Working Group or purport to represent the views of any of its other members. The DA-RT Working Group's recommendations were reviewed by the APSA Governing Council in September 2011, and passed to the Ethics, Rights, and Freedom Committee. The text was returned to Council in April 2012, and the language adopted as APSA policy.

[3] These terms were originally introduced into the current discussions in the Lupia and Elman memorandum referenced in note 2 above.

References

Lynn, Jonathan and Antony Jay. 1984. *The Complete Yes Minister: The Diaries of a Cabinet Minister, by the Right Hon. James Hacker MP*. Revised omnibus edition, London: British Broadcasting Corporation.

Wedeen, Lisa. 2010. "Reflections on Ethnographic Work in Political Science." *Annual Review of Political Science* 13, 255–272.

Closing the Infrastructure Gap: Qualitative Data Archiving

Diana Kapiszewski
University of California, Irvine
dianakap@uci.edu

Over the last 15 years, political science has witnessed a renaissance in qualitative research methods (see, e.g., Brady and Collier, eds., 2010).[1] The canon has been reworked, new areas of scholarship have appeared, and a rapidly expanding body of political science research now employs qualitative and multimethod analysis. Correspondingly, as noted in this symposium's introductory essay on the openness dialogue, although qualitative researchers have begun to explore ways to share their data and access those of other scholars, the lack of a dedicated venue or consensual set of practices for storing, sharing, and reusing qualitative social science data in the United States (Heaton 2004: 6) presents a significant obstacle. 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 practices of qualitative social scientists in other countries.[2] As a result of the lack of an appropriate data-sharing venue, the few American social scientists who *do* share their qualitative data generally do so via inefficient ad hoc arrangements.[3]

This brief essay argues for the development of generalized norms and specific practices for archiving and sharing qualitative data, and discusses an ongoing initiative to create a dedicated qualitative data repository. Archiving and sharing qualitative data will ease evaluation and replication of research, render research processes more transparent, and encourage secondary data analysis. Doing so will also provide valuable pedagogical tools and, by increasing researcher visibility, promote the formation of epistemic communities and research partnerships. Of course, not all qualitative data are shareable, and the establishment of norms of sharing could have unintended

and sometimes negative consequences. Nevertheless, the potential rewards of qualitative data archiving arguably compensate for the efforts required to address its difficulties.

The Promise of Qualitative Data Archiving

Qualitative data archiving enables scholars to store, search, access, and download electronic qualitative data of all types, from official documents, to interview transcripts, to photographic, audio, and video materials. Archiving qualitative data can produce several important benefits.

First, qualitative data archiving will allow for the vertical integration of primary data, secondary analysis, and scholarly output, allowing scholars to provide access to the data they used to arrive at their inferences and interpretations, and thus better demonstrate how they developed them. This transparency will encourage researchers to carry out data collection and analysis in a systematic, replicable way. It will also allow scholars to learn from others' experiences, help them to avoid reproducing mistakes, and facilitate discussion and critique of qualitative methods. Second and relatedly, by making data available and increasing the transparency and visibility of research processes, qualitative data archiving can dramatically reduce the costs of assessing and replicating empirically based qualitative analysis (Swan and Brown 2008: 7).

For instance, qualitative data archiving will facilitate instantiation of the "active citation" standard advocated by Andrew Moravcsik in his contribution to this symposium and elsewhere (e.g., 2010) by mediating between scholarly references and hyperlinked sources. This specific type of data archiving will allow scholars to make timely comments on and corrections to other scholars' use of primary sources. Consider, for example, the erroneous citation of a document as diagnostic evidence in the context of a process tracing narrative. Under the present state of affairs, the mistake would likely go unnoticed absent a subsequent publication on a closely related topic. The primary document's posting to a qualitative data archive would permit more immediate feedback at much lower transaction costs.

