

NORMAN BACKHAUS &  
ULRIKE MÜLLER-BÖKER

# Gesellschaft und Raum

Konzepte und Kategorien



## Gesellschaft und Raum

## Konzepte und Kategorien

Humangeographie

Herausgeberin:  
Ulrike Müller-Böker

**Norman Backhaus & Ulrike Müller-Böker**

**Gesellschaft und Raum**  
Konzepte und Kategorien

1. Auflage

Band 22  
Zürich 2006

Geographisches Institut  
Abteilung Humangeographie  
Universität Zürich  
Winterthurerstrasse 190  
CH-8057 Zürich  
Schweiz

**Gesellschaft und Raum: Konzepte und Kategorien /**  
Norman Backhaus & Ulrike Müller-Böker. – 1. Aufl. – Zürich, 2006  
(Schriftenreihe Humangeographie 22)

ISBN 3-906302-05-9

Erste Auflage 2006

© Norman Backhaus & Ulrike Müller-Böker  
Alle Rechte der Verbreitung vorbehalten

Druck:  
Druckerei der Zentralstelle der Studentenschaft der Universität Zürich

Umschlaggestaltung:  
Martin Steinmann

Layout:  
Norman Backhaus

Titelfoto:  
Ulrike Müller-Böker

Printed in Switzerland

ISBN 3-906302-05-9

# Inhalt

<b>Vorwort: Konzepte und Kategorien in der humangeographischen Forschung</b>	<b>9</b>
<i>Norman Backhaus &amp; Ulrike Müller-Böker</i>	
<b>Regionalisierung: eine konstruktivistische Perspektive</b>	<b>13</b>
<i>Norman Backhaus &amp; Urs Müller</i>	
<b>Regionalisierung und die Methode der Bildanalyse</b>	<b>31</b>
<i>Urs Müller &amp; Norman Backhaus</i>	
<b>Regionalisierung: Fallbeispiel Biosphäre Entlebuch</b>	<b>53</b>
<i>Urs Müller</i>	
<b>Stakeholder als sozialwissenschaftliches Konzept: Begrifflichkeit und Operationalisierung</b>	<b>73</b>
<i>Olivier Ejderyan, Urs Geiser &amp; Marc Zaugg Stern</i>	
<b>The household as an analytical category: concepts and challenges</b>	<b>103</b>
<i>Heidi Kaspar &amp; Michael Kollmair</i>	
<b>Data protection in qualitative research</b>	<b>125</b>
<i>Heidi Kaspar &amp; Ulrike Müller-Böker</i>	

## Autorinnen und Autoren

**Norman Backhaus** ist habilitierter wissenschaftlicher Mitarbeiter und Lehrbeauftragter der Abteilung Humangeographie des Geographischen Institut der Universität Zürich (GIUZ). Seine thematischen Schwerpunkte in Forschung und Lehre sind Sozialgeographie, Globalisierung, Naturschutz und Ressourcennutzung sowie nachhaltige Entwicklung und Tourismus. Seine regionalen Schwerpunkte liegen in Südostasien und in der Schweiz.

**Olivier Ejderyan** ist Doktorand am GIUZ. Er schreibt seine Dissertation über Mitwirkungsprozesse im Rahmen von Gewässer-Renaturierungsprojekten in der Schweiz. Ein spezieller Fokus liegt auf den Konsequenzen der kollektiven Wissensproduktion. Zurzeit ist er mit der Ausarbeitung des eLearning Projekts 'Globalisation and Livelihood Options of People living in Poverty' (GLOPP) beschäftigt.

**Urs Geiser** ist promovierter wissenschaftlicher Mitarbeiter am GIUZ. Er forscht und lehrt zu Lebensunterhaltsstrategien und Ressourcennutzung in ländlichen Gebieten Südasiens. Ländliche Entwicklung als Handlungsfeld unterschiedlicher Akteure sowie Diskurs und Praxis von 'Partizipation' sind weitere Forschungsfelder.

**Heidi Kaspar** ist diplomierte Geographin und arbeitet gegenwärtig am GIUZ im Rahmen des vom Schweizerischen Nationalfonds finanzierten Forschungsprojekts 'Nachhaltiges Entwerfen, Planen, Bewirtschaften und Aneignen öffentlicher städtischer Parkanlagen' (NFP 54) an ihrer Dissertation zur Thematik Raumaneignung und Interaktion im öffentlichen Raum.

**Michael Kollmair** ist promovierter Geograph und hat sich während seiner Forschungs- und Lehrtätigkeit an der Universität Zürich mit Nepal und Ressourcenmanagement auseinandergesetzt. Heute arbeitet er als Programm-Manager für 'Culture, Equity, Gender and Governance' am International Centre for Integrated Mountain Development (ICIMOD) mit Sitz in Kathmandu.

**Urs Müller** ist promovierter Geograph am GIUZ. In seiner Dissertation hat er sich – im Rahmen des Nationalen Forschungsprogrammes 48 'Landschaften und Lebensräume der Alpen' – mit der Rolle von Bildern im Prozess der nachhaltigen Entwicklung auseinandergesetzt. In der Lehre befasst er sich mit Sozialgeographie und speziell mit Regionalisierungen. Zur Zeit ist er Post-Doc-Stipendiat des Universitären Forschungsschwerpunktes 'Asien-Europa' der Universität Zürich.

