

Openness in Political Science: Data Access and Research Transparency

Introduction

Arthur Lupia, *University of Michigan*

Colin Elman, *Syracuse University*

In 2012, the American Political Science Association (APSA) Council adopted new policies guiding data access and research transparency in political science. The policies appear as a revision to APSA's *Guide to Professional Ethics in Political Science*. The revisions were the product of an extended and broad consultation with a variety of APSA committees and the association's membership.¹

After adding these changes to the ethics guide, APSA asked an Ad Hoc Committee of scholars actively discussing data access and research transparency (DA-RT) to provide guidance for instantiating these general principles in different research traditions. Although the changes in the ethics guide articulate a single set of general principles that apply across the research traditions, it was understood that different research communities would apply the principles in different ways. Accordingly, the DA-RT Ad Hoc Committee formed sub-committees to draft more fine-grained guidelines for scholars, journal editors, and program managers at funding agencies who work with one or more of these communities. The subcommittees have produced circulation drafts for APSA members' review and comment. The drafts are titled *Guidelines for Data Access and Research Transparency in the Quantitative Tradition* and *Guidelines for Data Access and Research Transparency in the Qualitative Tradition*² and are attached as Symposium Appendices A and B.

This article is the lead entry of a *PS: Political Science and Politics* symposium on the ethics guide changes described above, the continuing DA-RT project, and what these endeavors mean for individual political scientists and the discipline. Its content is as follows. In the first section, we offer a brief history of how the ethics guide changes came about and our understanding of the motivations of the diverse group of scholars who work on the DA-RT initiative. In the second section,

we present the changes to the ethics guide. In the third section, we work from these changes to offer a broader argument about the value of greater openness to individual political scientists and to the discipline. We conclude by providing a brief summary of themes developed in the symposium's seven subsequent articles and inviting feedback.

With this content in mind, we want to draw your attention to the fact that DA-RT is an open endeavor. While we are listed as authors on this particular article, the progress made in this domain in recent years is the result of the effort of numerous social scientists. In addition to being open, DA-RT is an ongoing effort in which any political scientist can participate. We hope that you will find in this symposium ways to increase the value and impact of your efforts as teachers, researchers, and public servants.

HISTORY

Political science is a diverse discipline comprising multiple, and sometimes seemingly irretrievably insular, research communities. We could spend much of this introduction (indeed fill several issues of the journal) on the sociology of academic disciplines and why they tend to fragment. But recent discussions about openness are a rare and welcome example of dissimilar scholars finding opportunities for collaboration and common action.

Several years ago, APSA's governing council, under the leadership of president Henry E. Brady, began an examination of research transparency. Its initial concerns were focused on the growing concern that scholars could not replicate a significant number of empirical claims that were being made in the discipline's leading journals. There were multiple instances where scholars would not, or could not, provide information about how they had selected cases, or how they had derived a particular conclusion from a specific set of data or observations. Other scholars refused to share data from which others could learn. Still other scholars would have been willing to share their data, but failed to archive them in effective ways, making the information unavailable for subsequent inquiries.

As political scientists described such episodes to each other, they realized that scholars from different methodological and substantive subfields were having similar experiences and

conversations. In a wide range of circumstances, professional customs and incentives for sharing information and data were less well developed than those for producing knowledge claims. An unusually diverse set of political scientists identified common concerns and aspirations, both in their reasons for wanting greater openness and in the benefits that new practices could bring.

What is political scientists' shared interest in openness? As Elman and Kapiszewski (2014) note, openness is best understood as a meta standard that applies to all social inquiry. All rule-based social inquiry is based on three notions: first, scholarly communities hold shared and stable beliefs that research designed and conducted in particular ways possesses certain characteristics. Second, both the conduct of social inquiry and the written products that represent its conclusions are designed to capture those characteristics. Finally, for any given piece of research in a particular tradition, the ability of scholars to claim the underlying warrants depends on their showing that it was designed and conducted in accordance with those rules. The view that social science is a group activity, requiring inter-subjective knowledge being created using public processes that are warranted to add value, is common to virtually every scholarly tradition.

Therefore, DA-RT operates from a “community standards” approach, where optimal means of data sharing and research transparency respect and build from the challenges and opportunities that characterize various research traditions.

Communities have very different beliefs about what constitutes useful knowledge and how such value is to be obtained. That said, there is substantial overlap about which attributes of openness contribute to accurate inter-subjective knowledge transfer. Our prescriptive methodologies all involve extracting information from the social world, analyzing the resulting data, and reaching a conclusion based on a combination of the evidence and its analysis. No matter whether the research is, for example, ethnographic field work, a laboratory experiment, or the statistical analysis of a large data set, they all combine assumptions, decisions, and actions that produce evidence and analysis. Sharing information about these assumptions, decisions, and actions is necessary for scholars to place one another's meanings in a legitimizing context. DA-RT is motivated by this premise—the principle that sharing data and information fuels a culture of openness that promotes effective knowledge transfer.

This justification for openness (the desire to establish a knowledge claim's validity) and its general content (showing both evidence and analysis) are epistemically neutral. They apply wherever scholars seek to use a shared logic of inquiry to reach evidence-based conclusions. To this end, a critical attribute of DA-RT is that it does not impose a uniform set of standards on political scientists. Instead, it begins from a simple premise about credibility and legitimacy. In short, schol-

ars who produce knowledge claims want others to have a rationale for believing those claims. Therefore, DA-RT operates from a “community standards” approach, where optimal means of data sharing and research transparency respect and build from the challenges and opportunities that characterize various research traditions. Because social scientists use different methods, how a knowledge claim achieves credibility and legitimacy depends on the type of work. For all research traditions in political science, our main focus is to better equip its scholars with incentives and mechanisms for making their knowledge claims easier for others to interpret and assess accurately.

That said, the shared commitment to openness places limits on practices that DA-RT can endorse. For example, DA-RT rules out claims about the credibility and legitimacy of scientific claims based solely on personality cults or on raw exercises in power (i.e., “the claim is true because my minions and I so testify”). What distinguishes scientific claims from others is the extent to which scholars attach to their claims publicly available information about the steps that they took to convert information from the past into conclusions about the past, present, or future.

The credibility of scientific claims comes, in part, from the fact that their meaning is, at a minimum, available for other

scholars to rigorously evaluate. In other words, the reason to believe a scientist's claim is not because he or she wears a lab coat, have a PhD, or have published a widely viewed paper in the past. Appeals to personality or faith, which facilitate information transmission in other domains, are not supposed to be required to access the content of a scientific claim. A claim's perceived legitimacy is grounded in the fact that the results are the product of publicly described processes that in turn are based on a stable and shared set of beliefs about how knowledge is produced. Such open access to the origins of others' claims is the hallmark of scientific ways of knowing.

Accordingly, when social scientists fail to document their assumptions, decisions, and actions and are unwilling or unable to share this information with others, it limits others' abilities to understand the meaning of the scientists' claims. When such failures are frequent in a research community, the credibility and legitimacy of the community as a whole are imperiled. Across the sciences, questions about data sharing and research transparency are now being increasingly and vigorously addressed. Advances in electronic communication not only expose scholars to a wider set of knowledge claims, but also give them reasons to expect that data and inferential information can be made more readily available. DA-RT is one of several efforts in the social sciences to advance the cause of transparency.

