

- Messy Data): A Modest Defense,” *Qualitative and Multi-Method Research* (Spring 2011) 8–17, 31–33.
- Gerring, John. 2012. “Mere Description.” *British Journal of Political Science* (forthcoming).
- Gerring, John and Rose McDermott. 2007. “An Experimental Template for Case-Study Research.” *American Journal of Political Science* 51:3 (July), 688–701.
- Glynn, Adam N. and Kevin M. Quinn. 2011. “Why Process Matters for Causal Inference.” *Political Analysis* 19:3, 273–286.
- Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. In Press. “Experimental Designs for Identifying Causal Mechanisms.” *Journal of the Royal Statistical Society Series A*.
- Kincaid, Harold. 1996. *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*. Cambridge: Cambridge University Press.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Laudan, Larry. 1996. *Beyond Positivism and Relativism: Theory, Method, and Evidence*. Boulder: Westview Press.
- Mahoney, James. 2008. “Toward a Unified Theory of Causality.” *Comparative Political Studies* 41:4–5 (April), 412–436.
- Mayo, Deborah G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
- Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge: Cambridge University Press.
- Reichenbach, Hans. 1938. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.
- Rubin, Donald B. 2008. “For Objective Causal Inference, Design Trumps Analysis.” *Annals of Applied Statistics* 2:3, 808–840.
- Seawright, Jason. 2010. “Regression-Based Inference: A Case Study in Failed Causal Assessment.” In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, 2nd ed. Henry E. Brady and David Collier, eds. (Lanham, MD: Rowman & Littlefield), 247–271.

Symposium: Research Openness in the Digital Age

Qualitative Data Access and Research Transparency

Colin Elman

Syracuse University
celman@maxwell.syr.edu

“...if people don’t know what you’re doing, they don’t know what you’re doing wrong.”¹

This brief essay draws on recent conversations about data access and research transparency. It discusses some of the issues involved, and describes a vocabulary to handle them. Finally, it explores some of the challenges of increasing openness in the context of qualitative research.

Data Dialogues

Issues of data and research transparency have long been considered important, particularly among political scientists who use quantitative methods. Recently, scholars have revisited a series of linked transparency issues—and expanded the conversations to include scholars who use qualitative techniques.² What responsibilities do scholars have to describe how they collected data and how they analyzed evidence to arrive at their conclusions? How (if at all) does the availability to subsequent scholars of the data and techniques scholars use impact the persuasiveness of the original scholars’ arguments? What is the value of allowing scholars other than those who collected the data to use them for analytic tasks beyond those of the original study?

Members of different research communities in the discipline arrived at these recent conversations with different perspectives. Quantitatively oriented scholars, who have been considering these issues for a longer period, commonly hold a commitment to replication and use data that are relatively less

problematic to share. Correspondingly, their concerns were to clarify these commitments, to consider improvements in infrastructure to facilitate data sharing, and to develop stricter standards, and in the meantime, ensure higher rates of compliance with already widely shared norms.

Qualitative scholars are for the most part both less well-equipped and less willing to join this dialogue. Yet the pressure for qualitative scholars to engage questions about research openness is increasing. Cognate disciplines in the United States like anthropology have made important progress on these issues, and European (and especially British) research communities have pushed further still. In addition, more and more data are being “born digital,” and younger generations of scholars arrive in the academy with expectations that information should be immediately accessible. Funding agencies and journals are also asking qualitative scholars to take a more considered approach to when qualitative data should be made accessible.

Political science as a discipline is constituted by a diverse set of research communities, with different views of the enterprise in which they are engaged. This diversity could easily derail progress, as each of the stakeholders seeks to frame the dialogue in terms that are most sympathetic to their research communities. A more promising approach might be to move forward with a two-stage approach: First, describe general concerns relevant for all types of social inquiry in the discipline, and second, consider how responses to those concerns might be instantiated in the context of research communities with different epistemological commitments.

Developing a Vocabulary to Discuss Data Access and Research Transparency

One challenge in generating an initial “whole discipline” dialogue is arriving at a vocabulary that is relevant to social inquiry in general *and* can accommodate alternate operationalizations for different research traditions. On the one hand, even

at this general level the central terms should have purchase on increasing data access and research transparency. On the other, the terms need to be sufficiently broad to have relevance to different research communities, and thus have the flexibility to be instantiated in various types of social inquiry.

