

**Central Questions of Anonymization:
A Case Study of Secondary Use of Qualitative Data**

Denise Thomson, Lana Bzdel, Karen Golden-Biddle, Trish Reay & Carole A. Estabrooks

Abstract: Anonymization—the removal of identifying information from data—is one way of preparing data for secondary use. This process has not received much attention from scholars, but close examination shows that it is full of methodological, ethical and theoretical tensions. Qualitative research focuses on how people live and act in very particular, situated contexts. Removing identifying information also, inevitably, removes contextual information that has potential value to the researcher. We propose to present a case study of working with anonymized data on the research project, *Knowledge Utilization and Policy Implementation*, a five-year program funded by the Canadian Institutes of Health Research. This project involves the secondary use of qualitative data sets from multiple separate research projects across Canada. Based on this case study, we provide useful recommendations that address some of the central questions of anonymization and consider the strengths and weaknesses of the anonymization process.

Key words: *ethical practice, collaborative research, confidentiality, privacy*

- 1. Introduction
 - 2. Secondary Analysis to Support Knowledge Utilization
 - 3. Literature Search
 - 4. Emergent Questions
 - 4.1 Question 1—What are the alternatives to anonymization?
 - 4.2 Question 2—What is anonymization in the context of secondary use of qualitative data?
 - 4.3 Question 3—How can researchers best anonymize qualitative data for secondary use?
 - 4.4 Question 4—How much anonymization is enough?
 - 5. Conclusions
- Acknowledgements
References
Appendix 1: Literature Search Methodology
Authors
Citation

1. Introduction

In secondary use, we need to find a balance between honoring commitments of confidentiality made to participants at the time the data was originally gathered, while still retaining the usefulness of the data in the development of further knowledge. Anonymization—the removal of identifying information from data—is one way of creating this balance. Anonymizing is a part of qualitative work that does not receive much attention, yet close analysis shows that the process is full of methodological, ethical and theoretical tensions. Qualitative research focuses on how people live and act in very particular, situated contexts. Removing identifying information also, inevitably, removes contextual information that has potential value to the researcher. [1]

We propose to present a case study of working with anonymized data on the research project, *Knowledge Utilization and Policy Implementation*, a five-year program funded by the Canadian Institutes of Health Research. This project involves the secondary use of qualitative data sets from multiple separate research projects across Canada. Based on this case study, we will provide useful recommendations that address some of the central questions of anonymization and consider the strengths and weaknesses of the anonymization process. [2]

2. Secondary Analysis to Support Knowledge Utilization

We became interested in anonymization and secondary use of data through our work on the Knowledge Utilization and Policy Implementation research project (KUPI). KUPI is a geographically-dispersed, interdisciplinary, collaborative research project involving researchers across Canada who are interested in the issues of knowledge utilization and policy implementation, particularly in health care settings. Dr. Carole ESTABROOKS, of the Faculty of Nursing at the University of Alberta, is the Principal Investigator. There are also three co-investigators, at various universities across the country, contributing the distinct disciplinary perspectives of organizational analysis, political science and sociology. [3]

The purpose of KUPI is to develop theoretical foundations for knowledge utilization (KU) to enable relevant knowledge use by health care practitioners and decision-makers. The work of this research will help in implementing successful strategies that increase the use of relevant knowledge in health care decision and policy-making processes. It is our hope that ultimately, improvements in health policy implementations at the clinical, organizational, and regional levels will lead to overall improvements in patient and system outcomes. [4]

The KUPI project depends on secondary use of data that was originally gathered for distinct projects conducted by the members of the research team. The researchers are applying multi-methods and multi-level analysis to these existing datasets, many of which are qualitative. Anonymization is crucial to the KUPI project, not only for coordination of data analyses across researchers and universities, but also and significantly so that data preparation meets the guidelines binding Canadian researchers. The three federal granting councils in this country have jointly published the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (henceforth referred to as the *Tri-Council Policy Statement*, Medical Research Council of Canada [MRCC], 2003). Canadian academic institutions have generally adopted the ethical standards of the *Tri-Council Policy Statement*. Its position on secondary use of data hinges on the question of whether identifying information is present in the data. If yes, the researchers must seek approval for the project from the relevant Research Ethics Board (REB), by submitting an application that outlines how the researchers propose to address the following requirements:

- demonstrating that the identifying information is essential to the research;
- documenting the measures that will be taken to "protect the privacy of the individuals, to ensure the confidentiality of the data, and to minimize harms to subjects"; and
- showing that the "individuals to whom the data refers have not objected to secondary use" (MRCC, 2003, p.3.5).
- Furthermore, the REB may also require that access for secondary use to data with identifying information be dependent on:
 - informed consent of those who contributed data, or authorized third parties; or
 - an appropriate strategy for informing the subjects; or
 - consultation with representatives of those who contributed data (MRCC, 2003, p.3.5). [5]

