SYMPOSIUM

Data Access and Research Transparency in the Quantitative Tradition

Arthur Lupia, *University of Michigan*
George Alter, *University of Michigan*

The number of people conducting scientific analyses and the number of topics being studied are higher than ever.[1] At the same time, there are questions about the public value of social scientific endeavors, particularly of federally funded quantitative research (Prewitt 2013). In this article, we contend that data access and research transparency are essential to the public value of the enterprise as a whole and to the credibility of the growing number of individuals who conduct such research (also see Esterling 2013).

By quantitative research, we mean work that includes survey research, experiments, and mathematical and computerized models of dynamic processes. In this work, scholars convert attributes of observations and events into symbols. These symbols are joined with a grammar—typically a set of logical rules from mathematics or statistics—to form an inferential language. The resulting language of quantitative social science can produce a more precise description of concepts and relationships than ordinary language. Quantitative conclusions about causal relationships and existential propositions are often offered as direct numerical expressions or as exact functional forms.

With this precision, however, comes a potentially important limitation. Information can be lost when converting observations into symbols and interpreting these symbols via logical rules. When scholars fail to document—and make available to others—information about how they selected cases to study, particular attributes of the cases on which to focus, specific ways of converting these attributes into numbers, and the choice of certain types of mathematics or statistics to convert these numbers into knowledge claims, the meaning and value of quantitative social science knowledge claims becomes increasingly uncertain.

To more effectively and rigorously answer questions about the value of quantitative social science, it is imperative that those of us who conduct such research take actions that reinforce its credibility and make it easier for others to interpret our findings accurately. This means sharing our data whenever possible. It also means making available a complete description of the steps that we used to convert data about the social world into quantitative claims about how it does and does not work. Such commitments will not only help others more accurately assess our claims about individual events but also increase the extent to which others will view as credible our attempts to draw generalizations about people, policies, and institutions from a series of numerical simplifications and logical transformations.

With such imperatives in mind, political science, like other disciplines, is seeking to increase its credibility by improving procedural transparency and data sharing. Supporting efforts to increase data sharing and research transparency is the fact that technical barriers to such openness are falling quickly. Data archives, for example, are becoming more numerous and archivists have found multiple ways to make them easier to use. Viewed from a technical perspective, depositing one's own data and documents and accessing others data and documentation has never been easier. Old explanations for why scholars need not share data or procedural information are becoming more difficult to support. Indeed, younger generations, who do not remember life before the Internet, expect greater transparency of all kinds (Pew Research Center 2013). At the same time, we recognize that there have often been few incentives for taking the time to document one's procedures or to share one's data. Unless greater incentives for sharing data and publicizing analytic procedures emerge, it is difficult to expect old patterns to change.

In what follows, we describe current and future activities that support greater data sharing and research transparency. We focus in particular on several efforts to make data sharing and research transparency more rewarding for individual investigators and larger research collaboratives. We contend that these and other efforts can help individual political scientists, and the discipline, more effectively demonstrate the evidentiary and intellectual foundations of their insights. In so doing, this new emphasis on clarifying the evidentiary and logical foundations of one's knowledge claims can increase the credibility of individual research projects, reduce uncertainty about the meaning of social scientific findings, and increase the value of quantitative social science to multiple constituencies.

DATA SHARING

Many scientific disciplines are having broad conversations about data sharing. While there is much interest in the topic among political scientists today, leading figures in our discipline have long been engaged this topic. Warren Miller, a principal architect of the American National Election Studies (ANES), also founded the Inter-university Consortium for Political and Social Research (ICPSR), the world's largest social

doi:10.1017/S1049096513001728

science data archive. When seeking funding for data sharing, he (1962) argued that: "The sporadic work of individuals with diverse interests and little or no association with the work of predecessors or contemporary colleagues often produces worthwhile results but through the Consortium the power of extended and cumulative programs of research can also be realized." Miller also provided an example for others to follow by sharing ANES data from early in the project's inception. Today, thanks to the work of Miller and others, many data archives exist and the emergence of the Internet makes accessing and sharing data more feasible to more people.

Despite these changes, data sharing remains a contentious issue. Despite widespread acceptance of replication as a nominal virtue in quantitative social science, as Hamlet might say, "it is a custom more honour'd in the breach than the observance." Many scholars do not share their data (Acord and Harley 2012). Nonsharing scholars sometimes describe the work that they have put into collecting such data and opine that sharing data limits their ability to claim credit for their labor. Other scholars face confidentiality issues and believe that sharing data would open them to legal risk or minimize their opportunities for future data gathering. Yet other scholars want to share their data but do not know how to do so, or begrudge the time required to prepare data for sharing. In the remainder of this section, we address each of the barriers to data sharing in turn.