**Ulrike Müller-Böker** ist zur Zeit Direktorin des GIUZ und hat den Lehrstuhl für Humangeographie inne. Sie leitet das Workpackage 'Livelihood options and globalisation' im National Centre of Competence in Research North-South (NCCR North-South) Ihre Forschungsschwerpunkte liegen in der geographischen Entwicklungsforschung, speziell in den Feldern 'Lebensunterhaltsstrategien', 'indigenes Wissen', 'Naturschutz' und 'Migration'. Ihre regionalen Schwerpunkte sind Süd- und Zentralasien.

**Marc Zaugg Stern** ist promovierter Oberassistent der Abteilung Humangeographie des GIUZ. Seine Schwerpunkte in Forschung und Lehre sind Definition und Umsetzung von raumwirksamen öffentlichen Politiken, gesellschaftlicher Umgang mit Risiken und der Einsatz von raumwirksamer Technologie sowie die Anwendung qualitativer und quantitativer Methoden.



© Heidi Kaspar

# **Data protection in qualitative research**

**Heidi Kaspar & Ulrike Müller-Böker**

## Zusammenfassung und Abstract

### Zusammenfassung

Die Diskussion um Datenschutz findet hauptsächlich statt, wenn grössere Datenmengen gesammelt und archiviert werden (siehe Corti et al., 2000; ESDS, 2004a; ESDS, 2004b; BstatG, 1992). In einer globalisierten und zunehmend vernetzten Wissenschaftsgemeinschaft (Parry and Mauthner, 2004: 140; Bishop, 2005: 335) gewinnt Datenschutz aber immer mehr an Bedeutung.

Während das Archivieren von quantitativen Daten in der Regel als relativ unproblematisch empfunden wird (siehe z.B. BstatG, 1992), fand in der qualitativen Forschung eine kontroverse Diskussion um die Datenarchivierung statt (Parry and Mauthner, 2004: 140). Der Grund dafür liegt darin, dass die Generierung von qualitativen Daten als ein gemeinsames Unterfangen von Interviewten und Interviewenden verstanden wird, weswegen alle beteiligten Parteien Urheber- und Eigentümerrechte haben sollten (Parry and Mauthner, 2004: 141). Fragen des Datenschutzes müssen jedoch nicht nur dort gestellt werden, wo Daten archiviert werden, sondern sie sind für jede Sozialforschung, die Daten erhebt und/oder auswertet von Bedeutung.

Dieser Beitrag weist auf zentrale Aspekte des Datenschutzes in der Sozialforschung hin. Vor allem möchten wir betonen, dass Datenschutz ein integraler Bestandteil des wissenschaftlichen Forschungsprozesses ist und also auch von Beginn an in das Forschungsdesign integriert werden soll. Neben allgemeinen und theoretischen Überlegungen liefert der Beitrag auch praktische Hinweise zur Anonymisierung von Datensätzen – dem wichtigsten Bestandteil des Datenschutzes. Die Ausführungen werden durch Beispiele v.a. aus der Forschungspraxis der Abteilung Humangeographie des Geographischen Instituts der Universität Zürich veranschaulicht.

### Abstract

The debate on data protection has so far been confined to institutions that collect and archive data in great quantities (see Corti et al., 2000; ESDS, 2004a; ESDS, 2004b; BstatG, 1992). In a globalised and increasingly networked (referring to geography, disciplines and institutions) scientific community (Parry and Mauthner, 2004: 140; Bishop, 2005: 335), data protection becomes more and more important. Previously, data from research was considered the property of the researchers. Nowadays, data sets are increasingly shared within the scientific community (Parry and Mauthner, 2004: 140). This development has two implications: the question of the ownership of scientific data has to be addressed afresh and data protection becomes ever more important.

Whereas, in quantitative research, archiving data is usually viewed as rather unproblematic (see e.g. BstatG, 1992), data protection has aroused controversy and debate in the qualitative research community (Parry and Mauthner, 2004: 140). The reason for this debate lies in the very nature of qualitative research. As Parry and Mauthner (2004: 141) state, “the construction of qualitative data is a joint endeavour between respondent and researcher” and therefore “both parties should retain authorship/ownership rights over the data” (Parry and Mauthner, 2004: 141). However, data protection cannot be confined to the archiving of data but has to be considered in every social research project that collects and/or analyses data.

This article covers some crucial issues of data protection for social research. Its particular aim is to emphasise the importance of data protection as a necessity to preserve scientific standards. We argue that data protection is an integral part of social research and therefore has to be included in the design of research projects. Besides general and theoretical considerations, the article gives practical advice for rendering data sets anonymous, which is the main procedure in data protection. The comments are accompanied by examples drawn mainly from the research practice of the Division of Human Geography of the University of Zurich's Department of Geography .

## Introduction

Various forms of interviews, ranging from structured questionnaires to non-structured narratives, are core methods of human geography. There are various challenges connected to these data collection methods. One of these challenges is data protection – an issue that is, in our opinion, insufficiently acknowledged and discussed in human geography, as is equally the case in other research disciplines.

Behind the results of social science there are human beings with a fundamental right to privacy. Therefore, good research should not only be oriented towards reliability, validity and intersubjectivity as academic standards, but must also answer to ethical standards such as the integrity of interviewed persons. Despite the paramount importance of this issue, little has been published on it (Corti et al., 2000). This is surprising since data protection is a part of methodology which concerns every social scientist. However, by contrast with data collection and data processing, data protection is hardly ever discussed in research.

