SYMPOSIUM

Qualitative Data Archiving: Rewards and Challenges

Colin Elman, *Syracuse University*
Diana Kapiszewski, *University of California, Irvine*
Lorena Vinuela, *Syracuse University*

Political science has witnessed a renaissance in qualitative research methods (Bennett and Elman 2006). Over the last 15 years, the canon has been reworked to systematize and expand the repertoire of qualitative methods, ground them more firmly in contemporary philosophy of science, and illuminate their strengths relative to quantitative and formal methods (Bennett and Elman 2007). A rapidly expanding body of political science research now employs qualitative and multimethod analysis, and institutions dedicated to qualitative and multi-method research have flourished.

New scholarship addressing field and archival research methods is an increasingly important facet of this renaissance. Books (e.g., Feldman, Bell, and Berger 2003; Magnette 2008; Trachtenberg 2006; Schatz 2009), journal symposia (e.g., Leech 2002; Lieberman, Howard, and Lynch 2004; Loaeza, Stevenson, and Moehler 2005; Clark 2006; Read, MacLean, and Cammett 2006; Ortbals and Rincker 2009), and individual journal or reference pieces (e.g., Harrison 1992; Collier 1996; Thies 2002; Wood 2006, 2007; Tansey 2007; Read, forthcoming) on field and archival research are being published with increasing frequency. Further, the Institute for Qualitative and Multi-Method Research (IQMR) has taught field and archival research methods each year since its inception in 2002, and a short course on conducting fieldwork has been offered at the annual meetings of the American Political Science Association (APSA) since 2001.

In tandem with advances in qualitative analysis and an increased focus on field research, an interest in data sharing has been growing among qualitative researchers. These dynamics have highlighted a significant infrastructure gap. In contrast with norms and practices associated with quantitative research, as well as with qualitative work in some other countries, there is currently no dedicated venue or agreed set of practices for storing, sharing, and reusing qualitative social science data in the United States (Heaton 2004, 6). As a result, most qualitative data generated by American social scientists are used only once. Scholars wishing to retrace the steps taken by others typically must repeat the research. To the limited extent that researchers share qualitative data, they generally do so via ad hoc arrangements that are neither efficient nor cost-effective.[1]

This brief article argues for a generalized norm of sharing qualitative data, and for the establishment of dedicated structures and practices for doing so. Increased sharing will ease replication and make research processes more transparent. This will assist the evaluation of research and encourage methodological rigor. Archiving qualitative data will promote secondary data analysis, allowing scholars to engage in comparative research that to date has been prohibitively time consuming or expensive. Finally, data archiving will provide valuable pedagogical tools and, by increasing researcher visibility, promote the formation of epistemic communities and research partnerships.

Of course, qualitative data archiving is not without its challenges. Emphasizing data sharing may interfere with scholars' ability to achieve their research goals. In addition, depending on the conditions of collection, not all data are shareable. Finally, for sharing and secondary analysis of qualitative data to become accepted, disciplinary norms will have to undergo significant change. As we note below, however, each of these challenges can be addressed. We believe that the potential rewards of qualitative data archiving will more than compensate for the effort that addressing its difficulties will require.

THE POTENTIAL OF QUALITATIVE DATA ARCHIVING

Qualitative data archiving empowers scholars to store, access, and download electronic qualitative data of all types. These include, for example, interview tapes, text files (transcripts from interviews, focus groups, and oral histories; case notes; meeting minutes; research diaries), scans of newspaper articles, images of official documents, and photographic, audio, and video materials. Registered users of the contemplated repository will be able to find and access potentially useful qualitative data sets through searching metadata elements as well as content.[2] User-access controls will regulate who is able to search and retrieve the data sets housed in the repository, with more sensitive materials being restricted. The "public face" of such a repository is likely to be a Web-based portal that makes visible portions of the metadata associated with the projects in the repository (within the bounds of confidentiality and human-subjects concerns).