Third, archiving qualitative data will provide valuable pedagogical and coordination tools. Students taking qualitative methods courses will be able to learn from and critique the data-collection techniques used by scholars who archived their data, better understand the analytic strategies such scholars used in their published work, 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, facilitating team research and the formation of epistemic communities around research areas and questions (Swan and Brown 2008: 26). Finally, archiving qualitative data will facilitate data accumulation, allowing scholars to undertake research in the context of a much larger universe of available data, and to make comparisons across space, time, policy areas, groups, and so on that could otherwise require additional research resources or assembling a research team (Corti 2000: 6.2).

One concern might be that expanding the practice of qualitative data archiving could weaken the current norm (or even

obligation) among qualitative researchers of close engagement with cases. At the limit, one could imagine qualitative research being undertaken without scholars directly engaging with *any* of the cases from which the data under analysis were drawn. Yet it seems very unlikely that qualitative researchers would become so disconnected from their cases either in the short- or long-term, not least because doing so would undermine one of the primary comparative advantages of qualitative research.

Given the potential benefits of archiving qualitative data, several scholars have initiated a project to develop a dedicated digital qualitative data repository.[4] The repository will be accessible via a website portal, with user-access controls to regulate how scholars search and retrieve archived data. This archive will of course be just one place qualitative researchers might store their digital data. They can also store them in general archives like Dataverse and ICPSR; in university repositories such as those offered by Cornell, Duke, the University of California, and many others; in archives focused on particular issue areas; and on scholars' own personal websites. The main goals of the dedicated qualitative data repository currently under construction—which will aim to link with these other venues—are to demonstrate the practical possibility and intellectual promise of sharing qualitative data broadly, encourage their sharing, and serve as a site around which best practices can begin to be developed.

Strategies for Addressing the Challenges of Qualitative Data Archiving

Qualitative data sharing presents a range of challenges— and requires a set of solutions—that differ in some ways from those associated with quantitative data archiving. A first set of issues concerns data-collection practices. Making research procedures more transparent may have the unintended consequence of encouraging researchers to engage in self-censorship, for instance, omitting from their analyses data collected using a technique they fear will not be considered rigorous. A related problem is that if the "shareability" of the underlying data becomes an important criterion for judging empirical qualitative research, scholars may focus solely on contexts where they can collect data that can be shared easily. Important information that could have been collected and used with discretion will go unsolicited, and important topics will not be researched.

A connected series of concerns regards the particular dynamics that might limit the sharing of qualitative data. Institutional review boards (IRBs) are likely to require that researchers who wish to store and share data collected using interactive techniques such as interviews and focus groups solicit subjects' permission (Mauthner, Parry, and Backett-Milburn 1998: 743; Heaton 2004: 79). 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). On the one hand, the conditions under which subjects offer information can affect the type and form of data scholars collect. On the other, some data may require a resource-intensive process of contextualization in order to be shared (or may need to be stored in partial form). All of these eventualities can have significant consequences for analysis, interpretation, and inference.

Protecting human subjects and conforming fully to standards set by IRBs are crucial imperatives. Achieving those goals may prevent some data from being shared. Yet 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; scholars who need to edit their data in this way can explicitly note what sort of information was removed and assess the impact of its deletion on the remaining data. Differential user access to digital archives can also help to address human-subjects concerns.[5] In sum, these important concerns need not stand in the way of making a great deal of material available to a broad range of scholars.

Non-interactive forms of data collection, such as archival research, may also produce data that cannot be electronically archived, 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. Like those related to human subjects, these challenges require specialized solutions that it is very likely social scientists can develop through sustained, thoughtful debate and collaboration.

In sum, while not all data social science researchers collect and produce as part of qualitative and multi-method research can be archived and shared, a considerable amount likely can. Sharing a significant subset of the qualitative data with which social scientists work is vastly superior to the status quo, in which practically no qualitative data are publicly available in the United States.