A helpful point of departure is to acknowledge that empirical social inquiry analyzes information about the world to generate intersubjective knowledge statements. While scholarly communities have different purposes and techniques, in the end all strive for the same overarching goal: to produce and present the most *persuasive* accounts possible, within the terms set by their epistemological commitments. A scholar writes for an audience, most often members of her immediate academic community.

A core claim made by those who favor research openness is that, *ceteris paribus*, *transparent* evidence-based research is more persuasive. Scholars try to convince others that their research advances a valid argument. The more information scholars provide about the evidence on which they base their claims and the analytical machinery they employed to arrive at their conclusions, the more swayed we should be. We might consider three dimensions along which the openness of empirical research can be evaluated.³

Are evidence-based knowledge claims backed up with citations to the data they rely upon, including information about where those data were found? If scholars collected their own data (whether via interviews, archival research, or any other data collection technique) do they provide *data access*? Do they let others see their data (be they interview transcripts, primary documents, etc.) or explain why they cannot?

If scholars collected the data on which they rely, do they engage in *production transparency*—do they offer an account of the procedures used to collect or generate the data?

Are evidence-based knowledge claims marked by *analytic transparency*, that is, do they strive to describe the process through which scholars drew their conclusions based on their data?

These questions can apply, *mutatis mutandis*, to all empirical social inquiry. They can be mapped on to different research traditions without forcing a false equivalence among them. In short, we can agree on the general principles underlying these dimensions while acknowledging that they will be instantiated differently depending on the style of social inquiry being conducted.

The Challenges to Openness in Qualitative Research

Qualitative scholars might have several reasonable concerns about research openness. An initial set is listed below, together with some preliminary thoughts about their probable impact.

Epistemologically, many qualitative scholars identify transparency with replication, and believe that repetition by a subsequent scholar for the purposes of assessing validity is

neither possible nor desirable.

An inclusive dialogue about openness should be sensitive to these types of epistemological concerns, and explicitly eschew imposing inappropriate norms on types of social inquiry where they do not belong. While transparency is a necessary condition for replication, the opposite is emphatically not so. To say that providing more information about a knowledge statement strengthens one's argument does *not* imply that another scholar armed with the same information must be able to arrive at the same conclusion in order for that argument to be considered valid. In this context, the case for transparency is that the more material scholars include, the more sense readers can make of their arguments and the more convinced they may be by them. As one scholar observed, a review essay on a Shakespeare sonnet is likely to be better understood and more persuasive if readers also get to read the sonnet.

Of course, some scholars who use interpretive methods are comfortable with at least some types of replication (see, for example, Wedeen 2010: 265). Moreover, many members of the qualitative and multi-method community who aim to produce objective social science are at least notionally committed to the idea that scholars using the same research tools in the same sites will collect similar data and arrive at similar conclusions and arguments. Consider, for example, the "positivist" qualitative dissertation which goes on to be published as a book by an academic press. Archetypically, the volume comprises six chapters: an introduction, a discussion of relevant theory and the study's research design, three case study chapters based on data gathered through interactive or archival techniques, and a conclusion. Although rarely put to a direct replication test, work of this type does implicitly claim to convey objective information and draw conclusions that other scholars employing the same research techniques would convey and draw. Yet while there is an abstract assurance of openness in this type of qualitative research, for the most part research transparency amounts to a few paragraphs in the text discussing where the author found her data and how she collected them. If subsequent scholars want to actively evaluate her claims, they have to pay the transaction costs of visiting the data sites themselves.

The point here is not to re-litigate the arguments for and against replication. It is to argue that research openness is a broader ideal, and one from which scholars can benefit regardless of which viewpoint they take on replication. For qualitative scholars who abjure replication, transparency strengthens the persuasiveness of their inquiries and their arguments. For researchers who believe that a necessary element of a truth claim is that other scholars should be able to duplicate conclusions, transparency is a prerequisite.

Pragmatically, researchers worry that committing to transparency will make data collection (be it through interaction or archival research) more difficult, with limited pay-offs for the additional work.

Valuing openness does not require blindly following the principle to the detriment of research. Transparency is one among several competing objectives researchers might pursue, and it will not always take priority. For example, field re-

searchers do not expect to be able to videotape or record every one (and sometimes any) of their interviews. In some instances contemporaneous notes (or simply listening!) will be the best they can do. Researchers' main responsibility is to get the best data they can, using the approach most appropriate to the circumstances. Where factors prohibit researchers from sharing their data, or require them to present it in a particular form, they simply need to offer a clear account of what those obstacles were. With respect to payoffs, any reasonable conversation about research transparency and data access has to include the establishment of guarantees of first-use of data by their collectors/generators, but also the introduction of disciplinary reward structures that incentivize scholars to take on the burdens imposed by transparency.