The options outlined in the *Tri-Council Policy Statement* for working with data with identifying information present were not feasible for us. Our data sets had been gathered from a number of different projects across the country and across a considerable period of time. It was not practical for us to undertake the task of re-contacting all the participants to ascertain whether they objected to secondary use of the data, and obtain their consent for doing so. Consequently, to meet both the letter and spirit of the guidelines, we had to *anonymize*—that is, remove the identifying information from—the data. We decided that we wanted to conduct the anonymization process according to the best practices identified in the literature to date, and therefore conducted an extensive search. [6]

3. Literature Search

Methodology. To orient us to the current issues and practices of secondary use of qualitative data, we conducted an extensive review of the literature dating from 1990. We searched bibliographic databases in: business, sociology, psychology, anthropology, linguistics, health sciences, medicine, and knowledge & information systems. We used the following search terms (and variants of them) to retrieve relevant articles related to: qualitative research methodology, secondary use, anonymization, privacy, confidentiality, ethical practice, collaborative research, and archiving. In addition, we searched university library catalogues for books on "qualitative research methodology." We hand-searched or reviewed electronically the table of contents and indexes of over 200 books either in print or online for relevant discussion of key concepts. Lastly, we also employed other search techniques, such as focused searches using Internet search engines, and "pearl growing," a process whereby the reference lists of exemplar articles are reviewed for additional resources of relevance. A detailed overview of our search strategy is attached as [Appendix 1](#). [7]

Through the course of our search, we developed a broad collection of articles to support our understanding of the secondary use of qualitative data. We found three main areas of helpful material. Some material dealt directly with how to anonymize qualitative data for secondary use (CORTI, DAY, & BACKHOUSE, 2000; ESDS QUALIDATA, 2004; HEATON, 2004; Inter-university Consortium for Political and Social Research [ICPSR], 2002). Also, we drew on three articles that dealt with anonymization specifically for publication (HOPKINS, 1993; NESPOR, 2000; SHULMAN, 1990) and one article that covered both uses (ROCK, 1999). Finally, we found it useful to consider some concepts from the arena of anonymizing survey data (ICPSR, 2002). This material was all helpful, and we will discuss it in more detail later in this paper. [8]

Ultimately, however, we found very little guidance to address the range of questions that were emerging for the KUPI team as data anonymization commenced. For our purposes, the silence in the literature about anonymization was compounded by a lack of institutional guidance in Canada for the re-use of qualitative data in general. Canada does not have an institution such as ESDS Qualidata (2004), a specialist service of the United Kingdom's Economic and Social Data Service that provides access and support for a range of social science qualitative datasets. Consequently, we had much work to do to find our own answers to the questions facing us. We believe a description of the process we went through may be of benefit to other researchers as they contemplate projects involving secondary use. [9]

4. Emergent Questions

As we searched the literature, conducted early anonymization of data and discussed our efforts, four central questions emerged for us:

1. What are the alternatives to anonymization?
2. What is anonymization, in the context of secondary use of qualitative data?
3. How can researchers best anonymize qualitative data for secondary use?
4. What is *enough* anonymization? [10]

The rest of this article examines these questions in the light of both our literature review and practical experience. [11]

4.1 Question 1—What are the alternatives to anonymization?

Anonymization of data is carried out in order to protect the privacy of research participants, while making the data accessible by researchers. We will discuss this point in more detail later in this paper, but here we want to review some alternatives to anonymization that have been discussed by other scholars. These alternatives address the need to balance the issues of

harm to subjects and the usefulness of material. In our particular situation, we were not able to use any of these alternatives for our research, given the nature of our ethics requirements and the practical realities of our data sets. However, because it is critical for research teams to explore all options of working with qualitative data in secondary use, we present them here for other researchers' consideration. [12]

The first alternative involves reliance on inter-researcher trust: a concept whereby other researchers are trusted to behave ethically. In other words, if we believe that we ourselves can be trusted with the data, we should extend at least some of that confidence to other researchers. This confidence can be strengthened with signed agreements. For example, the ESDS Qualidata service in the United Kingdom requires all secondary users sign an undertaking promising not to breach confidentiality, either by including identifying information in publications or by attempting to contact the people who participated in the original research (CORTI et al., 2000). Similar agreements could presumably be established between researchers working collaboratively. Although the KUPI team members signed confidentiality agreements guiding how they should treat the research data, this option clearly did not fall within the scope of the *Tri-Council Policy Statement* and thus was not a feasible option for us. [13]

A second alternative gives the original researchers a degree of control over the access to, and use of, the data by secondary researchers. For example, the ESDS Qualidata archive offers depositors "gate-keeping" options, limiting the possibility for misuse by restricting who can see the data (CORTI et al., 2000). Additionally, ROCK (1999, p.15) suggests that the original researchers could "flag" sections of data that are unsuitable for publication. According to ROCK, this compromise, while relying on the secondary researchers' goodwill, would allow a balance between preserving anonymity and, simultaneously, the integrity of the text. However, for both this and the previous alternative the substantial obstacle for Canadian researchers is the same: the lack of existing models in this country for such agreements would impose a considerable burden on pioneers who would be forced to incur the expense of time and money to figure out the practical, legal and ethical dimensions of this sort of contract in our jurisdiction. Committing these resources, which would come at the cost of financing actual research, is simply not practical. [14]