The issue of data protection plays a special role in development research. Whereas it seems obvious that data has to be anonymised in studies within the researcher's own society in order to protect the informants' privacy, the geographical distance involved in development studies often relegates this issue to a backseat. However, from an

ethical point of view, there is no reason, why ethical guidelines which are valid for the “home society” should not also be valid for other societies. It is impossible to foresee what will happen to the data, who will use it or which research projects might be conducted in the future. For example, emerging North-South research cooperation might allow people access to data that was not meant to be seen by people who are possibly very familiar with the research sites.

It is commonly acknowledged that it is part of the researcher’s responsibility to ensure that her or his informants are not exposed to potential injury or put at risk of harm, scandal or ridicule. But how is this need to be satisfied? This paper cannot provide a universal recipe of how to safeguard informants’ privacy; this is not possible as every research situation is unique and therefore requires special treatment. Instead, this paper aims to discuss the key points concerning data protection. Its objective is to make social scientists aware of this fundamental issue and to encourage academics to integrate privacy protection into their research designs from the outset and throughout the whole process of their projects.

The article first presents general considerations on data protection and then discusses details about obtaining informed consent and data anonymisation – the main elements of data protection. In the final chapter, there is a summary of the comments and an outlook for the future.

## Elements of data protection

Anonymisation of data is generally considered the most important component of data protection. However, anonymisation is only one element of data protection. There is a wide range of possible strategies to maintain informants’ confidentiality (Bishop, 2005: 334). Respecting ownership and copyright, warranting confidentiality of respondents and securing informed consent are other important elements of data protection in academic research (Parry and Mauthner, 2004: 141). Later in this paper, we will see that individual components of data protection can also be contradictory.

The UK Data Protection Act 1998 lists eight principles to make sure that information are handled properly. These basic principles state that data must be:

1. fairly and lawfully processed,
2. processed for limited purposes,
3. adequate, relevant and not excessive,
4. accurate,
5. not kept for longer than is necessary,
6. processed in line with one’s rights,

7. secure,
8. not transferred to countries without adequate protection.

In this paper, two elements are discussed in detail: obtaining informed consent and data anonymisation. These are the two components that are considered to be the most important and, at the same time, the most complex ones. Confidentiality is another important element. In several countries, such as the UK, there is a law on confidentiality (ESDS, 2004a). But, independently of whether such a legal background exists or not, researchers should feel obligated from an ethical point of view to handle information in a confidential manner. ESDS (2004a) emphasises that, when an explicit statement of agreement has been made regarding the extent of confidentiality provided by the researcher towards the informant, it constitutes a contract and the abuse of this confidentiality is equivalent to a breach of contract. Such an agreement is valid even when it is only verbal. Furthermore, under certain circumstances, it does not even require an explicit agreement to commit the researcher to handle the obtained information confidentially (ESDS, 2004a). ESDS (2004a) further states that the duty of confidentiality is also binding for situations where information is passed confidentially and where sensitive data is disclosed. But which data is sensitive? The Swiss law on data protection (Bundesgesetz über den Datenschutz (DSG) of 19th July 1992, DSG, 1992, Art. 3.c) lists the following data as particularly sensitive:

- data referring to religious, political, union activities and opinions, as well as those concerning a person's world view,
- data referring to medical details, a person's private sphere or ethnicity,
- data referring to social welfare provisions,
- data referring to administrative or penal sanctions.

It is evident that the lists above are not exhaustive. Before beginning an interview, interview partners must have given their agreement to participate. This issue will be tackled in the following section.

## Seeking consent

It is an ethical standard that informants have the possibility to give their consent freely to participate or not. Consent is defined by the EU Directive 95/46/EC/24/10/1995 as "any freely given specific and informed indication of his [sic] wishes by which the data subject signifies his agreement to personal data relating to him being processed" (ESDS, 2004a). According to this definition, there must be active communication between the parties. Hence consent cannot be inferred from a lack of response to a communication (ESDS, 2004a). Consistent with these insights, Corti et al. (2000: chap. 2.1) demand that interviewees must be aware of their right to refuse participation and be reminded of their right to renegotiate consent during the research process.

ture. Sin (2005: 286) alerts researchers to the fact that providing consent does not include a guarantee for completing the interview (see also Miller and Bell, 2002: 65). Rock (2001: 7) further stresses that even when consent is obtained, researchers should remain aware of the scope of the consent and of the expected use of the data (see also ESDS, 2004b).