DA-RT's distinction is that it is focused on political science. Our goal is to provide, through a community standards approach, individual scholars of every epistemic tradition opportunities for greater openness, transparency, legitimacy, and credibility. This goal has motivated a diverse set of scholars to contribute to the DA-RT project. These scholars have developed a wide range of mechanisms to increase professional incentives for data sharing and research transparency. They have also worked to make such activities easier for a growing range of scholars. DA-RT is a movement that anyone interested in political science can join.

The credibility of scientific claims comes, in part, from the fact that their meaning is, at a minimum, available for other scholars to rigorously evaluate. In other words, the reason to believe a scientist's claim is not because they wear a lab coat, have a PhD, or have published a widely viewed paper in the past. Appeals to personality or faith, which facilitate information transmission in other domains, are not supposed to be required to access the content of a scientific claim.

ETHICS GUIDE CHANGES

APSA's ethics guidelines now state that "researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated." The three constitutive elements are defined as follows:

- 6.1 *Data access*: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they themselves generated or collected, researchers should provide access to those data or explain why they cannot.
- 6.2 *Production transparency*: Researchers providing access to data they themselves generated or collected, should offer a full account of the procedures used to collect or generate the data.
- 6.3 *Analytic Transparency*: Researchers making evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data, i.e., clearly explicate the links connecting data to conclusions.
- 6.4 Scholars may be exempted from Data Access and Production Transparency in order to (A) address well-founded privacy and confidentiality concerns, including abiding by relevant human subjects regulation; and/or (B) comply with relevant and applicable laws, including copyright. Decisions to withhold data and a full account of the procedures used to collect or generate them should be made in good faith and on reasonable grounds. Researchers must, however, exercise appropriate restraint in making claims as to the confidential nature of their sources, and resolve all reasonable doubts in favor of full disclosure.

6.5 Dependent upon how and where data are stored, access may involve additional costs to the requesting researcher.

6.6 Researchers who collect or generate data have the right to use those data first. Hence, scholars may postpone data access and production transparency for one year after publication of evidence-based knowledge claims relying on those data, or such period as may be specified by (1) the journal or press publishing the claims, or (2) the funding agency supporting the research through which the data were generated or collected.

6.7 Nothing in this section shall require researchers to transfer ownership or other proprietary rights they may have.

6.8 As citizens, researchers have an obligation to cooperate with grand juries, other law enforcement agencies, and institutional officials. Conversely, researchers also have a professional duty not to divulge the identity of confidential sources of information or data developed in the course of research, whether to governmental or nongovernmental officials or bodies, even though in the present state of American law they run the risk of suffering an applicable penalty.

6.9 Where evidence-based knowledge claims are challenged, those challenges are to be specific rather than generalized or vague. Challengers are themselves in the status of authors in connection with the statements that they make, and therefore bear the same responsibilities regarding data access, production transparency, and analytic transparency as other authors.

While data access and research transparency are the "default" settings in the new guidelines, these expectations are contingent on the author not putting people at risk or breaking the law. Hence concerns about human subjects protections and copyright limitations are accounted for in the new language.

With these changes, APSA's ethics guide is more consistent with current and emerging standards across the sciences. Where APSA's previous language emphasized making data accessible *only* when findings were challenged, the new guidelines recognize data access and research transparency as an indispensable part of the research endeavor. It is also critical to notice that the updated language is epistemically neutral: it respects the integrity of different research traditions, and the diverse data collection and analytic steps that they take.

HOW POLITICAL SCIENCE BENEFITS FROM INCREASED OPENNESS

A more rigorous and self-conscious approach to openness promises several benefits to political scientists. One way to categorize these benefits is with respect to the different audiences for political science scholarship.

First, and most obviously, transparency offers an opportunity for members of a particular research community to understand and assess their own scholarship. Data sharing and research transparency allow a researcher's audience to evaluate claims and form an evidentiary and logical basis for treating the claims as valid.

The most widespread (although as we note below, not universal) way that this principle is pursued is through replication. For subfields that hold that inferential procedures are

repeatable, openness is a necessary condition for replication. For these communities, replication of another's claims provides increased confidence in the validity of that work.

Other audiences are not focally involved in research. Instead, they want to use research claims as the basis of action. Teachers, for example, want to use the claims for pedagogical purposes. Whether demonstrating substantive arguments about aspects of the social world, or training students to use research techniques, teaching is substantially improved by the availability of exemplary scholarship, with its data and reasoning on display.

Public and private sector decision makers comprise another audience. Their main interest is in using knowledge claims to improve the effectiveness and efficiency of valuable endeavors. Greater openness gives such audiences increased oppor-

For subfields that hold that inferential procedures are repeatable, openness is a necessary condition for replication. For these communities, replication of another's claims provides increased confidence in the validity of that work.

repeatable, openness is a necessary condition for replication. For these communities, replication of another's claims provides increased confidence in the validity of that work. When subfields have such confidence, they can devote their attention to evaluating competing theories of important phenomena. If, by contrast, opportunities for replication are diminished because of poor data availability or incomplete accounts of how results were reached, it is impossible to determine the strength or robustness of findings—which makes confidence harder to build.

Members of other research communities do not validate one another's claims by repeating the analyses that produced them. In these communities, the justification for transparency is not replication, but understandability and persuasiveness. The more material scholars make available, the more that they can accurately relate such claims to a legitimating context. When readers are empowered to make sense of others' arguments in these ways, the more pathways exist for readers to believe and value knowledge claims. Whether scholars privilege replication, context-specificity, or other ways of evaluating the meaning of a knowledge claim, sharing information that allow such evaluations facilitates knowledge transfer. Hence, research openness is a broader ideal, and one from which scholars can benefit regardless of which viewpoint they take on replication.

Second, openness is beneficial for scholars outside the immediate community in which the research is located. Political science is a methodologically diverse discipline, and we are sometimes unable to appreciate how other social scientists generate their conclusions. Mathematical modelers, for example, often know very little about how cases are selected in participant observation studies—and many people who seek meaning in texts have a limited understanding of how other social scientists try to seek meaning from surveys or com-

puter simulations of war. Higher standards of data access and research transparency will make cross-border understanding more attainable.

Beyond general openness, data sharing provides an important additional benefit—it allows secondary analysis. Shared data can be a valuable public good. Secondary analysts can use data in ways that data originators did not. In the best-case scenario, secondary data analyses allow authors to derive meaning from data that need not have occurred to the original researcher. When scholars can use research materials in these diverse ways, the data can become more valuable to science and society. Instead of a dataset producing one set of insights, data sharing gives other scholars the ability to multiply datas' value.

Many of these benefits of openness are widely known. We have found, however, that while the goals of greater data sharing and research transparency are generally accepted, they are less often followed in practice. Most political scientists to whom we have spoken find nothing radical or challenging about the notion that they show the information and analysis underpinning their evidence-based claims. But as the articles in this symposium show, there are multiple instances in which individual actions do not live up to our shared aspirations.