A third possibility is that researchers could seek permission from participants for secondary use at the time of the original contact. The consent form would have a section where the participant could explicitly agree to having the (for example) interview transcript deposited in an archive or otherwise made available for secondary use. A sample consent form given by the UK DATA ARCHIVE (2003) in their *Qualidata Acquisitions Pack* gives the participant three options: (1) giving permission to have the transcribed interview archived for research use; (2) archived for teaching use only; or (3) not deposited at all. Such detailed consent forms might help alleviate the concerns of those who feel that asking for permission for secondary use at the time of original contact, when secondary use is still hypothetical and undefined, stretches the concept of "informed consent" beyond reasonable bounds. "Obtaining unqualified, blanket consent for undefined future health research purposes is empty and meaningless and may sometimes reduce, rather than increase, privacy protection" (Canadian Institutes of Health Research [CIHR], 2003; see also CORTI et al., 2000; HEATON 2004). However, this option might pose problems for researchers seeking approval from Research Ethics Boards in disciplines without a tradition of archival depositing (particularly for health research), where it might raise concerns that would delay or prevent approval of the entire application. For KUPI, this option was not possible, as the data had already been gathered when secondary use was proposed. [15]

A final alternative involves a variant on the idea that "time heals all wounds"—the suggestion that the passage of time may well lessen the likelihood of secondary use causing any wounds to participants' privacy. CORTI et al. (2000, para. 15) suggest that the discussions about secondary use need primarily to be focused on the short term, as over time the material gradually becomes part of the "historical record" and therefore less problematic. However, the guidelines outlined by the *Tri-Council Policy Statement* do not encompass this possibility. Since none of these options were feasible for us, we focused on anonymization as the best means for preparing our data for secondary analysis. [16]

4.2 Question 2—What is anonymization in the context of secondary use of qualitative data?

Anonymization, most simply, is rendering research participants anonymous by removing identifying information from research data. Data can be anonymized for at least two main purposes: (1) *publication*—where excerpts quoted in published material have details changed so that the reader does not know who is being quoted—or (2) *secondary use*. Anonymization is often used in quantitative research, where it is a straightforward process, at least from the perspective of qualitative researchers. Removing identifying information from data collected through means such as surveys entails removing direct identifiers, including names, addresses and other linkable identification numbers. Indirect identifiers, such as geographical or educational information, can be treated so as to retain the information but lessen the linkage to an individual participant. Common ways of treating the data include:

- Removal: Eliminating the variable from the dataset entirely
- Bracketing: Combining the categories of a variable
- Top-coding: Restricting the upper range of a variable
- Collapsing and/or combining variables: Merging the concepts embodied in two or more variables by creating a new summary variable (ICPSR, 2002) [17]

What is the equivalent process for qualitative data such as interview transcripts or field observation notes? Typical descriptions of anonymization that we found focus on the notion of removing "information that breaches the confidentiality of the respondent or any other person or entity" (UK DATA ARCHIVE, 2002, p.12). In general, the recommendations are along the lines of those advanced by CORTI et al. (2000, para. 21.), who suggest removing "identifying details," such as proper names or street names, and replacing these with pseudonyms. In theory, then, the data remaining after anonymization tells us a story without telling whose story it is. [18]

All of the material we found on anonymization, regardless of the purpose, contains the same warning—the message that "efforts to disguise the identity of informants may also spoil and distort the data" (HEATON, 2004, p.83; see also CORTI et al., 2000; NESPOR, 2000; ROCK, 1999). This is because information that identifies a participant is often information that is of interest to the researcher; furthermore, removing portions of the data changes the context of the remaining text. A key issue here is perspective. Each of the researchers on the team had a different idea of what was important, and hence what had to be retained, and what could be anonymized. For example, in interdisciplinary discussions between researchers in business and nursing, the business researchers proposed replacing the words "nurse" and "physician" with the broader category of "health care professional" for the sake of preserving the individual professionals' anonymity. From an organizational analysis perspective, concerned with the professions as a group, this was a possibly acceptable solution. Our nursing colleague objected, however, saying that a participant's identity as a nurse or physician was crucial to data analysis. Thus, even though we were aware of the broad research questions for which this data is to be used, we still had difficulty reaching consensus about what was important to retain. [19]

In light of this background about what constitutes anonymization, we sought to establish the best way of proceeding, which lead us to our next question. [20]

4.3 Question 3—How can researchers best anonymize qualitative data for secondary use?

We found some interesting practical recommendations in the literature. The *Qualidata Process Guide* (UK DATA ARCHIVE, 2002) lists techniques that include:

- The removal of major identifying details, such as real names, place and company names, street names, and replacement with pseudonyms;
- the use of the same pseudonyms and place names used in any prior publication based on this data;

- a cross-referencing system linking pseudonyms to original names (this is not to be made available to secondary users);
- the deletion or withholding of problematic materials such as slanderous or libelous comments about third parties;
- the retention of the original data in all cases (p.13). [21]

As well, ROCK (1999) has created the prototype of a scheme for assessing individual datasets for anonymization. This scheme is developed primarily for anonymizing linguistic datasets, not specifically for secondary use, although ROCK recognizes this as one potential use for the anonymized result. Her framework works through a wide range of aspects to think about, which ROCK has broadly classified into "data-based considerations" and "research needs and wants." Examples of the former include the sensitivity of material under discussion, comments made about third parties, the situation in which the data was originally gathered and the form of consent given at the time of original collection. Under the second category, ROCK has created three sub-categories:

1. participant needs—the ethical and legal issues surrounding anonymization;
2. researchers' wants—methodological and theoretical considerations; and
3. other constraints, which are the practical aspects: financial and temporal (pp.21-23). [22]

ROCK's scheme is primarily useful for raising questions to consider, thereby implicitly showing the complexity of anonymization as an issue, rather than suggesting answers. [23]

Using this literature, particularly the *Qualidata Process Guide* (UK DATA ARCHIVE, 2002), as a basis, the KUPI team developed an initial protocol for anonymizing our text. Key points of this protocol were:

- Replacement
 - Replace direct identifiers such as place, organization, position title, number of years of experience, dates and any others.
 - Replace names with numbers to avoid the possibility of changing one name to the name of another participant who works in the same organization. Ensure that participants with the same name are not given identical numbers.
 - Establish a procedure for deleting and/or replacing text. Include a legend in the header of the document (e.g., D=Date, N=Name, L=Locations).
- Pilot test
 - Have assistants pilot test a sample of the interviews to get a sense of what identifiers need to be anonymized. Create an inventory of identifiers that need anonymizing for each sample document. This should save time in the long run.
- Proofread
 - After replacing identifying information through automated replacement techniques, additional proof reading will be required. This will prevent missing any names due to issues such as variations in the spelling of the replaced words, and will alert us to the presence of unanticipated identifiers that need to be changed.
- Database construction
 - Build a database to easily track and cross-reference names, numbers and documents (e.g. interviews and field notes). A note should be made in the database to indicate if information was deleted or altered for confidentiality reasons. [24]

This protocol assisted us greatly in doing our first anonymization work with interview data, and led to the idea of "levels" of anonymization. At each level, there are different decisions regarding how to treat identifying information in the interview text. For our work, the levels were connected to job-related identifiers, because we are studying people in the contexts of their work

environments. In other datasets, different identifiers may be more significant. Nevertheless, in our scheme, for every dataset there will be two extremes in the anonymization continuum: at one end, leaving the data untouched; at the other end, deciding that the entire item is too sensitive (i.e. induces "harm" to the participant) and omitting it entirely from the set of data made available for secondary use. We are currently envisioning this continuum as a ladder, with the central question in climbing the ladder being what identifiers to omit or replace. We have constructed a preliminary anonymization ladder with five levels. Level 0 concerns no change; leaving the data as is. Level 1 involves the removal of direct identifiers. Level 2 concerns the removal of indirect identifiers, particularly when considered in linkage with each other. Level 3 concerns the removal of some stories and life histories. Finally, Level 4—the last resort—is removing the data from the set. We are currently conducting tests of this ladder, which we will detail in a future manuscript. These tests will help us determine the integrity of the levels, especially their numbers and differentiation from each other. [25]

Interestingly, as we conducted this work, it has become increasingly clear to us that, even with the replacement of identifying details, there was still a possibility that the interviewees could be identified by the other researchers. Other researchers on our team are familiar with the health system in our province. They could easily have possessed enough contextual information to identify participants from details and stories that were remained in the interviews after the "first pass" of the basic anonymization process outlined above. Knowing this, we were not sure that we were meeting the ethical standard set by the *Tri-Council Policy Statement* (2003). Accordingly, we spent considerable time considering the question of "what is *enough* anonymization?" [26]