Qualitative data archiving will have several crucial benefits. At present, if a qualitative scholar wishes to check the findings of another, they typically have to do research on the same topic. By making data available and increasing the transparency and visibility of research processes, qualitative data archiving can dramatically reduce the costs of assessing empirically based analysis through replication (Swan and Brown 2008, 7). This is likely to be especially significant for multimethod studies (see e.g., Lieberman 2005), where the unavailability of the data underpinning the qualitative portion of the research has complicated evaluation and replication,

doi:10.1017/S1049096509990771

https://doi.org/10.1017/S104909651099077X Published online by Cambridge University Press

rendering the qualitative component the presumed weaker dimension of the research.

Second, qualitative data archiving will allow for the vertical integration of primary data, secondary analysis, and scholarly output.[3] This integration will generate transparency—researchers will be able to show more clearly how they arrived at their inferences and interpretations—with multiple positive effects on practice, pedagogy, and collaboration. In terms of practice, increased visibility will lead to increased rigor. Researchers who anticipate transparency will be motivated to carry out data collection and analysis in a systematic, replicable way. Having access to clear descriptions of how primary research and secondary data analysis were carried out will also allow scholars to learn from others' experiences and avoid reproducing mistakes—as well as to discuss, critique, and improve qualitative methods more readily. Data archiving will also provide valuable pedagogical and coordination tools. Scholars teaching qualitative methods courses can ask their students to identify and discuss a range of empirical research techniques employed in studies housed in such repositories, and practice the analytic techniques they are learning on "real" empirical data. Also, the publication of data will vastly increase the visibility of scholars working on particular topics. This greater visibility, in addition to the tools for collaborative research that such repositories can provide, will facilitate team research and the formation of epistemic communities around research areas and questions (Swan and Brown 2008, 26).

Finally, archiving qualitative data will allow data to accumulate, and will facilitate comparative research. Archiving qualitative data reduces the risks of data loss that personal storage poses, and minimizes the transaction costs associated with data sharing. Usually, qualitative data are researcher-centric, gathered in connection with a specific inquiry, and used just once (Swan and Brown 2008, 7).[4] Depositing data in a repository and making them available to other analysts will encourage efficiency by preempting the unnecessary repetition of research. Further, a repository will allow scholars to undertake research in the context of a much larger universe of available data. Hence, qualitative data archiving will help individual scholars make comparisons of many types—across space (countries, cities, etc.), across time, across policy areas, across groups, and so on—that could otherwise require additional research resources or assembling a research team (Corti 2000a, 6.2).

Data archiving raises the interesting and seemingly disconcerting possibility that qualitative research could be undertaken without the scholar directly engaging with *any* of the cases from which the data under analysis were drawn. While this is entirely unexceptional for quantitative scholars (who routinely start their analysis with another scholar's data set), it is a departure for qualitative scholars for whom close engagement with cases is typical, and perhaps obligatory. We do not anticipate that the introduction of qualitative social science data archiving in the United States will lead to either an immediate or eventual disconnect from cases for qualitative researchers. Nevertheless, beginning to archive qualitative data will doubtless provoke a spirited, healthy (and perhaps overdue) debate on the potential value of and risks involved in secondary data analysis.

ADDRESSING THE CHALLENGES OF QUALITATIVE DATA ARCHIVING

Qualitative data sharing presents a range of challenges that differ both in kind and degree from those involved in quantitative data archiving. Qualitative data are more heterogeneous than are quantitative data, their collection and sharing raise more human-subjects concerns, and they pose different metadata and indexing challenges (Heaton 2004, 29; Parry and Mauthner 2004, 144). In this section, we discuss some concerns regarding the effects that qualitative data archiving might have on data-collection practices, the "shareability" of certain types of information, and the evolution of disciplinary norms. We also highlight some possible solutions.

Increased sharing will ease replication and make processes more transparent. This will assist the evaluation of research and encourage methodological rigor. Archiving qualitative data will promote secondary data analysis, allowing scholars to engage in comparative research that to date has been prohibitively time consuming or expensive. Finally, data archiving will provide valuable pedagogical tools and, by increasing researcher visibility, promote the formation of epistemic communities and research partnerships.

