to narrative analysis, from the analysis of metaphors to that of conversation. The most widespread strategies, which may be applied both to texts and to images, include content analysis (qualitative and quantitative) and semiotic analysis, which we have already mentioned here in the paragraph on documentary analysis). Internet documents have the great advantage of being already "transcribed" or in any case registered in digital form. This saves time and money in the preliminary phase of the analysis.

At the same time, on-line documents need some stratagems in the analysis phase to take into account the particular nature of the communication channel they are using, i.e. Internet sites. In particular, for those undertaking a semiotic analysis, some suggestions may include distinguishing the space of the paratext¹⁹ from that of contents; to make constant reference to the users' shared language; and to distinguish the different roles called for in the typical situation of the site. Lastly, analysis must be made of the visual spaces, enunciative strategies and the general coherence of the site²⁰.

¹⁹ The paratext is made up of the set of auxiliary elements for understanding a text: titles, subtitles, notes, captions, key words, tables and images. The paratext also enriches and simplifies the text, in that, on the one hand, it offers the possibility to carry out in-depth examinations and, on the other, points out more rapid reading paths. Paratext elements play a key role in Internet.

²⁰ For an introduction, see <http://euphrates.wpunj.edu/faculty/yildizm/SP/> (accessed 24th april 2010).

6. Short conclusive notes

In this article we presented some of the main methodological issues that arise in doing research using the Internet (we made a distinction between the research *with* Internet and the research *on* Internet: this work is mainly devoted to the research on Internet).

The issues that have been presented raise complex and sensitive questions. This article contains no solutions to these questions, but rather a list, a sort of reminder, of the open issues. Nor is it possible to provide answers to many of the questions submitted: due to their complexity and relative newness, the scientific community has not yet given clear and shared answers to these questions.

A clear example in this direction consists of questions concerning the choice of the empirical documentation (we discussed this aspect in previous sections, providing some references to empirical studies that have adopted one of the available sampling strategies).

So what is the contribution of this work? First, this article aims to place properly the research using the Internet (research *with* Internet and research *on* Internet) within the tradition of sociological research, of which it is a full part. We believe that research using Internet can provide valuable research strategies: these research opportunities have equal dignity as the traditional ones.

As a second point, this article intends to emphasize that scholars doing research on Internet and with Internet cannot be considered “exempt” from the discipline of sociology and of social research methodology. Before beginning the research, all issues we listed in this paper must be considered and well planned. First, researchers should try to make use of what is already part of the toolbox of the sociologists. Researchers should be also aware that specific solutions and special precautions will be often necessary. We argue that it is important that all decisions are always declared, explained and discussed in the phase of presentation of the results. In research papers, a large methodological section must always be provided to allow the reader to reconstruct all the steps taken. It is important that the researcher will always give a comparison with the choices made in other similar studies, outlining his/her difficulties, acknowledging the pros and cons of his/her choices. Based on this report, confidence in the results proposed will be built, and a repertoire of good practice will be gradually built.

We hope that this contribution will be followed by other works: we refer not only to empirical researches, that in recent years seem to flourish, but even before to methodological reflections that in this moment are absolutely necessary.

REFERENCES

- Alaszewski, A. (2005). *Using Diaries for Social Research*, London, Sage Publication.
Altheide, D.L. (1996) *Qualitative Media Analysis*, Thousand Oaks, Sage.
Arosio, L. (2010) L'uso dei documenti. In: de Lillo, A. (ed.) *Il mondo della ricerca qualitativa*, Torino, Utet, in print.

