

---

# Research for Building a Better Data Community

I have come to believe that IASSIST members must seriously consider the value of conducting research about our own profession and field. If we do not initiate and value such research, the expectation that anyone else will undertake this task for us is unrealistic. I am not necessarily referring to “theory-driven” research, although this would be welcomed. Rather, I have in mind “issue-driven” research, that is, the kind of research that helps us understand the relationships, norms, and behaviours within our information and science cultures, including our own data sub-culture. For example, we need to conduct research on the issues behind data preservation and access. I am not talking about “how” to preserve data or provide access but instead investigating the norms and behaviours underlying the activities of data preservation and access.

A few recent events have led me to this new conviction. First, I was asked a couple of years ago by a major research council to review a grant application in which the principal investigators were proposing to study the economics of archiving data. In my enthusiastic endorsement of this application, I wrote that the principal investigators should expand their scope to explore the “data economy”, which I characterized as who gets access to which data, when and how. I thought that this would be a wonderful project for our profession. We would have research conducted about data archives and their role in the data economy. Here data archives would be the object of research rather than the sources of data for research. Unfortunately, the project was not funded. Nevertheless, it did stimulate my thinking about the possibilities of such research.

The next recent event took place at the 2000 IASSIST conference. I was excited by the research carried out by Karsten Boye Rasmussen and Repke de Vries about IASSIST as a virtual community. Their use of the IASSIST e-mail discussion list and the log files for the Web site and on-line issues of the IASSIST Quarterly clearly demonstrated ways of doing research about our organization and profession. A third and even more recent experience has been the research that I have been conducting in conjunction with the consultation underway in Canada about creating a national data archive.

*by Charles K. Humphrey\**

Before talking about specific research findings, let me briefly describe through an example one way in which “issue-driven” research might be performed in our field. One challenge we face, which has not changed over the last forty years, is how to get researchers to think about archiving their data at the beginning of a project rather than at the conclusion of

their research. And more than just thinking about archiving data earlier, how do we get researchers to conduct their projects so that their data products meet archival standards – rather than having to build an archival data product after the research has concluded. In other words, how do we mainstream data archiving in the research process? This is neither a new nor novel idea. However, is it an idea whose time has arrived? If our profession better understands the dynamics of current research practices that inhibit data archiving, can we be instrumental in bringing about the necessary changes to mainstream data archiving?

I am currently a co-investigator on a nationally funded project studying research utilization in the field of nursing. Specifically, we are studying how practitioners eventually use medical research findings about pain and pain control. How do research outcomes end up in practice? I introduced the principal investigator of this project to the value of data sharing and stewardship when she was a graduate student several years ago.<sup>1</sup> I also helped her preserve the research data from her dissertation. Subsequently, she has become a successful researcher who is a key proponent of data preservation and data sharing in Nursing research. This is how I came to be invited as a member of her team.

Out of this project, I was the lead author of a paper presented at the 2000 Conference on Social Science Methodology in Cologne, Germany entitled, “Archivist on board: contributions to the research team”.<sup>2</sup> This paper presented the role and value of an archivist on the research team. While I was not at the conference to present the paper, a colleague and co-author read it on my behalf. Ekkehard Mochmann was present at this session and came to the rescue of my colleague when one researcher, after hearing the presentation, protested that we were trying to turn researchers into archivists. My colleague reported that Ekkehard explained to the researcher that what we were proposing was to initiate partnerships between researchers

and archivists. When first hearing of this exchange, I was struck by the immediate reaction of the researcher. What were the attitudes and values behind such a response? Why were the attitudes of this researcher so seemingly different from those of the principal investigator with whom I work?

Turning to some findings from the research we are conducting in conjunction with the national data archive consultation, some insights into attitudinal differences underlying the principles of archiving data can be found. One of the four surveys that were administered was a sample of researchers who received a grant from the Canadian Social Sciences and Humanities Research Council between 1998 and 2000. One objective of this survey was to identify the number of researchers who produce data products as part of their research and to determine how many researchers have ever archived or intend to archive their data. A second objective was to investigate researchers' attitudes about data sharing and archiving.