One challenge is that quantitative and qualitative research traditions lack clearly specified guidelines as to what kinds of data and research information should be shared. Compounding this problem is a lack of professional incentives for documenting the evidentiary and logical foundations of knowledge claims, the temporal and monetary expense that can be involved in archiving research materials, and the potential for embarrassment that can come from having one's work

reexamined. These are all substantial headwinds confronting transparency movements. The question for individual investigators and the discipline as a whole is whether we can derive the benefits of greater openness while recognizing, and then minimizing, the costs.

The contributions to this symposium are motivated principally by such challenges and questions.

TOPIC OF THE SYMPOSIUM: NEXT STEPS IN DATA ACCESS AND RESEARCH TRANSPARENCY

This symposium contains seven articles on DA-RT-related activities. Each article is written by scholars interested in investigating the benefits of greater openness and offering ideas about how to make data access and research transparency more viable and incentive-compatible activities for all political scientists. The distinct contribution of each article to this cause is to identify where potential gains from openness are appar-

The distinct contribution of each article is to identify where potential gains from openness are apparent but not yet fully realized. In each case, the authors seek to reconcile individual incentives, existing norms, and possible ways of changing rewards and technology to increase the frequency and effect of greater openness.

ent but not yet fully realized. In each case, the authors seek to reconcile individual incentives, existing norms, and possible ways of changing rewards and technology to increase the frequency and effect of greater openness.

This introduction is followed by two articles focused on qualitative research. Colin Elman and Diana Kapiszewski discuss how openness is instantiated differently in diverse qualitative research traditions. They illustrate this discussion with a brief account of some concerns that arise when making process tracing research transparent. Andrew Moravcsik shows how a practice called *active citation* can be implemented to increase the credibility and legitimacy of a wide range of qualitative research.

The next two articles (Arthur Lupia and George Alter, and Allan Dafoe) concentrate on large-N observational studies. Lupia and Alter discuss general opportunities for, and challenges to, increased openness that face quantitative scholars. Dafoe cites the benefits of sharing complete replication files for scholars who base conclusions on various forms of high-N statistical inference.

Rose McDermott focuses on experimental research. She discusses several innovative openness proposals in that domain including experimental registries—a system where scholars commit to publicizing their research designs before collecting data so that readers can better evaluate the meaning and generalizability of experimental results. The symposium concludes with articles by Thomas M. Carsey and John Ishiyama on the topic of how to implement critical elements of the DA-RT agenda. Carsey, director of the Odum Institute, describes new and emerging archiving opportunities and makes a strong argument for how the success of such opportunities is tied to deci-

sions that we make about graduate student training. Ishiyama, lead editor of the *American Political Science Review*, describes the different ways in which journals are adapting to calls for greater openness. He concludes by offering a number of different ways that journals can better address demands for greater openness, including replication studies.

In many areas of the discipline, there are limited incentives to increase openness. At the same time, there are multiple levers the discipline can pull to increase openness's incentive compatibility for the purpose of augmenting political science's legitimacy. These levers include changing disciplinary norms so that data production is valued for promotion and tenure, developing software tools to lower barriers to entry for curating data (for example, the Active Citation Editor (ACE) and the Live Active Citation Editor (LACE) in qualitative research), and incentivizing graduate students for greater openness from the beginning of their careers.³

Each contributor to this symposium offers creative ideas about how to move forward and each of their views has informed our own. Taken together, the articles make the case that openness is an indispensable element of credible research and rigorous analysis, and hence essential to both making and demonstrating scientific progress. These articles represent the great energy for increased credibility that a deeper and more sustained commitment to DA-RT principles can bring.

If you are not yet familiar with DA-RT, the changes to the ethics guide, and their implications for future activity in our discipline, then this symposium is a good place to learn more about these topics. Having engaged the materials, we hope that you will join our effort. Admission is free, and we can use all the help that you can offer. ■

NOTES

1. The first DA-RT text was drafted by an Ad Hoc Committee, which consisted of Arthur Lupia (University of Michigan), Colin Elman (Syracuse University), George C. Alter (University of Michigan), Brian D. Humes (National Science Foundation), Diana Kapiszewski (Georgetown University), Rose McDermott (Brown University), Ron Rogowski (University of California, Los Angeles), S. Laurel Weldon (Purdue University), and Rick Wilson (Rice University). The suggested changes were reviewed and amended by APSA's Committee on Professional Ethics, Rights, and Freedoms, which consisted of Richard G.C. Johnston (University of British Columbia), Michael Lienesch (University of North Carolina, Chapel Hill), Marion Smiley (Brandeis University), Philip A. Schrodt (Pennsylvania State University), Sarah Birch (University of Essex), and Christian Davenport (University of Notre Dame). At the spring 2012 APSA Council Meeting in Chicago, the council adopted the language put forward by the Ethics Committee as APSA policy. The language was posted to APSANET and circulated to the membership. Following that consultation, the council at its October 2012 meeting formally voted to include the new language in the association's *Guide to Professional Ethics*.
2. To ensure continuity, and so that the process could benefit from the Ad Hoc Committee's expertise, the follow-on committees include a

combination of the original members and newly invited participants. The qualitative committee is chaired by Colin Elman (Syracuse University), and includes Diana Kapiszewski (Georgetown University), Rose McDermott (Brown University), Andrew Moravcsik (Princeton University), Brian Humes (National Science Foundation), Elizabeth Saunders (George Washington University), and Marc Trachtenberg (University of California, Los Angeles). The quantitative committee is chaired by George Alter (University of Michigan and Director of ICPSR), and includes Arthur Lupia (University of Michigan), Brian Humes (National Science Foundation), Gary King (Harvard University), Christopher Zorn (Pennsylvania State University), Rick K. Wilson (Rice University), Michael Alvarez (California Institute of Technology), Dara Strolovitch (University of Minnesota), Tom Carsey (University of North Carolina, Chapel Hill), and Valerie Martinez-Ebers (APSR and University of North Texas).

3. ACE and LACE are currently in development at Syracuse University's Qualitative Data Repository.

REFERENCES

- Elman, Colin, and Diana Kapiszewski. 2014. "Data Access and Research Transparency in the Qualitative Tradition." *PS: Political Science and Politics* 47 (1): this issue.
- Lupia, Arthur. 2014. "What Is the Value of Social Science? Challenges for Researchers and Government Funders." *PS: Political Science and Politics* 47 (1): this issue.

SYMPOSIUM AUTHORS

George Alter is professor of history at the University of Michigan and director of the Inter-university Consortium for Political and Social Research. His research interests lie in the history of the family, demography, and economic history. Recent work explores demographic responses to economic hardship in Europe and East Asia, and the effects of childhood experiences on health in old age. He can be reached at altergc@umich.edu.

Thomas M. Carsey is the Pearsall Distinguished Professor of Political Science at the University of North Carolina at Chapel Hill. His research focuses on representation in US state and national politics, campaigns and elections, party polarization, and quantitative research methods. He also serves as Director of the Odum Institute for Research in Social Science, which operates a large social science data archive, and is the editor for the academic journal *State Politics and Policy Quarterly*. He can be reached at carsey@unc.edu.

Allan Dafoe is assistant professor of political science at Yale University. His research examines the causes of war, with emphases on the character and causes of the liberal peace, reputational phenomena such as honor and tests of resolve, and escalation dynamics. He can be reached at allan.dafoe@yale.edu.