Furthermore, consent must be based on adequate information about the study the participants are going to be part of (Parry and Mauthner, 2004: 146). Hence, interviewees have to be informed as completely as possible and in terms comprehensible to them what the research is about, who is undertaking and financing it and why and how it will be disseminated (Parry and Mauthner, 2004: 146). But the extent to which participants can ever be fully informed is hotly disputed (Sin, 2005: 281). Corti et al. (2000: chap. 2.1) emphasise that “we should never assume that all participants have a detailed appreciation of the nature and aims of academic research”. On the other hand, it is problematic to make a judgment on who is able to give informed consent (Sin, 2005: 280). Furthermore, consent alone does not absolve researchers from the responsibility to anticipate and guard against potentially harmful consequences for participants (Parry and Mauthner, 2004). Finally, the traditionally reflexive and iterative process of qualitative research renders a “once only” consent insufficient, as new research questions arise during the research process (Parry and Mauthner, 2004: 146; Sin, 2005: 281). Miller and Bell (2002: 54-55) emphasise that the final research findings may have little in common with the initial aims of the project presented to the interviewees as a basis for them to decide whether or not to participate. For the authors, this situation raises the question of what the participant is consenting to. Is consent limited to the participation in research in terms of being interviewed? Or does participation include reading and commenting on transcripts and the analysis of the data? For this reason, a “process consent” is demanded and is indeed more appropriate. But the requirements of such a “process consent” might be impossible to meet during secondary use of a data set because informants can no longer be contacted since their identity has been concealed for data protection purposes (Parry and Mauthner, 2004: 146). This shows that one component of data protection can hinder the requirements of another element from being fulfilled; this is the complexity of data protection.

On the other hand, a high level of consent makes it possible to present sensitive information according to a high ethical standard. In their study on the norms and practice of the cantonal flood protection authorities, Zaugg et al. (2004) involved informants in the transcription and data preparation procedure. Telephone interviews were conducted with civil servants from the cantonal offices. Notes were made of these interviews. The minutes were then returned to the interviewees with the request that

they check it for accuracy – so as to avoid misunderstandings – and completeness – to allow for specifications and complement. In his study of institutional changes in Swiss flood protection since the 1970s, Zaugg (2006) even returned individual paragraphs, in which people were directly or indirectly quoted, to the respective informants. The aim of this procedure was to ensure that the researcher could not be blamed for having included false information. By giving explicit confirmation (i.e. by signing), participants stated that they had seen the respective paragraphs and gave their approval for their statement to be reproduced as fully representing their position in a clear and appropriate way. Moreover, this manner of seeking consent allows the scientist to verify and review his or her interpretations. Hence, when participants disagree with the scientist's interpretations, it gives the latter the chance to revise his or her views on the topic. But such situations can also become very complex and the researcher can find herself or himself entangled in contradictions where she or he has to decide what or who is "right". These procedures of seeking consent are obviously very time-consuming and complex and must be factored in from the very beginning of research projects. Despite this, they are indispensable for publications presenting statements or data that can be traced back to an individual by insiders or others. Furthermore, involving informants in analysis encourages their more active participation and allows for more reflected interpretations and hence results in scientific surplus.

Sin (2005) demonstrates through the example of a survey of elderly people that seeking consent is a highly dynamic process. The participants in a national survey agreed to take part in a follow-up study. When these participants were contacted to arrange a date for an interview, a considerable number refused to participate again. Sin therefore concludes that "seeking and giving consent is an extremely complex process" and proper procedure does not ensure final consent (*ibid.*: 285). The example further demonstrates that consent "cannot be taken for granted at any stage of the research process" (*ibid.*: 285-286). Sin (2005: 281-282, 286-287) lists four factors that constitute the ephemeral and ever-changing character of consent:

1. Depending on how definitions of consent are put into practice and on who is dealing with them, the expectations of what constitutes adequate consent may vary.
2. An interview usually consists of topics of varying degrees of sensitivity. Yet the sensitivity of issues is relative to and dependent on the perception and experience of the interviewees.
3. The different phases of a research project call for different forms of participation from the interviewees' side.
4. The extent to which researchers are committed to ethical issues varies from person to person. This might pose problems to team research.

The practical side of obtaining consent is often understood as ensuring that there is a signature at the bottom of a form where the purposes and requirements of the interview are listed. While this method renders the consent traceable by others, it should also be considered that such a procedure adds a very formal character to the interview. This formality can act as a deterrent and embarrass people. Therefore, the ESDS (2004a, section “Consent”) notes that consent should be obtained “preferably in writing, but clear and unambiguous verbal consent will suffice”. For further practical guidelines on securing consent, see ESDS (2004a, section “Consent”).

After the interview, considerations on data anonymisation have to be addressed – dependent on the issues discussed in interviews, the personal wishes of the interview partners and the intended use and dissemination of the data.

## Data anonymisation

Data anonymisation is generally understood as “the deliberate changing of, or concealment of, the name (and hence, identity) of someone or something” (Rock, 2001: 1). Names and all other details that make the identification of people possible are called identifiers. ESDS (2004b) differentiates between direct and indirect identifiers. Direct identifiers include names, addresses including postcode information, telephone numbers etc. Indirect identifiers include “information that when linked with other publicly available sources, could result in a breach of confidentiality”, such as geographical information, workplace/organisation, education, institution or occupation.

Hauser et al. (1998: 4) differentiate between absolutely and factually anonymised data: For absolutely anonymised data sets, “it must not be possible for anybody even with a vast additional knowledge about one or all units in the sample to find out the identity of any unit”. Of course, this is an extremely strong restriction and may result in unacceptable distortion of the data. Therefore, factual anonymisation is in many cases more appropriate. Factual anonymisation does not completely conceal identity. A person’s identity can be revealed if the reader has additional knowledge. Factual anonymisation can only be used when the researcher can assume that users of the data have limited knowledge, thereby ensuring that they will not be able to identify informants. Additionally, potential users should be limited to well-defined groups of scientists that can be held responsible if the ethical code of data protection is violated.