The first barrier pertains to claiming credit. Scholars who have invested a great deal of time and capital into collecting data want to reap benefits from their research. Some researchers resent the possibility that they may be "scooped" or even contradicted with their own data.

Scientific organizations are developing and implementing a number of creative ways to manage such difficulties. Underlying these endeavors are two premises. The first premise is that data that are shared widely can produce a larger number of meaningful inferences. Hence, sharing data is a public good that should be rewarded. The second premise is that data access facilitates replication that can be used to evaluate the truth-value and meaning of empirical claims. To this end, a growing number of journals are requiring authors who make empirical claims to make available the data from which such claims were derived.

At the same time, many professional associations, funding agencies, and journals acknowledge the need to balance the incentives of data collectors with the benefits to science and society from data sharing. One way that such a balance is sought is to offer a reasonable time to analyze data before making it available to others. Among social science journals, the most common requirement is for data to be available at the time of publication. Other journals and professional associations allow an additional "embargo period" after publication. Because the time from submission to publication is usually more than a year and often much longer and since the time from collection to submission also entails periods of multiple months or years, either rule provides several years for authors to exploit their data.

Moreover, journals typically require only the data needed to support empirical claims in a specific publication, so-called "replication data." At present, authors are not expected to release aspects of a data set not supporting a publication.[2] Such practices represent a compromise between the opportunities for data collectors to reap the fruits of their labor and the broader benefits that can come from making that data available to others.

Another means of providing credit to scholars who create and share valuable data is to develop better data citation practices (Altman and King 2007). For centuries, scholars have been expected to cite the evidence, theories, and conclusions on which their own research builds. Notwithstanding this long tradition, social science editors and publishers have been slow to recognize that data are intellectual products for which citation should be required (Mooney 2011). This pattern is changing. Numerous organizations actively promote data citation (see DataCite 2013; Data Preservation Alliance for the Social Sciences 2013; International Association for Social Science Information Services and Technology 2013; International Council for Science: Committee on Data for Science and Technology 2013; Inter-university Consortium for Political and Social Research 2013).

For centuries, scholars have been expected to cite the evidence, theories, and conclusions on which their own research builds. Notwithstanding this long tradition, social science editors and publishers have been slow to recognize that data are intellectual products for which citation should be required (Mooney 2011).

Funding agencies are also using data citation as a way to demonstrate the impact of their data collection investments. For example, in 2012, the ANES began to require people who download data from its website to sign an agreement to formally cite the ANES in any and all intellectual products they derive from its data. ICPSR has also expanded its ability to facilitate citation (http://www.icpsr.umich.edu/icpsrweb/ICPSR/citations/index.jsp). In 2012, Thomson Reuters unveiled a citation index for data, as a new feature in its Web of Knowledge platform (http://wokinfo.com/products_tools/multidisciplinary/dci/). With improvements in data citation, scholars who produce data that is used by others can expect to be cited for that work. Given the value that citation counts have for tenure and promotion reviews, better data citation norms should increase individual incentives to share data.

To this end, it should also be noted that a growing number of inquiries into citation counts find that "papers with publicly available datasets receive a higher number of citations than

similar studies without available data" (Piwowar and Vision 2013, 1). While we are not aware of such comprehensive studies being conducted in the social sciences, a study of more than 10,000 published studies in the gene expression literature funds that "studies that made data available in a public repository received 9% (95% confidence interval: 5% to 13%) more citations than similar studies for which data was not made available" (Piwowar and Vision 2013, 7). Dorch (2012) finds larger effects, albeit on a smaller sample of astrophysics articles.

The second barrier points to human subject harms and legal risks associated with some forms of data sharing. Protecting the privacy of research subjects is of paramount importance to the scientific community. Violations of confidentiality can damage public confidence in research and undermine the cooperation of subjects on which researchers depend. So the challenge becomes how to protect those rights while accruing, as much as possible, the individual and social benefits of data sharing.