Colin Elman is associate professor of political science, at the Maxwell School of Citizenship and Public Affairs, Syracuse University. His areas of interests are international relations, national security and qualitative methods. He is co-founder and director of the Institute for Qualitative and Multi-Method Research,

which offers intensive social science methods training, and co-editor of Cambridge University Press' *Strategies for Social Inquiry* series. He is co-director of the Qualitative Data Repository. He can be reached at celman@maxwell.syr.edu.

John Ishiyama, University Distinguished Research Professor of Political Science at the University of North Texas and lead editor of *American Political Science Review*, is a comparative politics scholar, who specializes in political parties and democratization in post-communist Russian, East Central European, and African (particularly Ethiopian) politics. He has also done considerable work on ethnic conflict and politics (particularly the role played by ethnic parties) and on the scholarship of teaching and learning. He can be reached at john.ishiyama@unt.edu.

Diana Kapiszewski is assistant professor of government at Georgetown University. Her research focuses on comparative judicial politics and qualitative methods in political science. She is a co-author of *Field Research in Political Science* which will be published by Cambridge University Press in 2014. Kapiszewski was recently awarded the David Collier Mid-Career Achievement Award by the APSA Section for Qualitative and Multi-Method Research. She is co-director of the Qualitative Data Repository. She can be reached at dk784@georgetown.edu.

Arthur Lupia is the Hal R. Varian Collegiate Professor of Political Science at the University of Michigan. He has served on APSA's Governing Council, Executive Council and as its treasurer. He is president of the Midwest Political Science Association and chair of

American Association for the Advancement of Science's Social, Behavioral, and Economics Division. He is principal investigator of EITM and has served as principal investigator of the American National Election Studies and TESS. He can be reached at lupia@umich.edu.

Rose McDermott is professor of political science at Brown University. Her main area of research revolves around political psychology in international relations. She has authored three books, co-edited two additional books, and has written numerous articles and book chapters on experimentation, evolutionary and neuroscientific models of political science, political behavior genetics and the impact of emotion on decision making. She can be reached at Rose_McDermott@brown.edu.

Andrew Moravcsik is professor of politics and director, European Union Program, at the Woodrow Wilson School at Princeton University. He has authored over 125 scholarly publications, including four books, on European integration, international relations theory, qualitative/historical methods, and other topics. He has served as trade negotiator for the US government, special assistant to the Deputy Prime Minister of the Republic of Korea, press assistant for the European Commission, editor of a Washington foreign policy journal, and on various policy commissions at Princeton University. He can be reached at amoravcs@princeton.edu.

APPENDIX A

DRAFT August 7, 2013

Guidelines for Data Access and Research Transparency for Qualitative Research in Political Science¹

In October 2012, the American Political Science Association (APSA) adopted new policies requiring transparency in political science research. The new policies have been integrated into Section 6 of the Association's *Guide to Professional Ethics, Rights and Freedoms* (and are reproduced in the Appendix to this document).

The new standards require researchers making evidence-based knowledge claims in their published work to provide data access, and engage in production transparency and analytic transparency.

- *Data access* requires authors to reference the data on which their descriptive and causal inferences and interpretations are based and, if they generated or collected those data, to make them available or explain why they cannot.
- *Production transparency* requires authors who collected and/or generated the data serving as evidence for their claims to explain the genesis of those data. Production transparency is necessary for other scholars to understand and interpret the data which authors have made available.
- *Analytic transparency* requires that authors demonstrate how they used cited data to arrive at evidence-based claims.

The promulgation of an APSA standard underscores a growing disciplinary (and multi-disciplinary) consensus that data access, production transparency and analytic transparency are all critical aspects of the research process. Transparency contributes to the credibility and legitimacy of political science research and facilitates the accumulation of knowledge. Assessing, critiquing, and debating evidence-based claims made in published research require access to the data cited to support them, documentation and metadata describing how those data were generated or collected, and an explanation of how the evidence and claims are connected. Providing access to data, and to documentation describing data generation or collection, also makes data more useful for testing new theories, for the development of new datasets and bodies of evidence, and for other forms of secondary data analysis.

Data access, production transparency, and analytic transparency are interconnected. Data access is a precondition for evaluating how data are used. Production transparency is a key prerequisite for evaluating author-provided data, and the connections that authors posit between those data and their inferences and interpretations. Conversely, one can more effectively evaluate an author's data generation or collection techniques (revealed through production transparency) when one knows for what analytical use the data are intended.

¹ We gratefully acknowledge the assistance of Louise Corti, Associate Director and Functional Director, UK Data Service, for helpful comments on earlier versions of this document.

This document is a resource for scholars, journal editors and academic evaluators (reviewers, funders or award committees) who seek assistance in satisfying these new data access and research transparency obligations in the context of qualitative research.² Accordingly, the document provides prospective guidance for meeting the obligations, as well as for retrospectively assessing whether they have been satisfied. While the new standards encourage as much data sharing and research transparency as possible, they should not be viewed in all-or-nothing terms: these activities often face friction, for example in the form of human subjects or copyright concerns. Sharing some data and being as transparent as possible, within those or other limits, will generally be better than doing neither at all.

The document's contents apply to all qualitative analytic techniques employed to support evidence-based claims, as well as all qualitative source materials.³ No matter which qualitative techniques scholars use, research-tradition specific standards of transparency allow scholars to demonstrate the richness and rigor of qualitative work, and make clear its considerable contributions to knowledge accumulation and theory generation.

The Argument for Research Tradition-Specific Transparency Practices

The need for transparency in qualitative political science research derives from the fundamental principles which underlie social science as a community-based activity. Enhancing transparency both augments the quality of qualitative political science and increases its salience in and contributions to the discipline. Transparency is best achieved in qualitative political science in ways that preserve and honor that research tradition. We argue each of these points in turn.

Why Adopt Transparency Practices?

Transparency is an indispensable element of rule-bound intersubjective knowledge. Scholarly communities in the social sciences, natural sciences and evidence-based humanities can only exist if their members openly share evidence, results and arguments. Transparency allows those communities to recognize when research has been conducted rigorously, to distinguish between valid and invalid propositions, to

² A parallel set of guidelines intended primarily for quantitative data has also been developed. Of course, the guidance concerning ways in which research can be made more transparent offered in these documents is not exhaustive. In particular, nothing here is intended to prevent or discourage the development of more fine-grained requirements attuned to a particular subset of research, such as registering research designs involving experiments prior to conducting research with the aim of preventing publication and reporting bias.

³ Such materials encompass traditional sources, such as primary textual documents and published primary sources; data from interviews, focus groups, or oral histories (in either audio or video form or transcripts from or summaries thereof); field notes (for instance from participant observation or ethnography); diaries and other personal records; and press clippings. The guidelines also apply to less conventional sources such as samples from bodies of secondary work; photographs; maps, posters and other representational work; and artwork.

better comprehend the subjective social understandings underlying different interpretations, to expand the number of participants in disciplinary conversations, and to achieve scientific progress.