An example of factually anonymised data is the thesis of Gamper (Gamper, 2004: 12). In his study of institutions of forest use in Nepal, he shows how different groups use different forest patches for different purposes. In Nepal, forests provide house-

holds with basic resources such as fodder, fuel wood, construction materials for agricultural appliances and medicinal plants, as well as serving as hunting and grazing land. The study by Gamper (2004) reveals that many people use forests illegally. Due to this illegal activity, respondents must not be identifiable. As the data was collected in small villages, only removing the names would not have been sufficient to conceal identity. Hence, the names of the villages were replaced by pseudonyms and only the district name was revealed. But the article contains a location map. Though the map does not contain location names, it does mean that the study area can be identified by people familiar with the setting. The map was considered indispensable as it illustrates the patterns of forest use by the different groups. Thus, the article cannot provide absolute anonymity, as this would have meant an unacceptable loss of information – but nevertheless informants are protected from direct consequences resulting from the publication of their illegal activity.

Steimann's (2004) study is another example from forest management research. Steimann investigated the relationships between the employees of the state forest department and the local population in the North-West Frontier Province (NWFP) of Pakistan. For this purpose, Steimann interviewed forest officers of lower levels in the hierarchy about their working conditions. In doing so, sensitive issues such as their relationships to colleagues and supervisors, salaries, bribery and other illegal practices relating to forestry came up. In this setting, guaranteeing confidentiality was a prerequisite for obtaining any information at all; it also had implications for the interview situation as well as for the presentation of the findings. Providing anonymity for the informants included Steimann having to work with an interpreter with absolutely no connections to the forest officers and making sure that no third party was around. For the presentation of the data, anonymisation included concealing details about the place and date of the interviews, since this information would have easily betrayed the informants' identity. The statements of local informants without any official function were also anonymised, but the location of the interview could be named, as many interviews were conducted in each village and hence identification of individuals was impossible. This example shows that the appropriate level of anonymisation depends to a great extent on the context and the issues to be discussed.

By contrast, in the study of Landolt (2004), location names were not concealed, as the research objective was a comparison of different municipalities of selected "socio-geographical milieus" (German: "sozialgeografische Milieus") in the canton of Zurich regarding how foreigners were perceived. The names of the places under study could have been replaced by pseudonyms, but this technique of data protection was unnecessary as other methods of anonymisation were applied. Firstly, one key criteria for the selection of the municipalities to be investigated through semi-structured group

discussions was the number of inhabitants: it must exceed 1,000. This population size prevents identification by people who have not participated in the group discussions themselves, even if they live in the community. But in a municipality with 1,000 inhabitants, people still know each other quite well and participants' identity could be betrayed by characteristic phrases, statements or opinions. Therefore, the results were presented as depersonalised statements of the whole group in order to make it impossible for non-participants to identify the informants. Thus place-names were maintained – allowing follow up studies, for example – without curtailing informants' privacy.

As a general principle, all data that contains sensitive personal information that may harm the respondent him- or herself or a third party – such as the illegal activity of informants mentioned earlier – have to be disconnected from data that could make it possible to identify an individual. In addition, data that opens someone to ridicule or scandal – as well as any the informant wishes to keep confidential – must be anonymised. Parry and Mauthner (2004: 148) state that the problem with this requirement is that often it is the researcher who differentiates between sensitive and non-sensitive data according to her or his consideration. In fact, respondents might place a very different emphasis on what they consider sensitive, because “these are personal, not just scientific and professional matters” (Rock, 2001: 4-5). It can be assumed that the gap between researchers' and informants' considerations widens the greater the cultural distance between them. Hence, this issue should be of special concern in development studies, since “we cannot know in advance the belief systems of the communities we are studying” (Rock, 2001: 5-6).

Grinyer (2002) supports the above statement by stressing that researchers should be careful about making judgements on behalf of others, however well-intentioned they may be. For her, it cannot be assumed, even for sensitive issues, that everybody wishes her or his information to be anonymised. This challenge to orthodox ethical (and legal) principles is a conclusion of her research with parents of young people suffering from cancer. The majority of the informants explicitly wished to have their own names published. Interviewees felt as if they were losing ownership of their stories and wished to be acknowledged by name. Thus Grinyer (2002) concludes that the issue of “identity/anonymity is more complex and less predictable than it might at first appear and may problematise some of the ethical and legal requirements for good practice”. The above example provides evidence of the complexity of data anonymisation. Researchers not only have to find a balance between providing privacy and preventing the loss of important information, but also between providing privacy and preventing a “loss of ownership” (Grinyer, 2002). To find an appropriate ba-

lance is the task of every researcher and has to be addressed one-on-one with each respondent (Grinyer, 2002).

Müller (2001) was confronted with another challenge to standard ethical principles. In his study on participative practices in nature conservation projects in Switzerland, he interviewed several stakeholders. In order to facilitate open and free discussions, interviews were not taped and they were assured of confidentiality. Furthermore, the issue of the extension of the conservation area was very present and highly controversial at the time of field-work. For these reasons, Müller concealed his informants' names. However, an important stakeholder criticised data anonymisation. According to this person, data anonymisation was unscientific for an issue of public importance, since informants could not be held responsible for what they said – the very aim of anonymisation being indeed to protect people from possible (negative) consequences resulting from their statements. Though in this particular case the criticism cannot be regarded as scientifically relevant, the example shows that data protection can also lead to animosity and disagreement (see Müller & Kollmair 2004).