The process of reconciling subject protection with data sharing begins with informed consent. Consent forms should promise to protect confidential information to the maximum extent allowed by law, but they need not exclude sharing data with other researchers.[3] After data are collected, a variety of techniques can minimize disclosure risks. Data "masking," for example, refers to techniques that modify data to prevent subject identification (see, e.g., Rubin 1993). Masking procedures include anatomizing (Xiao and Tao 2006), permuting (Zhang et al. 2007) or perturbing (Adam and Worthmann 1989) cells in a data matrix in ways that preserve the aggregate properties in which analysts are interested while decoupling identifying information from the data. When implemented successfully, these techniques allow analyses to be identical to what they would have been had individual cells not been altered. The ability to achieve such outcomes depends on relationships among properties of the data, how the cells are perturbed, and the kinds of analyses that individuals want to run. In a dataset with hundreds of variables, for example, it is typically impossible to implement a masking algorithm in ways that preserve all possible statistical relationships among all variables. If, however, a relatively limited set of relationships are of interest, successful masking possibilities emerge (Rubin 1993). Such techniques are expanding circumstances in which data can be shared while simultaneously protecting the privacy of individual respondents (Fung et al. 2010).

Other subject protection measures are also available (National Research Council 2003, 2005). Researchers can be required to sign data use agreements in which they provide detailed plans for data security and other measures to prevent disclosure of confidential information. Highly sensitive data can be shared in controlled environments, like the Research Data Centers operated by the Census Bureau, where outputs can be screened for disclosure risks. Indeed, many organizations now operate remote execution systems or "virtual data enclaves," which allow researchers to conduct such analyses without having direct data access.

Although a wide range of measures have been developed to protect confidential information from research subjects, restrictions imposed by private organizations are a growing concern. Many corporations are amassing vast quantities of data. Because these data are considered commercial assets, researchers are often required to sign nondisclosure agreements that prevent them from sharing the data with others. The *American Economic Review* exempts authors from sharing proprietary data but asks them to inform others how the data may be obtained (American Economic Association 2013). Political science journals should consider analogous policies.

With improvements in data citation, scholars who produce data that is used by others can expect to be cited for that work. Given the value that citation counts have for tenure and promotion reviews, better data citation norms should increase individual incentives to share data.

Regarding the third concern, planning and good data management practices can reduce burdens often associated with data sharing. Many universities employ "data librarians" who offer assistance with data management planning. Data archives, including the six partners in the Data Preservation Alliance for the Social Sciences (http://www.data-pass.org/), are also available to offer advice and assistance to a wide range of scholars (see also ICPSR 2012). Professional archives can also help scholars document their work so that it remains accessible and functional for scholars who seek accurate interpretations of shared data. Such practices can also benefit data producers because well-designed documentation of data and research "workflow" can reduce the time needed to respond when a journal issues a "revise and resubmit" (Long 2009).

Changes in data citation practices, the possibility that articles associated with data sharing are more often cited, statistical masking, and professional archiving services are all factors that make data sharing more rewarding and feasible. To the extent that scholars come to formally cite all data that they use, quantitative social scientists will not only find data sharing more rewarding, but they will also be able to benefit from the data that others are now sharing. If scholars further commit to making data accessible and following the best practices of professional archivists, they and others can benefit for years to come from data that has already been created.

RESEARCH TRANSPARENCY

Sharing data does not provide all of the information about a quantitative analysis that can advance science and benefit soci-

ety. When assessing the meaning of quantitative claims, audiences often want to understand the decisions and actions that produced the claim. This follows because the meaning of a conclusion depends on the premises and practices from which it was derived.

In 2012, the APSA responded to calls for greater transparency by revising its "Guide to Professional Ethics, Rights, and Freedoms." The guide now states 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 distinction of production and analytic transparency in the revision follows from definitions of Lupia and Elman (2010).

> Production transparency implies providing information about how the 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. Analytic transparency is a separate but closely associated concept. Scholars making evidence-based knowledge claims should provide a full account of how they drew their conclusions, clearly mapping the path on the data to the claims.

Now that the discipline is highlighting research transparency as a core ethical obligation for political scientists, the challenge is to help the scholarly community develop incentives and utilities that make research transparency more feasible and rewarding.

The work necessary to follow the guidelines will vary for different quantitative communities. In some areas of quantitative political science, providing such information is already the norm. Among game theorists, for example, formal proofs detail nonobvious relationships between premises and conclusions. Proofs, in this context, are like a computer code that others can use to verify that specific conditions produce specific conclusions. In game-theoretic research communities, nonobvious claims that lack proofs are not considered credible.

