Archivist On Board: Contributions To The Research Team

*Charles K. Humphrey, Carole A. Estabrooks, Judy R. Norris,
Jane E. Smith & Kathryn L. Hesketh*

Key words:
archiving, team research, data preservation, data sharing

Abstract: In this article, we demonstrate the advantages of including an archivist as a member of the research team based on our experience with two multi-site, multi-method, primarily qualitative projects. The Principal Investigator, committed to the principles of data sharing and preservation, recruited a data archivist at the inception of the projects. Several issues arose that are not typically encountered in a research project: investigators needed to agree to the principles of data preservation and sharing—concepts that are not typically discussed a prior; the research ethics application and approval had to incorporate the conditions of preservation and sharing; and we needed a comprehensive plan for preservation that would ensure the creation of high-quality data products worthy of deposition. This comprehensive plan required that we identify the standards of archiving, incorporating within the data management plan an appropriate inventory list and a design for tagged fields and a corresponding Document Type Definition (DTD) used in the mark-up of textual data. A plan for creating access to the data for secondary analysis was also developed. The conditions of use, cataloguing records, and citation guide are all part of preparing the data for access. Finally, the challenges of this approach are summarized.

Table of Contents

1. Introduction

The practice of archiving research data is rarely viewed as a component of the research process even though social science archives for quantitative data have been in existence and promoting these practices since the 1960s (FIENBERG & MARTIN, 1985). This oversight begins with the training of students where the subject of data archiving is not even discussed, let alone presented as an integral part of research. If archiving research data remains largely obscure to quantitative researchers, these practices are an even greater mystery to qualitative researchers. A pilot study in the UK estimated that approximately 65-70% of data collected in the 30 years previous had been stored in personal files and another 20% of data had been destroyed, thus only 10-15% of data was archived for other researchers to use (CORTI, FOSTER & THOMPSON, 1995). As a small step to address this gap in both research practice and training, the

principal investigator for a major study in the field of nursing intentionally incorporated data archiving into her design. The strategy to operationalize this component within the overall research design was to include a data archivist on the project team to: 1) teach data archiving practices to members of the research team, and 2) to incorporate archiving milestones into the project process to ensure that data products are prepared for archival deposit at the conclusion of the project. [1]

Canada does not have a national data archive nor does it have an institution actively promoting the archiving of research data. Consequently, few researchers in Canada think about archiving data. At the University of Alberta, researchers in the health sciences are required to maintain their data in secure and confidential form for five years. After this time, many researchers retain their data, but only for their own use. Others, because they do not intend to conduct secondary analyses, destroy their data. Until recently, researchers including a qualitative data collection in their research were required to destroy their data after a period of seven years, a practice that is now discretionary. [2]

Although one major Canadian research-funding agency, the Social Sciences and Humanities Research Council of Canada (SSHRC), mandates data archiving in their policies, there is no enforcement of the guideline to ensure compliance. At the same, time the nature of funded research has changed. Projects have become large multi-disciplinary, multi-site studies producing enormous amounts of data, and requiring more research money per project to conduct. Combining qualitative and quantitative data in one research project is a growing trend in nursing as well as other social science research. [3]

The Canadian National Archives, which focuses on the preservation of Canadian heritage, does not house academic research data. As this paper is being written, the first major national effort in Canada to address the absence of a formal institution to archive data in this country is underway (see http://www.sshrc.ca/english/resnews/presdesk/october2000.htm [Broken link, FQS, Sept. 2002]). The SSHRC and National Archives have jointly convened a national investigation into the need for a national data archive as well as to recommend the implementation of such services. This consultation is expected to take one year to complete. [4]

In addition to the shortcomings of institutional support in Canada for the archiving of research data, the literature of various scientific disciplines is inconsistent in advancing the fundamental norm of data sharing. Previously, HILGARTNER and BRANDT-RAUF (1994) described at least four perspectives from which data sharing had been examined: 1) the ethics of data sharing and ownership, 2) scientific findings as communal property, 3) intellectual property rights, and 4) university-industry relations (p.356). While there is quite an extensive history and resulting literature addressing data sharing in the social sciences (see for example: FIENBERG & MARTIN, 1985; HEDRICK, 1988; SIEBER, 1991), in nursing, minimal attention has been paid to the issue of archiving data. Since ESTABROOKS and ROMYN's 1995 article, there has been little published in the

nursing literature on this topic. Moreover, this literature completely overlooks the sharing of qualitative data in Canada. Consequently, the fact that, to our knowledge, no formal qualitative research archive exists in Canada is no surprise. [5]