A first set of issues concerns data-collection practices. As we noted above, by making research procedures transparent, qualitative data archiving can encourage scholars to be more self-conscious and rigorous in how they undertake data collection. However, this transparency may also have the unintended consequence of encouraging researchers to engage in self-censorship, omitting data collected using a technique they fear will not be considered rigorous, or failing to record relevant observations that they worry may reflect badly upon themselves or a subject.[5] A related problem is that if "shareability" becomes an overriding virtue by which data are judged, scholars may be influenced to focus solely on contexts where they can collect data that can be easily deposited. Important information that could have been collected and used with discretion will go unsolicited. The possibility of unintended collateral outcomes such as these highlights the importance of a sustained disciplinary conversation about the merits and limitations of data sharing.

https://doi.org/10.1017/S104909651099077X Published online by Cambridge University Press

Another series of concerns relates to limits on sharing qualitative data. Much field research involves interaction with human subjects—interviews and focus groups, oral histories, and ethnographic studies. Institutional review boards (IRBs) are likely to require that researchers who wish to store and share via the repository information gained through these interactions solicit subjects' permission (Mauthner, Parry, and Backett-Milburn 1998, 743; Heaton 2004, 79).[6] Inevitably, some subjects will be unable or unwilling to have their identity revealed and/or information they provide made available to other scholars (Corti, Foster, and Thompson 1995, 3). Further, interviews are sometimes given off-the-record and for background purposes only, and subjects are sometimes promised anonymity (Heaton 2004, 81; Parry and Mauthner 2004, 146). This conditional provision of data may also affect the form data take: unproblematic conversations may appear as audio and/or video recordings, whereas more difficult exchanges may appear as edited formal transcripts, redacted summaries, contemporaneous notes, or near contemporaneous recollections that may be less faithful renditions of scholar-respondent exchanges. In short, the conditions under which subjects offer information may require a resource-intensive process of contextualizing data (or the storage of partial or incomplete accounts and data sets), and can affect the type and form of data scholars collect. These outcomes can have significant consequences for analysis, interpretation, and inference.

Without minimizing these concerns, it is important to note that a range of measures can be taken to protect human subjects while sharing data collected from and about them. For example, washing qualitative data can help preserve anonymity, and information thereby removed can be explicitly noted and the impact of its deletion assessed. Differential user access can also help to address human-subjects concerns.[7] In sum, while protecting human subjects and conforming fully to standards set by IRBs may prevent some data from being shared, these concerns do not stand in the way of making a great deal of material available to a broad range of users.

Non-interactive forms of data collection, such as archival research, may also produce data that cannot be posted to qualitative data repositories, or that can only be posted with some time lag, or in attenuated form. For example, documents accessed in archives may have copyright restrictions that prevent them from being reproduced or made publicly available for secondary reproduction. Further, the legal frameworks guiding such restrictions differ by context, and may change over time (Heaton 2004, 83–85). While these challenges require specialized solutions, the sustained success of institutions such as the National Security Archive suggest that, with careful attention, they can be addressed.

In sum, while we recognize that not all data collected via qualitative and multi-method research can be deposited, a considerable amount likely can. Sharing a significant subset of the qualitative data that social science researchers collect and produce is vastly preferable to the status quo, in which practically no qualitative data are publicly available in the United States.

Finally, as anticipated above, disciplinary norms will need to change if sharing and reusing qualitative data—and producing scholarship relying (wholly or partly) on secondary data analysis—are to become accepted practices. There are several ways in which disciplinary norms can be influenced. For instance, the discipline's professional organization (APSA) could clarify the guidance it offers on data release and sharing for qualitative research.[8] Likewise, funding agencies could advise grantees to make qualitative data available,[9] and more journals could require authors to do so.[10] Training opportunities, outreach, and guidance on qualitative data sharing (see Corti 2000a, 6.3) could also be offered through IQMR or the short courses that are taught at the annual APSA meetings. In addition, professional associations and academic departments could consider revisiting career-based incentives for scholars such that, for instance, the publication of a qualitative data set receives due credit.