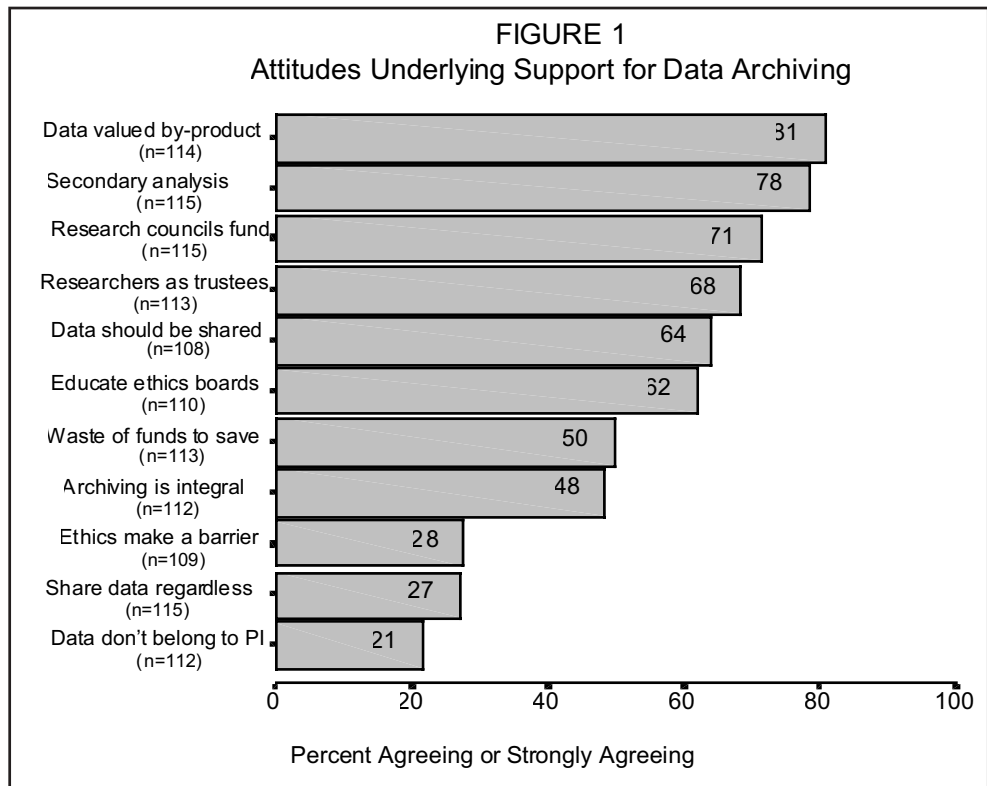
Eleven items were used to gauge these attitudes. These questions touch upon the legitimacy of secondary analysis as a research method, on the value of data as a by-product of research, on the issues of data ownership and data sharing, on research council funding to prepare data for sharing, and on the impact that ethics review boards have on data sharing (see List 1). Because the wording of five items (5a, 5d, 5e, 5h, 5j) does not support the principles of data sharing or archiving, the response categories for these items were recoded to correspond with the direction that supports the archiving principle.<sup>3</sup>

Figure 1 shows the combined percentage of respondents 'agreeing' or 'strongly agreeing' on each item. The items in this Figure have been arranged in decreasing order of support. Eighty-one percent agree that data should be a valued by-product of research, while only 21 percent agree that data do NOT belong to the researcher as her or his intellectual property. This decreasing order of agreement represents an increasing difficulty in support of sharing and preserving research data. Six steps of item-difficulty can be seen in this Figure. Eighty-one and 78 percent of the respondents

- LIST 1**
- Attitudinal Items Underlying Support for Data Archiving**
- 5a. Secondary data analysis is not a valid research method.
  - 5b. Data should be considered a valued by-product of .. research.
  - 5c. Data should be shared with other researchers, assuming it has been appropriately anonymized.
  - 5d. Data belong to the principal investigator as her or his intellectual property.
  - 5e. Data should only be shared if the principal investigator decides to share it.
  - 5f. Archiving data should be an integral part of conducting research.
  - 5g. Researchers who obtain information that cannot be easily reproduced from respondents are, to a degree, trustees of the data.
  - 5h. Spending resources to prepare the data from my research so that other researchers can use it would be a waste.
  - 5i. Research councils should include funds to cover the costs of preparing data for sharing.
  - 5j. Ethics review boards make it impossible to share confidential data on human subjects.
  - 5k. Ethics review boards need to be educated about the need to preserve data.

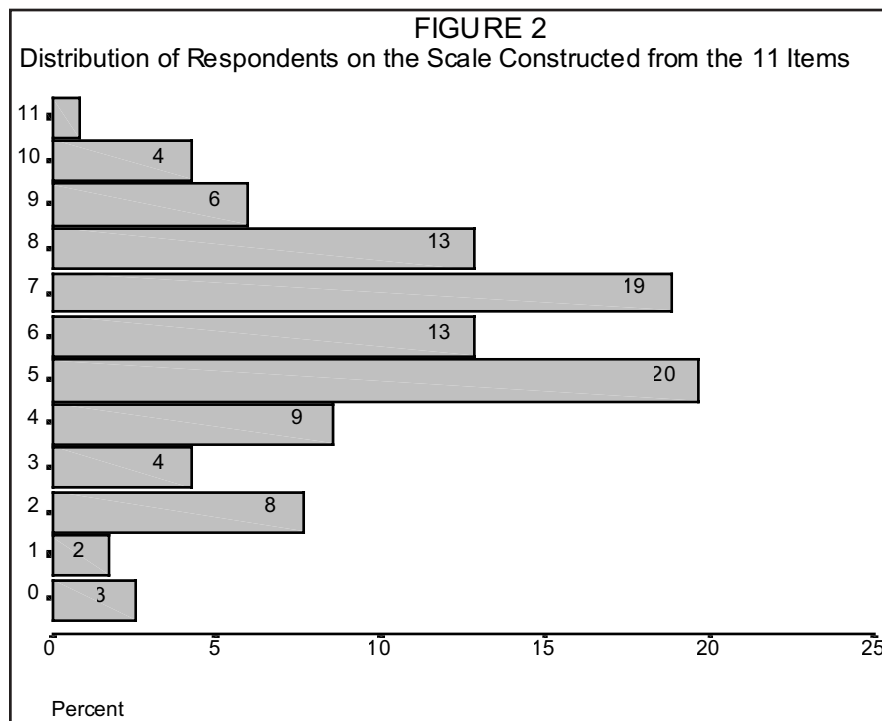
accept the first two items, data as a valued by-product and secondary analysis as a valid research method, respectively.