To date, this fundamental attribute of community-based knowledge generation has played out in political science primarily in the realm of replicating quantitative research. In contrast to the situation in legal academia, historical studies, classical philology and some other disciplines, in qualitative political science transparency norms have been weak or non-existent. To be sure, citations and references in qualitative research appear to assure openness. Nevertheless, imprecision in citation, the high transaction costs of actually locating cited evidence, and the opacity of links between data and conclusions, combine to make the critical evaluation of descriptive and causal inferences or cumulative deepening of data analysis a rare event.

The aim of transparency is to make the rigor and power of good qualitative research more visible, allowing and empowering each consumer to identify such research, and facilitating the awarding of appropriate credit. Further, increasing the ease with which a larger number of scholars can critically engage with qualitative research, and the depth with which they can do so, makes it more likely that such work will be incorporated into scholarly discussion and debate, and future research. In all these ways, enhancing understanding of the processes and products of qualitative research facilitates the accumulation of knowledge.

Why an Approach to Transparency that is Specific to Qualitative Research?

Transparency in any research tradition – whether quantitative or qualitative – requires that scholars show they followed the rules of data collection and analysis that guide the specific type of research in which they are engaged. That conformity is foundational to the validity of the resulting interpretations and inferences and its demonstration is a key component of social science.

A shared commitment to openness, however, does not oblige all research traditions to adopt the same approach. Rather, transparency should be pursued in ways and for reasons that are consistent with the epistemology of the social inquiry being carried out. There are several reasons why qualitative scholars should not (and sometimes simply could not) adopt the transparency practices employed by quantitative political scientists, but must instead develop and follow their own.

We begin from the position that qualitative research is invaluable, generating knowledge that could not be produced through any other form of inquiry. Such research generally entails close engagement with one or more cases, producing thick, rich and open-ended data. These data are collected and used by scholars with a range of epistemological beliefs, producing a wide variety of interpretations and inferences.

For qualitative scholars who are comfortable with replication (i.e., the repetition of a research process or analysis in an attempt to reproduce its findings), the case for transparency makes itself. Without transparency there can be no replication. Yet even qualitative scholars who do not share a commitment to replication should value greater visibility of data and methods. For instance, those who believe that an important social scientific task is to encourage recognition of the extent and importance of cultural, historical and social diversity should acknowledge the value of transparency in permitting the record of actors speaking in their own voices to reach readers of social scientific texts. In short, the more sense scholars can make of authors' arguments and evidence, the better they can engage them, the more varied techniques they can use to evaluate and document their legitimacy, and the more scholars can enter the conversation.

Transparency in qualitative research needs to be achieved and evaluated in ways that are sensitive to the nature of qualitative data, how they are gathered, and how they are employed. As the list offered previously suggests (see footnote 3), qualitative data take on more varied forms than quantitative data, and are less-structured. In terms of data collection/generation, qualitative scholars very commonly gather their own data, rather than rely solely on a shared dataset. Evaluating the processes used to obtain data is a key element in assessing qualitative work – not least because those processes have a critical effect on the research product. With respect to employment, qualitative data are used in a range of research designs, including single case studies, small-n case studies, and various mixed-method designs. A variety of methods are used to analyze qualitative data (e.g., narratives, counterfactual analysis, process tracing, Qualitative Comparative Analysis, content analysis, ethnographic analysis), and different inferential structures underpin each method. These fundamental facets of qualitative research have implications for how transparency can and should be achieved.

These epistemological considerations are reinforced by the especially acute ethical and legal imperatives, and the sociological framing of transparency, in qualitative research. The two most important ethical and legal imperatives with which transparency can be in tension in qualitative research are human subject and copyright concerns. Sometimes data are collected in circumstances that require discretion to protect the rights and welfare of subjects. This will, quite properly, limit transparency. Moreover, many sources are not, in their entirety, in the public domain, and there are limitations on how they can be shared. As noted below, scholars should only make qualitative data (and information about the decisions and processes that produced them) available in ways which conform to these social and legal imperatives.

Sociologically, no amount of written guidance will result in changes in transparency practices unless scholars believe that methods and research goals about which they care are being preserved and improved. A separate set of guidelines for qualitative research helps to establish that the aim of transparency is to demonstrate the power of qualitative research designs, data-collection techniques, interpretative modes, and analytic methods. In other words, rather than tacitly encouraging changes to qualitative research practices, the goal of enhanced transparency in qualitative research is precisely to *preserve* and *deepen* existing qualitative research traditions, render current qualitative

research practices more accessible, and make clearer the tremendous value-added qualitative research already delivers.

In short, while transparency is a universal principle, for epistemological, ethical, and sociological reasons, its instantiation in qualitative research needs to conform to traditions specific to qualitative work.

Data Access

Clause 6.1 in the revised APSA *Ethics Guide* obliges a scholar who makes evidence-based claims in her published work to reference the data she used to make those claims. If the scholar generated or collected the data herself, then she should also make those data available or explain why she cannot.

What data should be referenced and/or made available, and how?

Researchers making evidence-based knowledge claims should clearly and completely reference the data on which they base their interpretations or their descriptive or causal inferences. Generally, these are the data the author explicitly cites to support those claims.

Referencing textual data requires a full and precise bibliographic citation including page numbers and any other information necessary for readers to locate the material cited and find within it the passage an author suggests is evidence for his claims. For primary archival sources, for instance, information about the archive and collection, and the number of the box in which the document was found should be included. For non-textual sources, information allowing an equivalent degree of precision should be included.⁴ This information should be provided upon publication.

The new APSA standard entails a more stringent obligation for scholars who themselves generated or collected the data on which their evidence-based knowledge claims are based. Those scholars must, whenever possible, make those data available.⁵ Later in this document, we discuss strategies for, and issues involved in, sharing qualitative data.

Sharing cited data is sufficient to meet the APSA standards. Nonetheless, for many qualitative researchers, cited data are often a small subset of the information collected and used in a research endeavor. As such, researchers are strongly encouraged to share data which are implicated in their research but not cited in their publication – for

⁴ To give an example, when citing an audio tape, scholars might indicate the exact moment during the interview at which the cited material was mentioned (i.e., cite the time stamp), or might provide an extract of the recording and cite where it came from within the interview (e.g. this clip is six minutes in).

⁵ As noted later, scholars using Active Citation to achieve transparency must provide substantial excerpts from the data sources underlying their claims (and ideally provide the actual data sources) no matter whether they generated or collected those data or other scholars did so. If the source or the relevant portion thereof cannot be provided for ethical or legal reasons, a summary or redaction must be offered.

instance, additional data used to generate the argument (rather than test it), or to infirm alternative interpretations and inferences.

What limitations might there be on making qualitative data available?

It is critically important that scholars sharing data comply with all legal and ethical obligations. As paragraph 6.4 of the *APSA Guide to Professional Ethics* notes, while it is incumbent upon researchers to accurately represent the research process and study participants' contributions, external constraints may require that they withhold data, for instance, in order to protect human subjects or to comply with legal restrictions.

Confidentiality and Human Subjects: If scholars have promised the individuals whom they involved in their research confidentiality, it is incumbent upon them not to reveal those subjects' identities. Personal identity can be disclosed both directly (for example, through divulging a participant's address, telephone number, age, sex, occupation, and/or geographic location) or indirectly (for example, by disclosing information about the person that, when linked with publicly available information, reveals his/her identity).