With respect to persuading qualitative researchers to deposit and use shared data, the most effective solution is likely establishing a dedicated qualitative repository, rather than building on a pre-existing quantitative archive. A specialized repository will be more attuned to the particular challenges posed by archiving qualitative data and will have a larger demonstration effect. A dedicated repository can also be epistemically neutral. While a considerable amount of qualitative and multi-method research is situated in epistemes that value replication, this is by no means a universal value. A repository viewed broadly as a means of increasing the transparency of the evidentiary basis for interpretive, descriptive, or inferential arguments based on qualitative data will be better positioned to attract all types of qualitative researchers. Finally, while a dedicated repository will focus on developing expertise in storing, searching, and sharing qualitative data, it will not be a "silo." Rather, it will be designed to be visible to, and open to communication and interaction with, a wide audience, as well as with existing archives.

EXPERIENCES IN QUALITATIVE DATA SHARING

Efforts to increase the use of digital platforms to facilitate sharing qualitative data and encourage collaboration among researchers have had positive results elsewhere in the world. For instance, recognizing that qualitative data pose challenges that cannot be adequately addressed by archives designed to hold numerical data sets, and aiming to promote greater transparency and reduce duplication of research, funding agencies in many OECD countries initiated the push to close the infrastructure gap in qualitative research beginning in the 1990s. Several funding organizations adopted a mandatory sharing policy for grant holders,[11] and the resulting repositories receive regular deposits on a national scale and hold a wide range of qualitative materials. While there is variation in the degree of development across these institutions, their archives play an important role in training and support for researchers and students throughout the process of data deposit and reanalysis.[12]

The Qualitative Data Archival Resource Center (QUALIDATA), associated with the University of Essex, UK, is at the forefront of innovation. QUALIDATA was created in 1993, and is funded by the Economic and Social Science Research Council (ESRC)—the United Kingdom's largest sponsor of social

https://doi.org/10.1017/S104909651099077X Published online by Cambridge University Press

science research. QUALIDATA was created to save "endangered" qualitative social research that was held, at the time, in the National Social Policy and Social Change Archive (NSPSCA). In the process of digitizing the NSPSCA collection and building its electronic infrastructure, QUALIDATA established procedures for cataloguing and extracting metadata, depositing data, ensuring respondent confidentiality and abiding by agreements made at the time data were collected, and limiting user access (Corti 2000b). QUALIDATA has served as a model for other European countries, and its development offers important insights for the introduction of a repository in the United States.

CONCLUSION

Qualitative research makes vital contributions to political science. One common thread uniting qualitative research of all types is close engagement with subjects and cases, and data collection initiated at extremely close range. This approach produces thick, rich, and heterogeneous data that are used to advance an array of analytical goals including interpretation, descriptive generalization, and causal inference. Yet to date, there is no venue in the United States to store and share hard-won qualitative data. This infrastructure gap—which transcends scholarly differences over the contributions of qualitative research—contrasts sharply with well-established norms in quantitative research, and with the data-sharing practices of qualitative social scientists in other countries.

Qualitative data archiving holds the key to making qualitative and multi-method research more transparent and more replicable, allowing scholars to investigate the evidentiary basis for previous interpretive and inferential claims based on qualitative data, and thus to assess research more easily. Moreover, sharing allows data to be used as a basis for further research, and encourages scholars to engage in secondary data analysis, opening up a range of new research possibilities including cross-temporal and cross-context comparison. Of course, as occurs whenever new practices may be adopted, the challenges and risks of sharing and reusing qualitative data must be considered. A qualitative data repository may influence data-collection practices in unproductive ways. Further, there may be important limits on the sharing of certain types of data, and full use of the repository may require significant changes in longstanding disciplinary norms. Nonetheless, in our view, these challenges are relatively minor in comparison to the tremendous utility that sharing and reusing qualitative data will provide. ■

NOTES

This article draws extensively from a multi-researcher project in progress. The authors thank those who participated in the Planning Workshop on a Qualitative Data Repository, held March 28–29, 2009, at Syracuse University (funded by NSF Grant SES 0838716). We especially thank Michele Lamont, Howard Turtle, and Lisa Wedeen.