Parry and Mauthner (2004: 145) introduce an issue that is rarely discussed: *the researcher's anonymity*. As qualitative researchers view themselves as part of the research process, it is common for them to reveal personal information related to the research topic. However, researchers' identities are impossible to conceal and therefore their anticipation of the data being archived and hence open to outside scrutiny can influence their behaviour in the field, even to the point of compromising quality of the interviews.

Critical voices are raised about data anonymisation in general. Corti et al. (2000: chap. 4) have reservations about the feasibility of data protection. They ask:

*“Is it really possible to completely disguise a workplace or a village or the central characters in the drama? The situation is even more complex when we begin to think about other researchers re-using data they have not collected themselves. How can we ensure that respondents are suitably protected? Can we trust re-users to act responsibly?” Anonymisation can, in fact, be impossible in some cases. For instance, when the “number of the respective observation units is so small that it is possible to identify them already with very little additional information” (Hauser et al., 1998: 5).*

Even if names are removed or replaced, the identity of a person may be betrayed by a report, which describes in detail an individual's circumstances that may allow a third party to identify that person (Lancaster University, 2002). This is the case in Böhnisch's (1999) study of the social practices, roles and relationships of 20 spouses of top managers. The scientist faced difficulties in arranging interviews due to the informants' mistrust of researchers and their reluctance to be the object of public de-

bates. The data consisted amongst other things of information on family origins, their living circumstances and the milestones in their husbands' careers. This information is so specific it would have betrayed their identity; to tackle this and meet the informants' need for privacy, Böhnisch decided to anonymise his informants' names as well as information about their curriculum vitae and activities. This does not mean that the data is imaginary, however, as the structure of each specific life story was kept (Böhnisch, 1999: 245).

However, Parry and Mauthner (2004: 144) warn against resorting to falsifying non-essential information in order to protect participants in cases like the one mentioned above, where particularly sensitive research data is indispensable for analysis or where respondents are especially vulnerable. It should also be kept in mind that modifying the context renders secondary use of the data problematic; the difficulty lies in differentiating between essential and non-essential data and in who – the researcher or the respondent – should undertake this differentiation (Parry and Mauthner, 2004: 144).

The considerations discussed above show that data anonymisation is a process of balancing between different requirements. While informants' privacy must be assured, as much information as possible should be left in the data set. In fact, not all data needs to be anonymised and it is worth thinking, for all the reasons given above, about which data must be anonymised and which can be left unmodified.

After all these general, theoretical and critical considerations, practical advice is due. This is the focus of the following section.

## **Anonymisation in practice**

There are various existing guidelines for data anonymisation. Corti et al. (2000) present a list of guidelines of different research disciplines (such as anthropology and psychology). Parry and Mauthner (2004) especially commend the guidelines and expertise of oral history (British Social Association, 2002). Such guidelines can be used as orientation when designing data anonymisation procedures and the specification of anonymisation needs and requirements. Nevertheless, the usefulness of such guidelines should not be overestimated, as "one size can never fit all" (Corti and Wright, 2002, cit in. Bishop, 2005: 334). Therefore, guidelines often have to be considered "suggestive and general, not prescriptive and particular" (Bishop, 2005: 334). Furthermore, guidelines are not sufficient on their own, since they do not contain any practical hints on how one should actually conduct data protection procedures. Some practical hints are presented below.

In general, there are three questions for a good data protection design that have to be answered by scientists before embarking on a research project:

- At what point in the research process should data anonymisation be conducted?
- Which data has to be changed or removed?
- How is the data to be modified?

### **When within the research process should data anonymisation be conducted?**

Generally, data protection considerations should be an issue at every stage of the research process. Rock (2001: 16) states that anonymisation of data may be conducted

- a) from the very beginning (i.e. no names are noted),
- b) at the transcription stage (here a decision will be necessary if the raw data is to be removed or retained, about allowing access to names, etc.) or
- c) just before publishing.

Of course, the best procedure is to decide at an early stage, e.g. by designing the research outline, where the main data protection procedure will be realised and bearing the issue in mind throughout the whole research process. In the following table, there is a list of possible questions for each stage of the research cycle (see Figure 17).

Rock (2001: 1-2) advises against the blind implementation of simple definitions of data protection as introduced at the beginning of this section. He lists two potentially dangerous assumptions that underpin the definition of data protection as the concealment of the informant's name: firstly, by emphasising "name", it is implied that anonymisation is reached by just altering or removing names, whereas other personal details remain in the data. Secondly, altering and removing names is equated with concealing identity; but other features, such as typical terms or expressions, can also betray identity.

The aim of data protection is to save as much information as possible in a way that is compatible with the data protection requirements and ethical guidelines (without losing information, for reasons of effective analysis, that might prove to be useful at a latter stage). Names and addresses in particular, but also the date and place of recording, unusual topics or turns of phrase, are removed from, or modified, in the raw data set (Rock, 2001: 9). Rock (2001: 9) emphasises that whatever informants wish to be removed or replaced must, in any case, be concealed or modified. In his paper, Rock (2001: 21) presents a check-list for linguistic data, which is also useful for other qualitative research. He finally concludes that the items to be removed or replaced depend on the data, the data users, the research question and – last but not least – the informants' needs (2001: 9-10).