Data garnered from human subjects can often be shared legally and ethically if the appropriate informed consent is granted by project participants.⁶ Where necessary, additional protective steps can be taken including guaranteeing confidentiality when soliciting informed consent;⁷ employing anonymization strategies;⁸ carefully controlling access to data; and/or requiring that special measures to protect confidential information be clearly specified in a data-use agreement signed by anyone who wishes to view or analyze the data.

Documentary Data: Sometimes the owners or licensors of data collected through non-interactive techniques—archives or non-governmental organizations, for instance—place limitations on their use or dispersion. Likewise, such materials sometimes have copyright restrictions. Scholars should make every attempt to explain the value of data-sharing to those from whom they acquire documentary data, and investigate to what degree, and which, copyright law applies.

⁶ To be clear, there are instances in which a researcher who has obtained permission from a subject to share data should nonetheless not do so because, for example, the subject is not in a good position to evaluate the risk of information connected with him/her being made accessible, or the circumstances under which permission was granted have changed. Alternatively, the author may decide to impose additional safeguards not specified in the informed consent when sharing the data.

⁷ When seeking informed consent researchers should secure permission for data sharing where possible, and should avoid including statements in consent forms that purposefully preclude data sharing beyond the researchers or team.

⁸ For instance, on some occasions scholars will only be able to characterize their source ("a senior government official"), but will be able to attribute full quotations to him or her; on other occasions, they will be able to indicate that they consulted with particular people, but will not be able to attribute any specific information to them.

Proprietary Data: When research is based on proprietary data, authors should make available sufficient documentation so other scholars can evaluate their findings. Owners of proprietary data should be encouraged to provide access to bona fide researchers.

As the discussion of types of data ‘friction’ in this section makes clear, the exclusions and restrictions that can prevent authors from sharing the data that support their analytic claims are circumstantial, ethical and legal. Accordingly, where data cannot be shared, the author should clearly explain why not, and include as much information about those data as is ethically and legally possible, to help readers understand and evaluate the author’s inferential and interpretive claims.

When should data be made available?

The APSA standards recognize that “Researchers who collect or generate data have the right to use those data first.” A particular collection of data should be made available no more than one year after the earliest publication (either electronic or paper) of evidence-based statements made using that collection.

The APSA standards also recognize that journals and funding agencies may have different requirements (for instance, obliging researchers to make the data used in a book or article available prior to any publication). The one-year allowance specified by APSA does not alter any time limits established by journals and funding agencies.

Where and in what form should data be made available?

The best practice is for digital data (e.g. PDFs of documents, audio files, video files) to be made accessible online, at an established repository that can be discovered by standard Internet search engines. Standard and non-proprietary file formats are preferable, because they are more likely to remain accessible over time. For non-digital data, scholars should provide a metadata record identifying the source.

When deciding on a venue for making their data available, scholars should consider multiple desiderata. These include: the practices and rules of the publishing venue, the transaction cost for the reader of accessing the evidence in context, the potential storing venue’s ability to make the data accessible to all interested persons, as well as to support annotation of citations (on which, more below), the likely durability of the venue (i.e., whether it has stable and long-term funding sources), the availability and quality of assistance with curation, and the cost to data users.⁹

Scholars who anticipate incurring incremental costs when preparing data for sharing (e.g., for anonymizing to protect confidential information) should consider building those costs into funding applications, and/or they may request reimbursement (perhaps

⁹ Although university repositories will often meet these criteria, scholars are discouraged from hosting data themselves on local websites as such sites are notoriously unreliable. While doing so may be a good temporary measure, longer-term plans for storage in an established repository should be developed.

drawn from fees paid by researchers requesting to use shared data). Likewise, when distribution involves additional costs (e.g., for administration of special conditions of access to confidential information), data distributors may request reimbursement for the incremental costs of making data available (see Section 6.5 of the *Ethics Guide*).

What is a “persistent identifier”? Why should I get one? Where can I get one?

A persistent identifier is a permanent link to a publication, data collection, or unique metadata instance that points to (and records versioning of) a data collection on the Internet. The publisher of the resource agrees to maintain the link to keep it active. Over time the link behind the persistent identifier may be updated, but the identifier itself remains stable. There are several kinds of persistent identifiers (DOI, URN, Handle, etc.).

Persistent identifiers are “machine-actionable” and facilitate the harvesting of data references for online citation databases, like the Thomson-Reuters Data Citation Index. Scholars can easily track the impact of their data from citations in publications. An increasing number of journals are requiring persistent identifiers for data citations.

Persistent identifiers can be useful for keeping track of bodies of data. One way to obtain a persistent identifier for data is to deposit them in an established institutional or social science repository, for instance, members of Data-PASS (<http://www.data-pass.org/>).

Production Transparency

In order to achieve *production transparency*, researchers should provide comprehensive documentation and descriptive metadata detailing their project’s empirical base, the context of data collection, and the procedures and protocols they used to access, select, collect, generate, and capture data. To offer three specific examples, authors should address basic issues of how documentary sources were selected or sampled, the terms under which interviews were granted, and how participant observation or ethnographic work was conducted.

Production transparency is a prerequisite for an author’s data to be intelligible to other researchers. Providing information about decisions made and processes carried out in the course of collecting and generating data, selecting them for inclusion in published work, and presenting them makes it easier for other scholars to understand and interpret the data; allows them to assess whether those processes were carried out in an unbiased manner; and helps them to evaluate the validity of the claims made on the basis of the data.

The production transparency requirement is triggered when scholars themselves collected or generated the data that support their evidence-based claims. Accordingly, the same timetable and constraints that apply to making those data available apply to

production transparency in relation to those data. As noted previously, APSA allows scholars a one-year period for first use of data they collected and thus for describing the data-collection process.

If the data are subject to ethical or legal restrictions, it is likely that production transparency will be similarly constrained. Conforming production transparency to relevant limits helps to ensure that other scholars can evaluate or replicate authors' data-collection procedures legally and without threatening the privacy of human subjects.

Although documentation is often supplied in text files or spreadsheets, an advanced standard for documenting data (at the study level) in the social sciences is the Data Documentation Initiative (DDI). DDI is an XML markup standard designed for social science data. Since DDI is machine actionable, it can be used to create custom codebooks and to enable online search tools. A list of tools for creating DDI is available at the DDI [Tools Registry](http://www.ddialliance.org/resources/tools) (<http://www.ddialliance.org/resources/tools>). Original documents (e.g., technical reports, questionnaires, and showcards) can be submitted as text files or PDF/A.

Analytic Transparency

Achieving *analytic transparency* requires scholars to describe relevant aspects of the overall research process, detail the micro-connections between their data and claims (i.e., show how the specific evidence they cite supports those claims), and discuss how evidence was aggregated to support claims.

The APSA standard for analytic transparency prescribes no epistemology or methodology; it simply requires that authors be clear about the analytic processes they followed to derive claims from their data, and demonstrate how they followed the general rules that attend the interpretive or inferential approach they are using.

The Transparency Appendix and Active Citation

One way in which qualitative researchers can provide data access, achieve production transparency, and engage in analytic transparency, is by developing a transparency appendix to their published work. A transparency appendix typically consists of two elements: active citations and an overview section.