1. To be sure, several university libraries and research institutions have archives for data collected by their affiliated researchers and the facilities to store digitalized text and audio material. Nevertheless, in the American academy the overwhelming focus is on archiving quantitative data, or on quantitative redactions of qualitative data.
2. *Metadata* are data about data. Metadata elements describe an information resource without, necessarily, being part of the resource.
3. Andrew Moravcsik's (2010) essay in this symposium argues for an "active footnoting" mechanism, with hyperlinked primary source material. We anticipate that the repository will have the potential to serve an important role in this process, mediating between scholarly references and hyperlinked sources.
4. In fact, quite often researchers only use a sub-set of the data they bring home from the field, discarding some data as they refine their research questions through the writing process.
5. To offer one hypothetical example, a scholar may balk at publicly noting that a respondent was hostile or condescending, even though that characterization may be highly relevant for interpreting the information the respondent provided, and may even be a relevant datum itself for some kinds of research questions.
6. Consequently, data that were collected before the development of a repository (legacy data) pose particular challenges for data archiving, because of the need to obtain retrospective permissions.
7. For example, some data may be made available for online use by any registered scholar while other data may be kept in non-networked storage to be accessed only in person at the repository with the depositor's (and if need be, the respondent's) permission and in accordance with explicit data-sharing agreements (see Wolf et al. 2006, 10, 13).
8. The American Political Science Association's *A Guide to Professional Ethics in Political Science* (2008) envisions non-release as the default—except when funder mandates or challenges to findings trigger release. Although the guide establishes a general heuristic requirement to disclose non-confidential sources for replication and testing, it does not specify whether "sources" refers to the identity of interviewees or to data.
9. The MacArthur Foundation, for example, "encourages openness in research and freedom of access to underlying data by persons with a serious interest in the research" (2008, under "Intellectual Property").
10. Most political science journals that have data-release policies either explicitly or implicitly limit those mandates to statistical data.
11. In most cases, qualitative data repositories were developed to store research outputs from projects financed by the respective national research funding agency.
12. Some examples include WISDOM in Austria, FSD in Finland, GESIS in Germany, ADPSS Sociodata in Italy, DANS in the Netherlands, NSD in Norway, SND in Sweden, and UKDA in Great Britain.

REFERENCES

American Political Science Association. 2008. *A Guide to Professional Ethics in Political Science*. 2nd ed. Washington, D.C.: American Political Science Association.

Bennett, Andrew, and Colin Elman. 2006. "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science* 9 (1): 455–76.

———. 2007. "Case Study Methods in the Study of International Relations." In *Oxford Handbook of International Relations*, ed. Christian Reus-Smit and Duncan Snidal. New York: Oxford University Press.

Clark, Janine A., ed. 2006. "The Methodologies of Field Research in the Middle East." *PS: Political Science and Politics* 39 (July): 417–41.

Collier, David. 1996. "Data, Field Work, and Extracting New Ideas at Close Range." *APSA-CP Newsletter* 7 (Winter): 1–6.

Corti, Louise. 2000a. "Progress and Problems of Preserving and Providing Access to Qualitative Data for Social Research—The International Picture of an Emerging Culture." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 1 (December). http://www.qualitative-research.net/index.php/fqs/article/view/1019.

———. 2000b. "Qualitative Archives: Short Descriptions: ESRC Qualitative Data Archival Resource Centre (Qualidata), University of Essex, UK." *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 1 (December). http://www.qualitative-research.net/index.php/fqs/article/view/1019.

Corti, Louise, Janet Foster, and Paul Thompson. 1995. "Archiving Qualitative Research Data." *Social Research Update* 10 (Autumn). http://sru.soc.surrey.ac.uk/SRU10.html.

Feldman, Martha S., Jeannine Bell, and Michele T. Berger, eds. 2003. *Gaining Access: A Practical and Theoretical Guide for Qualitative Researchers*. Walnut Creek: AltaMira Press.