4.4 Question 4—How much anonymization is enough?

The question of "what is enough" highlights the practical and ethical issues surrounding anonymization. The practical aspects are discussed by CORTI et al. (2000) and ROCK (1999). ROCK (1999), in her study of anonymization of linguistic data, suggests that the desirable end result of anonymization may be that only the participant can recognize him or herself in the finished data. As ROCK acknowledges, however, reaching this level of anonymization is a "large, complex task" (p.9). Elsewhere in her article she emphasizes that both financial and temporal constraints affect the level of anonymization that can be accomplished, a highly important point (p.23). CORTI et al. (2000, para. 22) suggest that the "appropriate level of anonymization" depends on the "history and nature" of the study. Furthermore, they say, each case needs to be considered individually; "in some cases, revealing the names of regions and towns may not be problematic, in other case the consequences of disclosure could be damning" (CORTI et al., 2000, para. 22). The difficulty lies in knowing the difference. Here we are faced with a core problem in anonymization—the difficulty of knowing both what will be useful to future researchers, and what will constitute identifying information to those who have not been involved in conducting the research. Even more important is the fact that researchers are required to consider the nature of harm that could accrue to participants if identification occurs. This is a key ethical issue to which we now turn our attention. [27]

The discussions about the ethics of anonymization focus on its purpose, which, we would argue, is centrally to protect research participants from the harm of having their identity revealed to anyone other than those to whom consent has been given. It is generally considered that anonymization as a process benefits research participants by protecting their anonymity and privacy, key guarantees that are central to ethical research. The *Tri-Council Policy Statement* explicitly states that, "as a general rule, the best protection of the confidentiality of personal information and records will be achieved through anonymity" (MRCC, 2003, p.3.2). For individuals to maintain their privacy, they must be able to control information about themselves (MARX, 1999, p.100). The principles of informed consent, anonymity and confidentiality limit researchers' access to the lives of research participants, in that it is the participant who decides whether, and under what conditions, to participate in the research, and does so on the basis that the information stay within tightly-controlled boundaries (HOMAN, 1991). Anonymization maintains the boundaries around personal information while allowing the rest of the data to be

used in other research contexts. Consequently, it also benefits researchers, by broadening the base of data from which they can draw (ROCK, 1999). [28]

However, this perspective is sharply critiqued by NESPOR (2000), HOPKINS (1993) and SHULMAN (1990), who all discuss anonymization in the context of preparing data for publication. NESPOR argues that anonymization is a taken-for-granted technique that needs to be scrutinized for its unacknowledged ontological and political effects. NESPOR (2002, p.564) concludes that "we have failed to adequately analyze how anonymization works as a representational practice—what it allows, what it hinders—because we have assumed that it was an obligatory ethical tactic" and argues that "even if anonymization practices worked perfectly and hid identities completely, I think we should discard them as automatic default positions and instead articulate a clearer politics behind our strategies of identification or masking." SHULMAN (1990, p.14) argues, in a discussion of research on educational issues, that the assumption that participants need to be "invisible" casts them as "powerless and in need of protection." SHULMAN (1990, p.14) rejects this assumption in favor of a stance which treats participants as "professional colleagues who deserve as much recognition as the traditional scholar" and research as a vehicle "for the professionalization and empowerment of teachers." [29]

NESPOR (2000) suggests that *not* anonymizing should be the default option for qualitative researchers, with the decision to protect identity being a decision made after consultation between researcher and participants, for cases where both sides deem it necessary. However, this approach brings with it its own tensions based on the ethical and practical issues raised by balancing the welfare of the individual participant against that of his/her own community. SHULMAN points out that the research participant may feel empowered by being named, but at the same time this involves also naming the participant's community, the members of which may be embarrassed by such publicity. HOPKINS (1993) discusses the other side of this dynamic, in which, even if the individual him/herself is embarrassed, the community may benefit if reporting on its living conditions leads to change. However, HOPKINS (1993, p.125) adds the cautionary note that "once a community is publicly identified, that revelation cannot be retracted when the ethnographer wants to publish more sensitive material at a later date." [30]

There is an interesting convergence, in all the work we reviewed, around the key role that research participants can play in anonymization decisions. ROCK (1999) reminds researchers to be cautious of relying on their own subjective opinions of what information will be considered sensitive by participants, and suggests that researchers should consider asking participants what should be anonymized. Indeed, the guidelines set out in the *Tri-Council Policy Statement* (2003) implicitly accommodate some of these thoughts by allowing the presence of identifying information in research data if participants explicitly consent to this, a position which allows participants to choose to be named if they so wish. For the purposes of the KUPI study, where the datasets were not originally gathered with the notion of secondary use, and where recontacting participants was too large a task, we could not consider using non-anonymized data. [31]

5. Conclusions

What we have found, and presented here, shows the complexity of anonymization and the number of methodological, practical and ethical issues that need to be considered by researchers who expect to undertake this work. Summarizing our conclusions from working through the four questions explored in this paper, we found that researchers need to resolve the following issues when preparing data for secondary use:

- *In secondary use, what measures do we need to take to protect our participants and retain the usefulness of the data?* For us, anonymization was the best alternative given the environment in which we work. Others may be in a position to consider options such as contacting participants for consent or obtaining consent for secondary use at the time of the initial interview.