2. The Study

In this article, we relate the experiences of one Canadian research team that is archiving both quantitative and qualitative data. The Principal Investigator and the data archivist shared a common belief that researchers are stewards, rather than owners, of research data; and that publicly funded research belongs ultimately in the public domain. As a result, they undertook to incorporate data archiving activities from the study's inception. The study, which is currently underway, is examining the determinants of research use in both adult and pediatric nursing settings, using pain management as the context (see http://www.ualberta.ca/~kusp). The goals of this research are to develop strategies that will enhance the utilization of research by health professionals, and to make policy recommendations targeted at organizational levels. Data collection began in September of 1999 and will conclude in June 2001. The study involves two provinces, four tertiary-level hospitals and eight separate adult and pediatric nursing units. Data collected include observations, individual and group interviews, documents, and several quantitative measures. The data are captured in transcribed field notes, transcribed interviews, documents, and numerical data files. [6]

As part of the research process, our team is developing and adapting data preservation standards and procedures. We hope that what we have learned will encourage researchers in many disciplines to share and exchange data. As we disseminate our findings in papers and presentations, we intend to raise awareness of the benefits of data sharing. [7]

3. The Role of the Archivist

What is the archivist's role on a project? Primarily, and certainly in the beginning, the archivist's role is to educate the research team. Researchers need to see the wider use of data, especially the value of data for secondary research purposes. In our project, some members were genuinely horrified at the idea of archiving data:

- What?! These data are going to be available to others? Forever?
- Do we have to tell the participants? What will we tell them?
- What about ethics approval? How do we explain this in our request for ethical approval?
- How will we handle confidentiality?
- Will we lose control over the data—how will we know how it is being used?
- What?! Someone is going to see my fieldnotes? Isn't that an appropriation of my intellectual property? [8]

Members of our team became increasingly excited about archiving data once the possibilities for replication, creation of new data, and the potential for testing models across data collections became apparent. As researchers with a new understanding of the research potential of archived data, we believe that there are opportunities for archivists to advance the idea of archiving to other researchers. With more awareness, researchers will become interested in archiving and sharing data as part of the research process. We now see that archiving data will provide new possibilities for research in our specific area of interest: research utilization and evidence-based practice amongst nurses and other health care practitioners. As individuals studying the value of research itself, this idea is very gratifying to us. [9]

Establishing actual data archiving practices within a conventional research method is a further educational task of the archivist. Creating an awareness of the value of data archiving among the members of a research team is not enough. The implementation of practices that lead to archival products is also required. Often, many of the established methods employed during the data collection are driven by a sense of urgency. For example, the desires and pressures to publish are often so intense that shortcuts are taken in documenting the details of a data collection. In the long term, this oversight sacrifices the completeness of the research record. It is the archivist's role to ensure that the data are prepared with an eye toward long-term preservation. Archival practices will not necessarily delay the time required to complete research if archiving procedures are incorporated as part of the method from the start of the research project. [10]

By beginning a research project with the clear objective of producing an archival product, the researcher sets standards for her or his team that would otherwise have been overlooked. By asking, "how will someone else use these data?" at the various junctures of collecting data, the researcher is cognizant of the level of detailed documentation that must be prepared while the data are created. From an archiving perspective, the whole orientation of the researcher requires an assessment of the future of the data products. [11]

4. The Role of the Researcher

We believe that most researchers fail to archive their data simply because they have not had the proper training. For others, in the rush to publish, present and write reports for funding agencies, archiving may appear to be a frill that requires a lot of effort at the end of a project, with little pay-off for the researcher. In addition, some researchers may find it difficult to accept the idea of "their" research as public property. For example, a researcher might say:

> But I would need to write a grant in order to prepare the final products for the archivist. And I would need to revise my ethics proposal to reflect the new uses of the data. I think it would be easier just to drop this! [12]

The first responsibility of researchers is to protect the confidentiality of their participants. For most researchers, the easiest way to protect privacy is to destroy the data. This may be true if archiving arises as an afterthought. Archiving requirements necessitate proper documentation and care to ensure that confidentiality is maintained throughout the research process. Participants need to be informed of the intention to archive and share data with other researchers, as well as the steps that will be taken to ensure privacy. [13]

The steps to creating an archival product overlap a great deal with the process of creating a properly coded, well documented dataset for primary analysis. These steps include: 1) verifying accuracy and integrity of the data sets; 2) creating codebooks to allow for proper coding, and later interpretation of the coding; and 3) ensuring all ethical and procedural documentation is in order. In addition, archiving a data set involves determining a comprehensive data management plan including conditions of access to the data, as well as preparing sets of the data in forms that are universally readable by software (e.g., using ASCII format for numerical data, ensuring a standardized textual structure is identified by means of a universal tagging system such as XML for textual data). Creating an archival product is mostly a matter of following sound data management procedures from the outset of the project, rather than creating more work at the end of the process. [14]

If archiving is included in the planning process and continues throughout the project, the extra work requirements to preserve data are greatly reduced. Also, if the funds to archive data were included in grants, archiving would not be viewed as such a drain on existing research resources. Funding agencies need to mandate and enforce archiving activities as part of the research process. Not only do researchers need to be educated, but the major funding agencies also require training in the importance of archiving and sharing data. [15]