The second step consists of the items about research councils covering the costs to prepare data for sharing and



about researchers serving as trustees of data that cannot be easily reproduced from respondents. Seventy-one and 68 percent endorsed these items, respectively. The third step is made up of the item stating that data should be shared if it has been properly anonymized (64 percent) and the item that ethics review boards need to be educated about the need to preserve data (62 percent).

A slightly larger step occurs with the next two items. Fifty percent agree that spending resources to prepare the data from their research would not be a waste and 48 percent agree with the statement that archiving data should be an



integral part of conducting research. An even larger drop occurs with the fifth step. Twenty-eight percent disagree that ethics review boards make it impossible to share confidential data on human subjects, while 27 percent disagree that data should only be shared if the principal investigator decides to share it. As mentioned above, the smallest percentage of agreement (21 percent) is that data do NOT belong to the principal investigator as her or his intellectual property.

A scale was constructed based on the total number of items on which each respondent agreed with the principles of data archiving (see Figure 2). A score of zero indicates that the respondent did not support any of the items endorsing data archiving, whereas a score of 11 represents someone who supported all of the items. Seventeen percent were low supporters of data archiving (those with scores from

zero to three), while 24 percent were high supporters (those with scores from eight to 11). Fifty-nine percent are in the middle. The correlation between this scale and a question asking how important it is for Canada to establish national services for the preservation of research data is 0.50, which is corroborative evidence that this scale measures some aspect of support for data archiving.

What do we make of these findings in light of the question asked earlier about

how to mainstream data archiving in the research process?

First, only around a quarter of Canadian researchers in this study appear to be strong advocates of archiving data. While only 17 percent seem to be protagonists, close to 60 percent are in the ambivalent middle. Apparently, this rather substantial group requires further education on the principles of archiving.

Secondly, one of the eleven items rather succinctly summarizes the notion of mainstreaming data archiving. This is the item that states, "Archiving data should be an integral part of conducting research." Looking at the results of this item, 12 percent agreed strongly, 36 percent agreed, 30 percent were unsure, 18 percent disagreed, and 4 percent disagreed strongly. The 30 percent who are unsure is as alarming in this distribution as the 22 percent who disagree. One concern raised by this finding is that at least half of the researchers do not view data archiving as part of the normal practices of conducting research. The concept of archiving may be generally understood,

but archiving as part of the research process has not become routine.

An explanation for these results may be directed at incomplete training of researchers in their graduate school years or at senior researchers who are not mentoring junior researchers about data archiving. Another possible explanation is the failure of our profession to promote archiving as part of the research process. If the importance of the practice is not taught as part of the research method, data archiving will not be discussed or perceived as a generally important activity.

Turning away from this specific example, I would like to conclude with a couple of observations about changing our thoughts in IASSIST toward research. First, our organiza-

tion is well positioned to conduct comparative, cross-national research. We are an international organization with opportunities to investigate the generalizability of national findings. Are the attitudes described above held only by Canadian researchers, or are these attitudes commonly found among researchers across societies? Will we discover underlying attitudes about archiving data that are held by researchers around the globe? These are challenges that we can undertake together. Secondly, we live in a world increasingly calling for evidence-based decision-making. We should approach the issues we face in our profession by building evidence through research. Not only will we be stewards of data, but we will be contributing to the knowledge about the research process.

### Footnotes

<sup>1</sup> While completing her doctorate, she and a fellow graduate student published the following article about data sharing: Estabrooks, C.A. and Romyn, D.M. (1995). Data sharing in nursing research: advantages and challenges. In *Canadian Journal of Nursing Research*, 27(1):77-88.

<sup>2</sup> Humphrey, C.K., Estabrooks, C.A., Norris, J.R., Smith, J.E., Hesketh, K.L. (2000). Archivist on board: contributions to the research team. In *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [Online Journal], 1(3).

<sup>3</sup> “Strongly agree” and “agree” are the responses supportive of these principles.

\* Paper presented at the IASSIST/IFDO Conference 2001, Amsterdam Charles K. Humphrey May 2001.