Fig. 17: Questions for designing data protection procedures within the research process

Stages in the re-search process	Questions relevant for data protection
Defining research questions	Is the research question basically compatible with data protection? Will there be enough informants prepared to talk about this issue?
Formulation of hypothesis and operational definitions	Can the required data be processed and presented in a way that does not violate informants' privacy?
Methodological considerations	How can confidentiality be established? Which information is passed to participants before obtaining consent? How is consent obtained? (oral or written; either way, it must be explicit)
Data collection	Should names, income, and other potentially sensitive data be collected? Is this information really needed? Are there any questions which require a reconfirmation, during or after the interview, of the consent given initially?
Data preparation	Which items have to be removed or replaced? Should data anonymisation be made during transcription or afterwards? Has raw data been removed or maintained (in order to allow access to original data)? If the data will be made accessible for secondary use: are the people who have access to the data set reliable? Which control mechanisms are needed?
Analysis and interpretation	In the case of changing research questions and aims, or if secondary use is planned: can the informed consent be extended to the new research questions and aims, or is new consent needed?
Reporting and publication	Is there any sensitive data that still remains in the text? Is it necessary to consult informants again before publication? Who will read the article? Are special sanctions or measures required for this specific audience?

Source: own draft.

### How is the data modified?

Identifying details can simply be deleted or concealed; or they can be modified by replacing them with pseudonyms, for instance. Rock (2001: 11-15) presents a variety of suggestions for pseudonyms, as shown in Figure 18. He basically differentiates between the alternatives of using codes or names as pseudonyms on the one hand, and removing a part of the identifier as a partial anonymisation on the other hand. While research reports with names instead of codes are easier to read, codes allow the introduction of a certain logical structure into the pseudonyms. A basic code, for instance,

can include the information of location, sex and individual. This information is transparent only to the researcher. The code VP3 may stand for the third female participant interviewed in the city of Zurich.

Fig. 18: Typology of pseudonyms

Type	Name	Explanation	Example
Code	basic codes	Any variable or combination of variable can be used	'XXX', 'XXY', 'XYX', ...
	sex-based codes	The sex of the informant appears in an indexed code	'f1', 'm1' (f for female, m for male)
	role codes	This pseudonym indicates the role or function of the informant during the interaction	'd1', 'p1' (d for doctor, p for patient)
	name-based codes	Real names are replaced by the initials	'K. M.' for Karin Meier
Replacing names with names	fictitious names	Imaginary names, which do not exist in the real world are used. This method makes any confusion impossible	'Brunnenwil' for Hergiswil
	non-fictitious names	Existing names different from the informant's or place's name. Regarding personal names, the sex representation is often maintained, i.e. female names are always replaced by female names.	'Sarah Müller' for Karin Meier
Partial anonymisation		One part of the name is maintained.	'Karin' for Karin Meier

Source: own draft based on Rock (2001: 11-15).

For digital text, search-and-replace techniques render the procedure quite efficient. Finally, each transcript and interview note should be carefully proof-read to ensure that other more subtle but obvious clues to a character, place or institution have been removed (Corti et al., 2000: chap. 4.3). This, of course, is the most resource-intensive process.

What may not on any account be missed out of an anonymised research report is the indication that names (or whatever) have been anonymised.

## Conclusion

This contribution has shown the manifold aspects and nuances of data protection in social science. It has hopefully been made clear that data protection is a complex and dynamic endeavour that might require intensive communication between researchers and informants. Data protection is a complex issue because, on the one hand, people have different needs of privacy and, on the other hand, data protection may conflict with the research requirements of information. Additionally, data protection is a dynamic process because ethical considerations should form an ongoing component of the research process. In this article, we have particularly focused on seeking consent and data anonymisation as the two main elements of data protection.

The dynamic character of seeking consent is of special importance for qualitative research. As the focuses of research often shift during the research process, it might be necessary to obtain consent a second time. Other reasons for reconfirming consent are requests from other researchers to use the data set too, or because the study will be published, when this was not the initial intention.

However, data protection does not mean achieving absolute anonymisation in every case. For each research project, an adequate level of data protection has to be found. Data sets that will be archived and made accessible to other scientists or published require a higher level of data protection than data sets for exclusive analysis by the researcher who collected the data or ones that will not be published. Furthermore, we should bear in mind that anonymisation is not the only strategy to preserve confidentiality (ESDS, 2004b). Combining it with other strategies might allow a lower level of anonymisation and thus save important information. To choose the appropriate level of anonymisation is a task that requires flair and sensitiveness on the researcher's part. It is an ethical responsibility of each scientist. The key issue of data protection in social science is to address the issue when interviewing people and readdress the issue as and when required. For example, if an atmosphere of confidentiality is established during an interview and the informant reveals very personal information, consent should be obtained again at the end of the interview. However, as we have stressed in this article, obtaining consent does not absolve researchers from their responsibilities, as we can never assume that our informants can fully assess the consequences of our research, however well informed they may have been at the time.