- *If anonymization, what amount of identifying information needs to be removed—in other words, what is "enough" anonymization?* Here the balance is between retaining context and protecting participants. We return to the statement of CORTI, DAY and BACKHOUSE (2000), quoted above, that we must consider the consequences of disclosure, a concept that is rooted in the idea that identification of participants can cause them harm. This possible harm can accrue not only to the individual participant him/herself, but also to the community in which s/he lives or works. We eventually settled on a range of options, including pseudonyms, selective deletion of particular passages, and completely omitting particular interviews from the dataset to be shared if we felt that anonymization was not possible in that instance. This framework is a work in progress for us, not a final product.
- *What is the impact of anonymization on the secondary use process?* There is no doubt that anonymization lessens the situated, contextual nature of qualitative data. The crucial issue, and one we have not seen explored elsewhere, is: what effect does the use of anonymized data have on our data analysis process? As we continue with our data analysis phase, we are monitoring our impressions of how the use of anonymized data is affecting our analysis. This work will be of immediate benefit to our research, as we refine our anonymization decisions, and will also, we expect, benefit other researchers who can learn from our experiences. [32]

All of these questions remain very much alive for us and we expect to devote considerable attention to them as the KUPI project unfolds. We hope that other researchers will feel moved to contribute to the dialogue about the role of anonymization in preparing qualitative data for re-use. [33]

Acknowledgements

This paper was made possible in part by the CIHR Knowledge Utilization and Policy Implementation research project (KUPI), the Health Organization Studies (HOS) research group at the University of Alberta School of Business, and the Knowledge Utilization Studies in Practice (KUSP) research group at the University of Alberta Faculty of Nursing. We would like to thank Chuck HUMPHREY for his insights, as well as a professor in the Faculty of Arts at the University of Alberta who wishes to remain anonymous.

References

Canadian Institutes of Health Research (CIHR) (2003, April). *Executive summary of secondary use of personal information in health research: Cases studies*. Ottawa, Canada: CIHR. Retrieved May 21, 2004, from <http://www.cihr-irsc.gc.ca/e/about/6827.shtml>.

Corti, Louise, Day, Annette, & Backhouse, Gill (2000). Confidentiality and informed consent: Issues for consideration in the preservation of and provision of access to qualitative data archives [46 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], 1(3), Art. 7. Retrieved June 27, 2004, from <http://www.qualitative-research.net/fqs-texte/3-00/3-00cortietal-e.htm>.

ESDS Qualidata (2004, June). *ESDS Qualidata* [Website]. Retrieved June 27, 2004, from <http://www.esds.ac.uk/qualidata/about/introduction.asp>.

Heaton, Janet (2004). *Reworking qualitative data*. London: Sage.

Homan, Roger (1991). *The ethics of social research*. London: Longman.

Hopkins, MaryCarol (1993). Is anonymity possible? Writing about refugees in the United States. In Caroline B. Brettell (Ed.), *When they read what we write: The politics of ethnography* (pp.121-129). Westport, CT: Bergin & Garvey.

Inter-university Consortium for Political and Social Research (ICPSR). (2002, March). *Guide to social science data preparation*. Retrieved June 27, 2004, from <http://www.icpsr.umich.edu/access/dpm.html>.

Marx, Gary T. (1999). What's in a name? Some reflections on the sociology of anonymity. *The Information Society*, 15(2), 99-112.

Medical Research Council of Canada, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada (2003). *Tri-council policy statement: Ethical conduct for research involving humans*. Ottawa: Medical Research Council of Canada.

Nespor, Jan (2000). Anonymity and place in qualitative inquiry. *Qualitative Inquiry*, 6(4), 546-569.

Rock, Frances (2001). Policy and practice in the anonymisation of linguistic data. *International Journal of Corpus Linguistics*, 6(1), 1-26.

Shulman, Judy H. (1990). Now you see them, now you don't: Anonymity versus visibility in case studies of teachers. *Educational Researcher*, 19(6), 11-15.

UK Data Archive (2002, April). *Qualidata process guide*. Retrieved June 27, 2004, from <http://www.qualidata.essex.ac.uk/docs/dataprocess.pdf>.

UK Data Archive (2003). *Qualidata acquisitions pack: IASSIST Workshop*. Essex, UK: University of Essex.

Appendix 1—Literature Search Methodology

The following is an inventory of the search strategies employed, as well as the resources that were reviewed to orient us to the current issues and practices of secondary use and anonymization of qualitative data.

1. Search Strategies

The selected *key concepts* and search strategies were intended to retrieve a reasonable number of relevant items while minimizing the number of irrelevant items retrieved. While many searches were conducted for this review, the following represents the general approach taken when searching for conceptual and empirical works related to the secondary use and anonymization of qualitative data.

Concept 1	Concept 2	Concept 3	Concept 4	Concept 5
Qualitative research	Secondary use	Ethical practice	Collaborative research	Context
Related Terms				
Analysis Inquiry Methodology Methods Interview transcripts	Archiving	Anonymization Privacy Confidentiality Informed consent Harm	Interdisciplinary teams	Meaning

Table 1: Key concepts

Where appropriate, different search strings were used in conjunction with different bibliographic databases and library catalogues to make efficient use of the controlled vocabularies of the various databases. The literature search strategy for this review excluded items published in languages other than English, and most items prior to 1990.