The Road Ahead

Disciplinary norms will need to change if sharing and reusing qualitative data—and producing scholarship relying on secondary data analysis—are to become accepted practices. As noted in this symposium's introductory essay on the openness dialogue, the discipline could benefit from a sustained conversation about the merits and limitations of data sharing, and from clarifying guidance on data release and sharing in qualitative research.[6] Likewise, funding agencies could advise grantees to make qualitative data available, and more journals could require authors to do so.[7] Training opportunities, outreach, and guidance on qualitative data sharing (see Corti 2000, 6.3) could also be offered through the Institute for Qualitative and Multi-Method Research or the short courses taught at the annual APSA meetings.

Establishing a dedicated qualitative data repository like the one currently being developed, rather than building on a pre-existing quantitative archive, may be the most effective way to encourage qualitative researchers to deposit and use shared data. Such a specialized repository will be more attuned to the particular challenges posed by archiving qualitative data

and will have a larger demonstration effect. The repository will also be epistemically neutral—viewed broadly as a means of increasing the transparency of the evidentiary basis for interpretive, descriptive, or explanatory work based on qualitative data—and designed to be visible to, and open to communication and interaction with, a wide audience. And of course, as an electronic resource, the repository will be linked to the broad range of existing institution-specific and specialized archives that already exist.

Qualitative research makes vital contributions to political science, and qualitative data archiving holds the key to making qualitative and multi-method research more transparent and more replicable. 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 carefully considered and addressed. Nonetheless, those challenges may prove to be relatively minor in comparison with the tremendous utility that sharing and reusing qualitative data can provide.

Notes

[1] This piece draws extensively on an article that appeared in the January 2010 issue of *PS*, co-authored with Colin Elman and Lorena Vinuela. I would also like to thank the broader set of scholars who participated in a workshop convened to explore the idea of building a qualitative data repository held March 28–29, 2009, at Syracuse University (funded by NSF Grant SES 0838716).

[2] For instance, funding agencies in several OECD countries adopted a mandatory sharing policy for grant holders in the 1990s, and the repositories constructed as a result receive regular deposits on a national scale and hold a wide range of qualitative materials. Some examples include QUALIDATA in the UK, WISDOM in Austria, SDA of the Czech Republic, DDA of Denmark, FSD in Finland, Réseau Quetelet in France, GSDB-EKKE in Greece, GESIS in Germany, ADPSS Sociodata in Italy, CEPS in Luxembourg, DANS in the Netherlands, NSD in Norway, ARCES/CIS in Spain, and SND in Sweden.

[3] To be sure, several university libraries and research institutions have archives for data collected by their affiliated researchers and the facilities to archive 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.

[4] The project has been funded by NSF Grant SES 1061292.

[5] 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 original source's) permission and in accordance with explicit data-sharing agreements.

[6] 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.

[7] Most political science journals that have data-release policies either explicitly or implicitly limit those mandates to statistical data.

References

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

Brady, Henry and David Collier, eds. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed. Lanham, MD: Rowman & Littlefield.

Corti, Louise. 2000. "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.

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.

Elman, Colin, Diana Kapiszewski, and Lorena Vinuela. 2010. "Qualitative Data Archiving: Rewards and Challenges." *PS: Political Science & Politics* 43:1 (January), 23–27.

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

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), 733–745.

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

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): 139–152.

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.

Active Citation and Qualitative Political Science

Andrew Moravcsik
Princeton University
amoravcs@princeton.edu

This article presents a proposal for the adoption of "active citation," together with a discussion of why it is necessary, its possible advantages, and some potential concerns.[1] Active citation envisages the use of rigorous, annotated citations hyperlinked to the sources themselves. The goal is to provide opportunities for scholars to be rewarded not just for more rigorous but also for richer and more diverse qualitative scholarship.

The Problem: The Evidence in Qualitative Research Remains Invisible

Qualitative research dominates political science. While the use of statistical and formal methods is spreading, historical, qualitative, or textual research remains strong. In the field