To be sure, these types of pragmatic concerns are real and deserving of serious attention. But the absence of an American tradition of openness in qualitative social inquiry may also have generated a tendency to dwell on difficulties. Some of these concerns are likely to seem less worrisome once more public examples of transparent qualitative research are presented. In addition, because openness is achievable by a range of possible alternatives, once some of these pathways to transparency are described and used, more are likely to be developed. With a wider range of different approaches, it is likely that some may be better suited to address particular pragmatic concerns. This is another reason why it is better to think of the endpoint of this conversation as a menu, not a mandate.

Of course research openness will likely involve its own pitfalls and unintended consequences. For example, at the moment, scholars whose main data sources are primary documents cite, but do not provide links or access to, the primary sources underlying their claims. By accessing an author's primary sources by visiting an archive, a second researcher can both investigate whether those sources really do support the argument, *and* check for other sources that undermine it. Once authors begin to provide direct access to the primary sources they cite (or at the very least, substantial redactions), checking referenced materials will become much easier than exploring what materials might have been ignored. While this is not a good argument against openness, it is a strong reminder that transparency is a matter of degree and one step in a broader evaluative process. We should not confuse a scholar's provision of materials that underlie her claims with undeniable proof that those claims are valid, nor make the assumption that she has provided every piece of evidence that might bear on the argument.

Ethically, qualitative scholars quite properly worry that providing others access to their data may involve human subjects or legal concerns.

No conversation about openness in qualitative research can proceed without carefully considering ethical and legal obstacles to transparency, and including those concerns as valid grounds for withholding information. With regard to finding areas on which scholars can agree, this is the most straightforward of all the obstacles. Protection of human subjects, for example, is a concern which unites all researchers. The rule should be correspondingly straightforward: if the protection

of human subjects constrains what scholars can safely say, they should work within those limits. But they should also explain to readers why the data are unavailable, and (depending on the epistemic community in which they are operating) perhaps consider making the data available in an anonymized form.

Conclusion

Increasing openness is likely to have positive effects on qualitative research in the long run. As data become more immediately visible, and as the bar is raised for requiring explicit connections between authors' data, analysis, and conclusions (i.e., as they are encouraged to demonstrate how particular techniques were used to make inferences from particular data), we anticipate research designs and execution will become more rigorous.

In addition, there will be substantial pedagogical opportunities from being able to demonstrate the clear application of analytical techniques. To be clear, this does *not* require or even encourage convergence on a single method of data collection or data analysis. But it will need scholars to be much clearer about the choices they make among the alternatives.

We understand that there is some reticence about openness among different groups of qualitative researchers, but we nevertheless believe that this is a conversation worth having. Otherwise the discipline will push ahead without the active participation of qualitative researchers. We will either be left behind entirely or, worse, be squeezed into an unhelpful one-size-fits-all template. In addition, the wider discipline's view of qualitative research is likely to be heavily shaped by technological possibilities rather than epistemological imperatives. The worst case scenario is that qualitative research will be the passive ingredients to big data's synthesis of combine harvester and food blender, indiscriminately harvested and homogenized. A better solution is to instead begin with the research traditions, and develop transparency tools that fit.

Broad participation in shaping the dialogue about openness will maximize its benefit across the discipline, and allow us to shape application of transparency norms to fit our research styles. Accordingly, this short memorandum argues that qualitative researchers will be better off if they have a positive agenda that is the product of a conscious conversation, grounded in our epistemological, pragmatic, and ethical concerns.

Notes

¹ Lynn and Jay (1984: 21). Jonathan Lynn and Anthony Jay authored the BBC series *Yes Minister*, and wrote a book drawn from the series in the form of Jim Hacker's fictional diaries (but actually comprising a combination of those diaries with fictional memoranda and notes by other participants). The quotation in the epigram is from a note by Humphrey Appleby, and attributes the observation to Sir Arnold Robinson, Secretary to the Cabinet.

² In August 2010 Arthur Lupia of the University of Michigan and I co-authored a memorandum to the APSA Council, which resulted in the formation of an APSA Working Group on the topic of research transparency. Although the current essay draws on and reflects several themes of that dialogue, it does not constitute a report from the

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