We initially ran a simple search strategy across multiple databases to determine the relevance of each to this review. To determine relevance we looked at the number of relevant retrievals from the particular database and also the uniqueness of the items retrieved from the database (i.e., if a database contained a small number of highly relevant items that were not indexed in

other databases, it was included for review). We then focused on detailed search strategies on the most appropriate databases.

1.1 Simple search string

1. "qualitative*"
2. "secondary use" OR "anonymi*"
3. 1 AND 2
4. limit 3 to title and abstract

1.2 Extended search string

1. qualitative* AND (research OR analys* OR inquiry OR method*)
2. "secondary use" OR archiv*
3. ethic* OR anonymi* OR privacy* OR confidential* OR "informed consent"
4. 2 OR 3
5. 1 AND 4
6. limit 5 to title and abstract; 1990-present; English language
7. collaborat* OR context*
8. 6 AND 7

2. Resources

2.1 Bibliographic databases

Due to the multidisciplinary nature of secondary use of qualitative data, a wide variety of bibliographic databases, key journals, libraries and library catalogues, and Internet resources were investigated:

Database	Date Range	Discipline
Business Source Premier	1922-Present	Business
ABI Inform	1970-Present	Business, Interdisciplinary
Academic Search Premier	1975-Present	Interdisciplinary
ISI Web of Science ¹	Art & Humanities Citation Index® (1975-present) Science Citation Index Expanded™ (1945-present) Social Sciences Citation Index® (1956-present)	Arts & Humanities, Science and Social Sciences
MEDLINE	1966-Present	Medicine and Health Sciences
CINAHL	1982-Present	Nursing and Allied Health
PsychINFO	1985-Present	Psychology

Table 2: Bibliographic databases

¹ The ISI *Journal Citation Reports* and the *Cited Reference Search* were also searched to help determine and gauge the relative quality of various journal publications, and associated "works" by key authors and contributors.

2.2 Key journals

Throughout the course of our search, we identified the following journals as key outlets for publishing relevant articles relating to the issues associated with the secondary use and anonymization of qualitative data:

- *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*
- *International Journal of Nursing Studies*
- *International Journal of Social Research Methodology*
- *Qualitative Health Research*

2.3 Special journal issues

The individual article abstracts from the following special journal issues/supplements were scanned for relevance based on the *key concepts* identified above:

- Text . Archive . Re-Analysis (2000, December). *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1 (3).
- Privacy, Data and Health Research (2003, July). *Journal of Health Services Research & Policy*, 8 (3 Suppl).
- Special Issue—Celebrating Classic Sociology: Pioneers of Contemporary British Qualitative Research (2004, February). *International Journal of Social Research Methodology: Theory and Practice*, 7 (1).

2.4 Libraries and catalogues

The following library catalogues and libraries were searched for books on "qualitative research methodology" in addition to the *key concepts* identified above. The "tables of contents" of over 200 books were either hand-searched or reviewed electronically online for relevant discussion of *key concepts*.

- NEOS Library Consortium Catalogue, Alberta, Canada, <http://www.neoslibraries.ca/index.aspx>
- University of Alberta Libraries, Edmonton, Alberta, Canada, <http://www.library.ualberta.ca>
- International Institute for Qualitative Methodology Library, Edmonton, Canada, <http://www.ualberta.ca/~iiqm/>
- OCLC WorldCat

2.5 Key authors

The following key authors were searched individually in the stated bibliographic databases and on the Internet (i.e., professional websites):

Author	Affiliation
Corti, Louise	University of Essex, UK
Fielding, Nigel G.	University of Surrey, UK
Hammersley, Martyn	Open University, UK
Heaton, Janet	University of York, UK
Humphrey, Charles K.	University of Alberta, Canada
Mauthner, Natasha S.	University of Aberdeen, UK
Morse, Janice M.	University of Alberta, Canada
Nespor, Jan K.	Virginia Tech, USA
Rock, Frances	University of Surrey, UK
Thompson, Paul	University of Essex, UK
Thorne, Sally	University of British Columbia, Canada

Table 3: Key authors

2.6 Internet

Google was the primary search engine used to search for information about Canadian and international data archives, services and institutions. The following is a sample list of key sites that were identified and scanned for information about the secondary use and anonymization of research data (both qualitative and quantitative):