Active citations follow the format of traditional footnotes or endnotes, but are digitally augmented to include:

- a precise and complete reference and any additional information that scholars will need to locate the cited source and find the relevant information within it;
- excerpts from cited sources;
- the cited sources themselves if the author possesses them and is in a position to share them, and/or hyperlinks thereto;

- annotations that
 - explain how individual pieces of data, sources, citations, and facts were interpreted and why they were interpreted as they were;
 - illustrate precisely how those individual pieces support claims in the text;¹⁰
 - address any important interpretive ambiguities or counter-arguments;
 - explain how individual pieces of data aggregate to support broad interpretative and theoretical conclusions.

Because active citations follow the format of traditional footnotes or endnotes, they are ideally suited to elucidate particular inferences or interpretations in the author's text. Certain aspects of research that should be explained if transparency is to be achieved, however, do not comfortably attach themselves to a particular subsection of text or footnote. These matters are instead best dealt with holistically. When such overarching concerns cannot be addressed in the main text, authors should include a brief "overview" in the transparency appendix clarifying their overall research trajectory (e.g., how interpretations and hypotheses were generated and evaluated); outlining the data-generation process; and demonstrating how the analysis attends to the inferential/interpretive rules and structures that underlie the type of analysis the author is doing.

Information provided in a transparency appendix *supplements* rather than replaces or repeats information offered in the text and footnotes of a book or article: it supplies additional context and background to authors' research efforts, offering an opportunity for authors to describe the rigor and thoroughness of their research (and field research), and allowing other scholars to understand and evaluate the appropriateness of their use (and, where relevant, generation) of data. What is "appropriate" depends upon the interpretive or inferential structures implied by the author's underlying epistemology and employed in the type of qualitative research he or she is conducting.

With respect to data access, scholars using active citation provide excerpts from the data sources underlying their claims (and ideally provide the actual data sources). In terms of production transparency, authors who cannot provide basic information about data collection in the main text of their publications due to length-limitations can include additional information in an introductory overview.

As for analytic transparency, the traditional representation in qualitative research—elaboration of an argument in the text combined with a simple citation—is often inadequate to make the link between an argument and evidence apparent. The critical element in the evidence is often difficult to discern, and the evidence is often interpretable in multiple ways. Likewise, a passage in a source can often only be properly interpreted within a broader textual context. Moreover, abbreviated ("scientific" or endnote) footnote formats, shrinking word limits for published work, and

¹⁰ The standards the author may seek to meet or rules he may follow when detailing these micro-connections can vary: they may include the scholar's interpretation of the relative persuasiveness and consistency of evidence, explicit process-tracing, discussing narrative plausibility, making arguments about the plausibility of counterfactuals, advancing a systematic scheme for weighting data, mixed method approaches, etc.. Analytic transparency requires only that scholars be consistent and transparent, so that the reader can follow how their overall conclusions follow from smaller-scale findings.

unfamiliarity with careful textual interpretation have rendered traditional journals (and even books) inhospitable forums for achieving rigorous analytic transparency.

In sum, the introductory overview component of a transparency appendix empowers authors to enhance readers' understanding of the context, design and conduct of research. Using active citation empowers authors to clarify the micro-connections between data, analysis, and conclusions. Both enhance the rigor and persuasiveness of qualitative research.

Publishers' Responsibilities

Journals, editors, and publishers should assist authors in complying with data access and research transparency guidelines.

Publishers should:

- inform authors of options for meeting data access and research transparency requirements;
- host scholars' cited sources and transparency appendices on line, or guide authors to online archives which will house these materials, and provide links from articles (at the level of the individual citation, if needed) to those materials;
- provide guidelines for bibliographic citation of data;
- include consistent and complete data citations in all publications.

Resources

- 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> (August 13, 2009).
- Corti, Louise. 2005. "Qualitative Archiving and Data Sharing: Extending the Reach and Impact of Qualitative Data." *IASSIST Quarterly* (Fall): 8-13.
- 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-%20ofinal.pdf> (August 13, 2009).
- UK Data Archive
 - [Managing and Sharing Data: A Best Practice Guide for Researchers](http://data-archive.ac.uk/media/2894/managingsharing.pdf) (<http://data-archive.ac.uk/media/2894/managingsharing.pdf>)
 - [Create and Manage Data](http://data-archive.ac.uk/create-manage) (<http://data-archive.ac.uk/create-manage>)
- UK Data Service
 - [Advice and Training](http://ukdataservice.ac.uk/use-data/advice.aspx) (<http://ukdataservice.ac.uk/use-data/advice.aspx>)
 - [Prepare and Manage Data](http://ukdataservice.ac.uk/manage-data.aspx) (<http://ukdataservice.ac.uk/manage-data.aspx>)

- Van den Eynden, Veerle and Louise Corti. 2009. "Tensions between data sharing and data protection in research with people." *SRA News* (May): 12-15.
- 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.

Appendix

Section 6 of the American Political Science Association's *Guide to Professional Ethics, Rights and Freedoms* as amended in October 2012:

"6. Researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated.

6.1 Data access: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they themselves generated or collected, researchers should provide access to those data or explain why they cannot.

6.2 Production transparency: Researchers providing access to data they themselves generated or collected, should offer a full account of the procedures used to collect or generate the data.

6.3 Analytic transparency: Researchers making evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data, i.e., clearly explicate the links connecting data to conclusions.

6.4 Scholars may be exempted from Data Access and Production Transparency in order to (A) address well-founded privacy and confidentiality concerns, including abiding by relevant human subjects regulation; and/or (B) comply with relevant and applicable laws, including copyright. Decisions to withhold data and a full account of the procedures used to collect or generate them should be made in good faith and on reasonable grounds. Researchers must, however, exercise appropriate restraint in making claims as to the confidential nature of their sources, and resolve all reasonable doubts in favor of full disclosure.

6.5 Dependent upon how and where data are stored, access may involve additional costs to the requesting researcher.

6.6 Researchers who collect or generate data have the right to use those data first. Hence, scholars may postpone data access and production transparency for one year after publication of evidence-based knowledge claims relying on those data, or such period as may be specified by (1) the journal or press publishing the claims, or (2) the funding agency supporting the research through which the data were generated or collected.

6.7 Nothing in this section shall require researchers to transfer ownership or other proprietary rights they may have.

6.8 As citizens, researchers have an obligation to cooperate with grand juries, other law enforcement agencies, and institutional officials. Conversely, researchers also have a professional duty not to divulge the identity of confidential sources of information or data developed in the course of research, whether to governmental or non-governmental officials or bodies, even though in the present state of American law they run the risk of suffering an applicable penalty.

6.9 Where evidence-based knowledge claims are challenged, those challenges are to be specific rather than generalized or vague. Challengers are themselves in the status of authors in connection with the statements that they make, and therefore bear the same responsibilities regarding data access, production transparency, and analytic transparency as other authors.”

APPENDIX B

DRAFT July 28, 2013

Guidelines for Data Access and Research Transparency for Quantitative Research in Political Science

The *APSA Guide to Professional Ethics, Rights and Freedoms* recognizes that

6. Researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated.

6.1 Data access: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they themselves generated or collected, researchers should provide access to those data or explain why they cannot.