## Bibliography

- Barnes, J. (1979): *Who Should Know What? Social Science, Privacy and Ethics*. Cambridge University Press.
- Bishop, L. (2005): Protecting Respondents and Enabling Data Sharing. Reply to Parry and Mauthner. In: *Sociology*, 39(2), pp. 333-336.
- British Sociological Association (BSA) (2002): Statement of Ethical Practice. From: [http://www.britisoc.co.uk/new\\_site/user\\_doc/Ethicsguidelines2002.doc](http://www.britisoc.co.uk/new_site/user_doc/Ethicsguidelines2002.doc), 24-03-2005.
- Böhnisch, T. (1999): *Gattinnen. Die Frauen der Elite. Kritische Theorie und Kulturforschung*, Vol. 2. Westfälisches Dampfboot, Münster.
- Corti, L., Day, A. & Backhouse, G. (2000): Confidentiality and Informed Consent. Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives. From: [http://www.qualitative-research.net/fgs\\_texte/3-00/3-00cortietal-e.htm](http://www.qualitative-research.net/fgs_texte/3-00/3-00cortietal-e.htm), 23-03-2005.
- Economic and Social Data Service (ESDS) (2004a): Ethical and Legal Considerations. From: <http://www.esds.ac.uk/aandp/create/ethical.asp>, 31-03-2005.
- Economic and Social Data Service (ESDS) (2004b): Identifiers and Anonymisation. Dealing with Confidentiality. From: <http://www.esds.ac.uk/aandp/create/identguideline.asp#back>, 31-03-2005.
- Eidgenossenschaft, B.d.S. (1992): Bundesgesetz über den Datenschutz (DSG) vom 19. Juni 1992 (Stand am 3. Oktober 2000).
- Eidgenossenschaft, B.d.S. (1992): Bundesstatistikgesetz (BStatG) vom 9. Oktober 1992 (Stand am 18. April 2000).
- Gamper, S. (2004): *Institutional Arrangements of Forest Use in Nepal. Local Realities Beyond Community Forestry*. Master Thesis, University of Zurich, Zurich.
- Grinyer, A. (2002): The Anonymity of Research Participants. Assumptions, Ethics and Practicalities. In: *Social Research Update*, 36. From: <http://www.soc.surrey.ac.uk/sru/SRU36.html>, 23-03-2005.
- Hauser, R., Wagner, G.G. & Zimmermann, K.F. (1998): Memorandum on the Requirements for Empirical Economic Research and for Empirically Supported Economic and Social Policy Advice. A revised and updated version of a statement published in *Allgemeines Statistisches Archiv*, 82, pp. 369-379.
- Kluge, S. & Opitz, D. (2000): Computer-aided Archiving of Qualitative Data with the Database System "QBiQ". In: *Forum Qualitative Research*, 1, 3. From: [http://www.qualitative-research.net/fqs\\_texte/3-00/3-00klugeopitz-e.htm#g41](http://www.qualitative-research.net/fqs_texte/3-00/3-00klugeopitz-e.htm#g41), 02-05-2003.
- Lancaster University (2002): Research. From: <http://www.dpa.lancs.ac.uk/approved/Research.doc>, 02-05-2003.
- Landolt, S. (2004): *Das sozialgeografische Milieu und die Wahrnehmung des Fremden*. Unpublished master thesis, University of Zurich, Zurich.
- Miller, T. & Bell, L. (2002): Consenting to What? Issues of Access, Gate-keeping and "Informed" Consent. In: Mauthner, M., Birch, M., Jessop, J., Miller, T. (Editors), *Ethics in Qualitative Research*. Sage Publications, London, Thousand Oaks, New Delhi, pp. 53-69.
- Müller, U. (2001): *Wie funktioniert Partizipation bei Naturschutzvorhaben in der Schweiz? Untersucht am Beispiel der Erweiterung des Schweizerischen Nationalparks*. Unpublished master thesis, University of Zurich, Zurich.

## Gesellschaft und Raum – Konzepte und Kategorien

- Müller, U. & Kollmair, M. (2004): Die Erweiterung des Schweizerischen Nationalparks: Der Planungsprozess 1995-2000, betrachtet aus partizipationstheoretischer Sicht. *DISP*, 159(4/2004): 44-51.
- Parry, O. & Mauthner, N.S. (2004): Whose Data are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. In: *Sociology*, 38(1), pp. 139-152.
- Rock, F. (2001): Policy and Practice in the Anonymisation of Linguistic Data. In: *International Journal of Corpus Linguistics*, 6(1), pp. 1-26.
- Sin, C.H. (2005): Seeking Informed Consent. Reflections on Research Practice. In: *Sociology*, 39(2), pp. 277-294.
- Steimann, B. (2003): Decentralization and Participation in the Forestry Sector of NWFP, Pakistan. The Role of the State. Unpublished master thesis, University of Zurich, Zurich.
- Zaugg Stern, M. (2006): Philosophiewandel im schweizerischen Wasserbau – Zur Vollzugspraxis des nachhaltigen Hochwasserschutzes. Schriftenreihe Humangeographie, Vol. 20. Geographisches Institut, Universität Zürich, Zürich.
- Zaugg, M., Ejderyan, O. & Geiser, U. (2004): Normen, Kontext und konkrete Praxis des kantonalen Wasserbaus. Resultate einer Umfrage zu den Rahmenbedingungen der kantonalen Ämter oder Fachstellen für Wasserbau bei der Umsetzung der eidgenössischen Wasserbaugesetzgebung. Schriftenreihe Humangeographie, Vol. 19. Geographisches Institut, Universität Zürich, Zürich.