In summary, by becoming involved at the beginning of a research project, the data archivist has an opportunity to introduce standards for the preparation of data and documentation. One of the challenges facing archivists is the wide variety of formats in which data and documentation are prepared. Some of these practices actually hinder the sharing of files or the secondary use of data and documentation. Early involvement allows the archivist to promote standards that support access and usability, without creating more work for the researcher. If the researcher leaves the task of archiving to the end of the project when the motivation and resources to complete the task are waning, archiving is more likely to be overlooked. [16]

5. Conclusion

We hope that our experiences in this research project encourage other research-ers in our field to incorporate a similar strategy of data preservation in their studies. While there is excellent information available about how to archive data, examples of research projects where archiving is an integral component of the research planning and operations are missing from the literature. We hope that eventually our study serves as a prototype for other researchers, and to initiate a discussion in the literature about archiving strategies. The current gap existing in the research literature needs to be addressed. We have concluded that

- the archivist needs to be in at the beginning of the project, not the end;
- in Canada, we need to develop guidelines for archiving data; and
- we need to raise awareness about data preservation among the academic and funding communities. [17]

Furthermore, researchers need a greater awareness of the relationships that need to be built with the data archive community. One recent reaction from a researcher upon hearing about this project was that the authors were trying to convert all researchers into archivists, which sadly misses the point. Rather, we are promoting the collaboration between researchers and archivists and encouraging new partnerships in research projects. Toward this end, we are calling for the inclusion of the archivist as a member of the research team. [18]

In an era where research dollars are available, but with a renewed mandate for fiscal responsibility, a perfect opportunity exists for archivists to renew their efforts to educate researchers about why archiving is an important consideration in their projects. We believe that once health researchers realize the potential benefits of archiving data, they will be more receptive to building infrastructure into their research programs. We also recognize that we have considerable work to do before funding agencies are prepared to mandate the sharing of data. Such sharing requires the development and acceptance of international standards for archival products, and it requires that granting agencies permit grant budgets that include data preservation costs. [19]

Acknowledgments

References

Corti, L.; Foster, J. & Thompson, P. (1995). Archiving qualitative research data. *Social Research Update, 10*. [Online: http://www.soc.surrey.ac.uk/sru/SRU10.html].

Estabrooks, C.A. & Romyn, D.M. (1995). Data sharing in nursing research: advantages and challenges. *Canadian Journal of Nursing Research, 27*(1), 77-88.

Fienberg, S.E.; Martin, M.E. & Straf, M.L., National Research Council (U.S.) Committee on National Statistics., & National Research Council (U.S.). Commission on Behavioral and Social Sciences and Education (1985). *Sharing research data*. Washington, D.C.: National Academy Press.

Hedrick, T.E. (1988). Justifications for the sharing of social science data. *Law and Human Behavior, 12*(2), 163-171.

Hilgartner, S. & Brandt-Rauf, S.I. (1994). Data access, ownership, and control. *Knowledge: Creation, Diffusion, Utilization, 15*(4), 355-372.

Sieber, J.E. (1991). Openness in the social sciences: Sharing data. *Ethics & Behavior, 1*(2), 69-86.

Web Sites of Interest

Qualidata: http://www.essex.ac.uk/qualidata/

The Data Document Initiative: http://www.icpsr.umich.edu/DDI/

ICPSR Guide to Social Science Data Preparation and Archiving:
http://www.ifdo.org/archiving_distribution/datprep_archiving_cfr.htm [Broken link, FQS, August 2005]

Authors

Charles K. HUMPHREY is the head of research data services in the Library at the University of Alberta

Contact:

Charles K. Humphrey

1-01 Rutherford South
University of Alberta
Edmonton, Alberta T6G 2J4
Canada

Phone: 780-492-9216

E-mail: humphrey@datalib.library.ualberta.ca

Carole A. ESTABROOKS is an assistant professor in the faculty of nursing at the University of Alberta

Contact:

Carole A. Estabrooks

E-mail: Carole.Estabrooks@ualberta.ca
URL:
http://www.nursing.ualberta.ca/homepage.nsf/name/estabrookscarole

Judy R. NORRIS is an associate professor in the faculty of nursing at the University of Alberta

Contact:

Judy R. Norris

E-mail: Judy.Norris@ualberta.ca

Jane E. SMITH is a doctoral student in the faculty of nursing at the University of Alberta

Contact:

Jane E. Smith

E-mail: Jane.Smith@ualberta.ca

Kathryn L. HESKETH is the project coordinator for the Knowledge Utilization Studies in Practice Program (http://www.ualberta.ca/~kusp) at the University of Alberta

Contact:

Kathryn L. Hesketh

E-mail: katie.hesketh@ualberta.ca

Citation

Revised 2/2007