6.2 Production transparency: Researchers providing access to data they themselves generated or collected, should offer a full account of the procedures used to collect or generate the data.

6.3 Analytic Transparency: Researchers making evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data, i.e., clearly explicate the links connecting data to conclusions.

Data Access, Production Transparency, and Analytic Transparency describe key stages of the research process. Data access is not sufficient without documentation of how data were prepared and how analysis was conducted. By meeting these requirements, researchers contribute to the credibility and legitimacy of Political Science.

While evidence comes in many forms, these guidelines refer primarily to numerical data that can be analyzed with quantitative and statistical methods.¹

Data Access

What data should be accessible to other scholars?

When an author makes evidence-based knowledge claims, all data required to replicate the results serving as evidence for statements and conclusions should be open to other scholars. Researchers who have generated or created their own data have an obligation to provide access to the data used in their analysis whenever possible. When the data were collected by others, an author is responsible for providing a clear path to the data through a full bibliographic

¹ A parallel set of guidelines is in preparation which are intended primarily for qualitative data.

DRAFT July 28, 2013

citation. In both cases, the steps involved in deriving conclusions and inferences from data should be fully described.

Researchers are strongly encouraged to share data beyond those required for replication of published findings. It is particularly important for researchers to provide access to data used in the process of generating conclusions but not included in the final analysis. More generally, providing as much access as possible to existing data can increase its value and often attracts greater attention to the work of the people who produced it.

When should data access be provided?

The *APSA Guide to Professional Ethics* recognizes that “Researchers who collect or generate data have the right to use those data first.” Data access should be provided no more than one year after public dissemination of evidence-based statements. Journals and funding agencies may have different requirements. Moreover, some funding agencies may require researchers to provide data access *prior to any publication*. Nothing in these guidelines should be read to contradict such requirements.

Where should data be made available?

Data should be made available online at an established repository or a website that can be discovered by standard Internet search engines. When deciding on a venue for making their data available, scholars should consider multiple desiderata, including the venue’s ability to make the data available to all interested persons, the likely durability of the venue (does it have stable and long-term funding sources), the availability of assistance with curation, and the cost to data users.

How should data be made available?

All data should be accompanied by:

1. Documentation describing the data in full
2. A complete citation including a “persistent identifier,” like “digital object identifiers” (DOIs).

Standard and non-proprietary formats are preferable, because they are more likely to remain accessible over time.

When distribution involves additional costs (e.g. for protection of confidential information), data distributors may request reimbursement for the incremental costs of making data available.

DRAFT July 28, 2013

How do I share data that includes confidential information?

As paragraph 6.4 of the *APSA Guide to Professional Ethics* notes, researchers may need to withhold access to data to protect subjects and comply with legal restrictions. However, secure methods of sharing confidential data are often available. When respondents might be re-identified by combining information in the data (e.g. age, sex, occupation, geographic location), a data use agreement specifying measures to protect confidential information can be required. Access may also be provided in a “data enclave,” where information derived from the data can be reviewed before it is released.

What if I used proprietary data?

When research is based on proprietary data, researchers should make available documentation that would allow other scholars to replicate their findings. Owners of proprietary data should be encouraged to provide access to all qualified researchers.

What is a “persistent identifier”? Why should I get one? Where can I get one?

A persistent identifier is a permanent link to a publication or a dataset on the Internet. The publisher of the resource agrees to maintain the link to keep it active. Over time the link behind the persistent identifier may be updated, but the identifier itself remains stable and does not change. There are several kinds of persistent identifiers (DOI, URN, Handle, etc.)

Persistent identifiers are “machine-actionable” and facilitate the harvesting of data references for online citations databases, like the Thomson-Reuters Data Citation Index. You will be able to easily track the impact of your data from citations in publications. An increasing number of journals are requiring persistent identifiers for data citations.

The best way to obtain a persistent identifier is to deposit your data in an established repository. Social science repositories, like the members of Data-PASS (<http://www.data-pass.org/>), and institutional repositories assign persistent identifiers to their holdings. There are also agencies that will issue a persistent identifier to a website that you maintain yourself.

What are the obligations of scholars who use data collected by others?

When the data were collected by others, an author is responsible for providing a full bibliographic citation in the same way that a publication or other scholarly product would be cited. Data citations should include author, title, date, and a persistent identifier (or other location information).

Production Transparency

Production transparency implies providing information about how original data were generated or collected, including a record of decisions the scholar made in the course of transforming their labor and capital into data points and similar recorded observations. In order for data to be understandable and effectively interpretable by other scholars, whether for replication or secondary analysis, they should be accompanied by comprehensive documentation and metadata detailing the context of data collection, and the processes employed to generate/collect the data. Production transparency should be thought of as a prerequisite for the content of one scholar's data to be truly accessible to other researchers.

What should documentation include about the overall research project?

- Principal Investigator
- Title
- Purpose of the study
- Scope of the study
- Study design
- Sample
- Mode of data collection
- Instruments used
- Weighting
- Response rates
- Funding source

What should the codebook provide about each variable?

- Variable description
- Instrument, question text, or computation formula
- Valid values and their meanings
- Cases to which this variable applies
- Methods for imputing missing values

How should I prepare documentation?

Although data producers often supply documentation in text files or spreadsheets, the standard for documentation in the social sciences is the [Data Documentation Initiative](#) (DDI). DDI is an XML markup standard designed for social science data. Since DDI is machine actionable, it can be used to create custom codebooks and to enable online search tools. A list of tools for creating DDI is available at the [DDI Tools Registry](#).

DRAFT July 28, 2013

Analytic Transparency

Scholars making evidence based knowledge claims should provide a full account of how they drew their conclusions, *clearly mapping the path from the data to the claims*. This path can be documented in many ways such as computer programs and scripts.

Researchers should make available materials sufficient to allow others to reproduce their results. For example, when providing computer programs to satisfy an analytic transparency requirement, questions about sufficiency can be answered as follows:

Is the program that produced my tables enough?

Transparency involves documenting all of the steps from the original data to the results supporting your conclusions.

I have lots of programs. Do I need to provide all of them?

The best practice is to consolidate all data transformation and analysis steps in a single program. Program steps may be developed separately, but they should operate as an integrated workflow.

Publisher's Responsibilities

Journals, editors, and publishers should assist authors in complying with data access and research transparency guidelines.

Publishers should

- inform authors of options for meeting data access and research transparency requirements;
- verify that data and program code are accessible, when appropriate;
- provide guidelines for bibliographic citations of data;
- include consistent and complete data citations in all publications.

Resources

- Australia National University - [ANU Data Management Manual, September 2010](#)
- Columbia University, Center for International Earth Science Information Network (CIESIN) - [Guide to Managing Geospatial Electronic Records, June 2005](#)
- Council of European Social Science Data Archives (CESSDA) - [Sharing Data Website](#)
- Gary King - [Data Sharing and Replication](#)
- Inter-university Consortium for Political and Social Research (ICPSR) - [Guide to Social Science Data Preparation and Archiving, 2009](#)
- UK Data Archive - [Managing and Sharing Data: A Best Practice Guide for Researchers, September 2009](#)
- UK Data Archive - [Create and Manage Data Website](#)