Canada	
National Research Data Archive Consultation	http://mmsd1.mms.nrcan.gc.ca/archives/default-e.asp
University of Alberta Data Archive	http://www.library.ualberta.ca/datalibrary/index.cfm
United Kingdom & Europe	
Council of Social Science Data Archives (CESSDA)	http://www.nsd.uib.no/cessda/index.html
UK Data Archive (UKDA)	http://www.data-archive.ac.uk/
ESDS Qualidata	http://www.esds.ac.uk/qualidata/about/introduction.asp http://www.esds.ac.uk/
United States	
Inter-university Consortium for Political and Social Research (ICPSR)	http://www.icpsr.umich.edu/
Institute for Social Science Research (ISSR) Data Archives	http://www.sscnet.ucla.edu/issr/da/
Murray Research Center	http://www.radcliffe.harvard.edu/murray/
International	
IASSIST—International Association for Social Sciences Information Service and Technology	http://datalib.library.ualberta.ca/iassist/

Table 4: Useful Internet resources

Authors

<p><i>Denise THOMSON</i> recently completed a Master of Business Administration degree with the Health Organization Studies research group at the University of Alberta School of Business. She is now the Program Administrator for the Child Health Field of the Cochrane Collaboration, an international not-for-profit organization providing up-to-date information about the effects of health care. She holds a MA and a BA in History from the University of Alberta.</p>	<p>Contact: Denise Thomson ARCHE, Department of Pediatrics University of Alberta 9419 Aberhart Centre One 11402 University Ave Edmonton, AB T6G 2J3, Canada E-mail: denise.thomson@ualberta.ca URL: http://www.bus.ualberta.ca/hos/</p>
---	---

<p><i>Lana BZDEL</i> is the Research and Content Coordinator for the Health Organization Studies research group at the University of Alberta School of Business. Lana's interest in the anonymization of qualitative data stems from her involvement with the Knowledge Utilization and Policy Implementation (KUPI) research initiative—a multi-disciplinary, collaborative research program investigating the linkage between knowledge use and policy implementation in health organizations. Lana BZDEL holds a Master's degree in Library and Information Studies and a Bachelor's degree from the University of Alberta.</p>	<p>Contact: Lana Bzdel Health Organization Studies University of Alberta School of Business 3-23 Business Building Edmonton, Alberta T6G 2R6, Canada E-mail: ana.bzdel@ualberta.ca URL: http://www.bus.ualberta.ca/hos/</p>
<p><i>Karen GOLDEN-BIDDLE</i> is Professor, Department of Strategic Management and Organization, and Director of Health Organization Studies at the University of Alberta School of Business. She is a co-investigator on the Knowledge Utilization and Policy Implementation (KUPI) research program discussed in this paper, and the Principal Investigator on a multi-year qualitative research program studying organizational change in Alberta's health care system. Her main research interests are in the areas of <i>knowledge making in science</i>, and <i>organizational change</i>, specifically how organizational change is implemented and sustained over time, and how cultural systems shape change.</p>	<p>Contact: Karen Golden-Biddle Health Organization Studies University of Alberta School of Business 3-23 Business Building Edmonton, Alberta T6G 2R6, Canada E-mail: karen.golden-biddle@ualberta.ca URL: http://www.bus.ualberta.ca/hos/</p>
<p><i>Trish REAY</i> is Assistant Professor, Department of Strategic Management and Organization at the University of Alberta School of Business, and a co-investigator on the Organizational Change in Health Care program of research. In addition to teaching both undergraduate and MBA students in the School of Business, she is also a core faculty member of SEARCH Alberta. This program is supported by the Alberta Heritage Foundation for Medical Research, and is a unique educational program for health care professionals who want to deliver services based on an Evidence Based Decision-Making approach.</p>	<p>Contact: Trish Reay Health Organization Studies University of Alberta School of Business 3-23 Business Building Edmonton, Alberta T6G 2R6, Canada E-mail: trish.reay@ualberta.ca URL: http://www.bus.ualberta.ca/hos/</p>
<p><i>Carole A. ESTABROOKS</i> is Professor, Faculty of Nursing, University of Alberta. She is the Principal Investigator on the CIHR funded Knowledge Utilization and Policy Implementation (KUPI) research program. As well, Dr. ESTABROOKS is Principal Investigator of the Knowledge Utilization Studies Program (KUSP), and Academic Co-Director of the national training <i>Centre for Knowledge Transfer</i>. She holds appointments as an Adjunct Scientist at the Institute for Clinical Evaluative Sciences (ICES), and a Research Affiliate at the Alberta Centre for Active Living, University of Alberta.</p>	<p>Contact: Carole A. Estabrooks 5-112 Clinical Sciences Building Faculty of Nursing University of Alberta Edmonton, AB T6G 2G3, Canada E-mail: carole.estabrooks@ualberta.ca URL: http://www.ualberta.ca/~kusp/</p>

Citation

Please cite this article as follows (and include paragraph numbers if necessary):

Thomson, Denise, Bzdel, Lana, Golden-Biddle, Karen, Reay, Trish & Estabrooks, Carole A. (2005, January). Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data [33 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6(1), Art. 29. Available at: <http://www.qualitative-research.net/fqs-texte/1-05/05-1-29-e.htm> [Date of access: Month Day, Year].