Harrison, Hope M. 1992. "Inside the SED Archives: A Researcher's Diary." *Cold War International History Project Bulletin* 2 (Fall): 20, 28–32.

https://doi.org/10.1017/S104909651099077X Published online by Cambridge University Press

Heaton, Janet. 2004. *Reworking Qualitative Data*. London: Sage Publications.

Leech, Beth L., ed. 2002. "Symposium: Interview Methods in Political Science." *Political Science and Politics* 35 (4): 663–88.

Lieberman, Evan. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (August): 435–52.

Lieberman, Evan, Marc Howard, and Julia Lynch. 2004. "Symposium: Field Research." *Qualitative Methods* 2 (Spring): 2–15.

Loaeza, Soledad, Randy Stevenson, and Devra C. Moehler. 2005. "Symposium: Should Everyone Do Fieldwork?" *APSA-CP* 16 (2): 8–18.

MacArthur Foundation. 2008. "Intellectual Property Arising out of the Use of Foundation Funds." http://www.macfound.org/site/c.lkLXJ8MQKrH/b.4804425/k.6FE5/Intellectual_Property.htm.

Magnette, Paul. 2008. *Elite and Specialized Interviewing*. Essex: European Consortium for Political Research Press.

Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. "The Data Are Out There, or Are They? Implications for Archiving and Revisiting Qualitative Data." *Sociology* 32 (November): 733–45.

Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research." *PS: Political Science and Politics* 43 (1): 29–35.

Ortbals, Candice D., and Meg. E. Rincker, eds. 2009. "Symposium: Fieldwork, Identities, and Intersectionality: Negotiating Gender, Race, Class, Religion, Nationality, and Age in the Research Field Abroad." *PS: Political Science and Politics* 42 (2): 287–328.

Parry, Odette, and Natasha S. Mauthner. 2004. "Whose Data Are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data." *Sociology* 38 (February): 139–52.

Read, Benjamin. Forthcoming. "More than an Interview, Less than Sedaka: Studying Subtle and Hidden Politics with Site-Intensive Methods." In *Chinese Politics: New Sources, Methods, and Field Strategies*, ed. Allen Carlson, Mary Gallagher, Kenneth Lieberthal, and Melanie Manion. New York: Cambridge University Press.

Read, Benjamin, Lauren MacLean, and Melani Cammett. 2006. "Symposium: Field Research: How Rich? How Thick? How Participatory?" *Qualitative Methods* 4 (Fall): 9–18.

Schatz, Edward, ed. 2009. *Political Ethnography: What Immersion Contributes to the Study of Power*. Chicago: University of Chicago Press.

Swan, Alma, and Sheridan Brown. 2008. "To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs." A report commissioned by the Research Information Network. School of Electronics & Computer Science, University of Southampton. http://www.rin.ac.uk/files/Data%20publication%20report,%20main%20-%20final.pdf.

Tansey, Oisín. 2007. "Process Tracing and Elite Interviewing: A Case for Non-Probability Sampling." *PS: Political Science and Politics* 40 (October): 765–72.

Thies, Cameron. 2002. "A Pragmatic Guide to Qualitative Historical Analysis in the Study of International Relations." *International Studies Perspectives* 3 (November): 351–72.

Trachtenberg, Marc. 2006. *The Craft of International History: A Guide to Method*. Princeton: Princeton University Press.

Wolf, Virginia A., Joan E. Sieber, Philip M. Steel, and Alvan O. Zarate. 2006. "Meeting the Challenge When Data Sharing Is Required." *IRB: Ethics and Human Research* 28 (March-April): 10–15.

Wood, Elisabeth. 2006. "The Ethical Challenges of Field Research in Conflict Zones." *Qualitative Sociology* 29 (September): 372–86.

———. 2007. "Field Methods." In *The Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan Stokes. Oxford: Oxford University Press, 123–46.

https://doi.org/10.1017/S104909651099077X Published online by Cambridge University Press