- Arosio, L. & Ruspini, E. (2010) Unico, indimenticabile, da sogno. Il viaggio di nozze tra mito ed esigenze di mercato. *Sociologia Urbana e Rurale*, in print.
- Barthes, R. (1964) *Eléments de sémiologie*, Paris, Seuil.
- Barthes, R. (1982) *L'obvie et l'obtus*, Paris, Seuil.
- Berelson, B. (1952) *Content Analysis in Communication Research*, Glencoe, Free Press.
- Bichi, R. (2004) *I documenti personali*. In: Corposanto, C. (eds.) *Metodologia e tecniche non intrusive nella ricerca sociale*, Milan, FrancoAngeli.
- Chalfen, R. (1987) *Snapshot Versions of Life*, Bowling Green, OH.: Popular Press.
- Cohen L., Manion L. & Morrison K. (eds.) (2000) *Research Methods in Education*, London, Routledge.
- de Lillo, A. (ed.) (2010) *Il mondo della ricerca qualitativa*, Torino, Utet, in print.
- Dillman, D. A. & Bowker, D. K. (2001) The Web questionnaire challenge to survey methodologists. In: Reips U. D., & Bosnjak M. (Eds.), *Dimensions of Internet science* (pp. 159-178). Lengerich: Pabst Science Publishers.
- Dillman, D. A., Tortora, R. D., & Bowker, D. (1999) Principles for constructing web surveys Pullman, Washington: Washington State University; SESRC Technical Report 98-50.
- Girard, A. (1963) *Le journal intime*, Paris, Puf.
- Glaser, B. G & Strauss, A. L. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago, Aldine Publishing Company.
- Greenbaum, T. (2008) *Internet focus groups are not focus groups - so don't call them that*. Available from: <http://www.groupsplus.com/pages/qmro798.htm> [Accessed 24th April 2010]
- Greimas, A. J. (1970) *Du sens*, Paris, Seuil.
- Greimas, A. J. (1983) *Du sens II - Essais sémiotiques*, Paris, Seuil.
- Henzinger, M. R., Heydon, A., Mitzenmacher, M., & Najork, M. (2000) *On Near-Uniform URL Sampling*, Proceedings of the 9th international World Wide Web conference on Computer networks: the international journal of computer and telecommunications networking. Available from: <http://www9.org/w9cdrom/88/88.html> [Accessed 24th April 2010]
- Hine, C. (2008) *The Internet and Research Methods*. In: Gilbert, N. (eds.) *Researching Social Life*, London, Sage.
- Jung, T., Youn, H. & Chang, T. (2005) Understanding Diet Websites Function on the Web: Content Analysis of Diet Websites, Paper presented at the annual meeting of the International Communication Association, Sheraton New York, New York City, NY. Available from: http://www.allacademic.com//meta/p_mla_apa_research_citation/0/1/3/2/3/pages13236/p13236-1.php [Accessed 24th April 2010]
- Kracauer, S. (1952) *The Challenge of Qualitative Content Analysis*. *Public Opinion Quarterly*, 16, pp. 631-642.
- Krippendorff, K. (1980) *Content Analysis: An Introduction to its Methodology*, London, Sage.
- Lee, R. M. (2000) *Unobtrusive methods in Social research*, Open University Press.
- Lejeune, P. (1989) "Cher cahier..." *Temoignages sur le journal personnel*, Gallimard.
- Lejeune, P. (2000) "Cher écran..." *Journal personnel, ordinateur, Internet*, Editions du Seuil.
- Lejeune P. (2009) *On Diary* (edited by J. D. Popkin & J. Rak), University of Hawaii Press.
- Macdonald, K. & Tipton, C. (1993) *Using documents*. In: Gilbert, N. (eds.) *Researching social life*, Sage, London.
- McCulloch, G. (2004) *Documentary Research in Education, History and the Social Sciences*, London and New York, Routledge.

- Mayring, P. (1988) *Qualitative Inhaltsanalyse: Grundlagen und Techniken* [Qualitative content analysis. Basic knowledge and techniques]. Weinheim, Germany: Deutscher Studien Verlag.
- Norris, P. (2001) *Digital Divide. Civic engagement, information poverty, and the Internet Worldwide*, Cambridge, University Press.
- Paldacci, M. (2003). Les quatre mondes du journal intime en ligne. In: *Terrains & Travaux* , 2 (n 5), pp. 7-30.
- Plummer, K. (1983) *Documents of Life*, London, Allen & Unwin.
- Prior, L. (2003) *Using Documents in Social Research*, London, Sage.
- Rheingold, H. (1993) *The Virtual Community: Finding Connection in a Computerized World*, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.
- Ruspini, E. (2002) *Introduction to Longitudinal Research*, London, Routledge.
- Saussure de, F. (1916) *Course in General Linguistics*.
- Scott, J. (1990) *A Matter of Record: Documentary Sources in Social Research*, Cambridge, Polity Press.
- Scott, J. (2006) *Documentary Research*, London, Sage Publications.
- Sorapure M. (2008) *The Lifewriting of Dataselves: Autobiographical Acts in New Media*. In: *Reading (and Writing) New Media*, C. Ball & J. Kalmbach (Eds.) Hampton Press.
- Thomas, W. I., & Znaniecki, F. (1918-20) *The Polish Peasant in Europe and America*, Chicago, University of Chicago Press.
- Webb, E. J., Campbell D. T., Schwartz R. D. & Sechrest L. (1966) *Unobtrusive Measures: Nonreactive Research in the Social Sciences*, Chicago, Rand McNally.
- Weber, P.R. (1990) *Basic Content Analysis*, London, Sage Publications.

Laura Arosio is Assistant Professor in Sociology in the Faculty of Sociology of the University of Milan Bicocca. She has a Ph.D. in Sociology (University of Milan). Her main interests concern families, marriage markets, conjugal instability in contemporary societies. She is very interested in the field of methodology and empirical research (both qualitative and quantitative). Lately she has intensively studied the potential of documentary analysis for social sciences. She also is interested in longitudinal and life course analysis.