In other fields, the documentation and sharing of "do-files" or "code" is less common. Consider, for example, survey research. For survey producers, procedural transparency entails descriptions of case selection, question selection, interviewer selection, interviewer training, and strategies for managing nonresponse. Each of these decisions can affect the meaning of specific data points as well as the aggregate conclusions drawn from survey data. For survey analysts, transparency includes descriptions of how variables were coded and how analysts chose among different inferential methods and model specifications.

Today, information on the data production and analytic decisions that underlie many published works in political science is unavailable. This is one reason that many graduate courses in political science are unsuccessful in their attempts to replicate published empirical claims. Even when students have access to the same data as the original researchers (e.g., the same version of the ANES), they have not always had access to how data producers and analysts collected data, created variables, or knowledge of the exact code (i.e., statistical model) that produced published findings. When this material is not available, replication is undermined as is other scholars' ability to evaluate what a quantitative empirical claim actually means. A website called Political Science Replication now collects such instances and, in so doing, reveals many difficulties associated with contemporary replication attempts.[4]

Despite the disappointing record for data sharing in some quantitative communities, promising signs indicate that research transparency is being taken more seriously in important areas of political science.

Despite the disappointing record for data sharing in some quantitative communities, promising signs indicate that research transparency is being taken more seriously in important areas of political science. Archiving of procedural materials, research design registries, and revised data citation practices are three ideas that political scientists are pursuing to make research transparency more rewarding and feasible. We describe each of these ideas in turn.

Of the three ideas, archiving of procedural materials is farthest along. Entities such as ICPSR, Dataverse, and the Open Data Project provide means for scholars to share not only their data, but also supplementary materials that allow others to replicate existing findings. Among survey providers, the ANES (production) has provided unprecedented documentation of this kind. For its 2008 studies, the ANES produced dozens of reports on many steps of its data production processes. Its Online Commons provides histories of the evaluative procedures that the ANES used to choose which of more than 3,000 proposed questions to include on its surveys (Aldrich and McGraw 2011). The site also describes many ways in which questions were evaluated including alternate weighting algorithms (DeBell and Krosnick 2009), and how it developed new code frameworks for open-ended responses (Berent, Krosnick, and Lupia 2013).

A second idea that is growing in popularity is requiring research designs to be registered before rather than after such research is conducted (Humphreys et al. 2013). A benefit of registries for researchers is that it allows them to lay claim to

a set of procedures. To see why this is valuable, note that in much empirical research today, only the final version of a multistep analysis is published. This final version is often influenced by well-known publication biases. Because many journals are hesitant to publish null results, scholars tend to send journals only analyses that produce statistically significant findings. Patterns in published articles (Gerber, Green, and Nickerson 2001) suggest that scholars suppress analyses that do not feature significant coefficients. As King (1986) and Lupia (2008) have written, such incentives may lead scholars to engage in "stargazing," the practice of running data through different model specifications until finding a specification that produces statistically significant results on the variables that the scholar wanted to feature. Stargazing is a problem for many reasons, not the least of which is that stargazing can cause the standard errors underlying common measures of significance to lose the attributes that make statistical significance meaningful (Rubin 2007). So, when scholars show only "significant" results, it can be impossible for readers to determine whether the results have the substantive meaning that the authors claim.

Research registries, by contrast, allow scholars to document practices and findings at many stages of a research agenda. Researchers can post experimental designs, regression models, computer simulation programs, or a list of hypotheses in a registry. Readers of an article or book can then use the registry to determine whether a result is a true characterization of a focal social relationship or whether it is the product of publication-related biases that have no clear theoretical foundation. In other words, they can see whether published designs, models, programs, and hypotheses represent a larger set of inquiries or are cherry-picked because they produce a desired finding. While registries run the risk of embarrassing researchers who are reticent to reveal that they did not derive the best solution to a problem on their first try, such outcomes are a public good. Scholars, particularly those who eventually succeed in discovering important relationships, can help others advance research more quickly by revealing initial and seemingly sensible specifications that turned out to be suboptimal.

The third idea is oriented toward making research transparency more rewarding. Scholars may ask why they should allow others to reap the benefits of research agendas or analytic strategies to which they devoted substantial time and effort. As was the case for scholars who work hard to accumulate data, there are limited professional incentives to share one's "code." To make transparency more rewarding some scholars have proposed revising citation practices. In addition to data citation practices described in the previous section, scholars are also pursuing "code" citation practices. In computer science, for example, many people recognize the value of code. If scholars expected one another to cite their code, there would be greater incentives to make such code available to others.

A complementary endeavor is the Open Science Collaboration's "badge" system (https://opensciencefframework.org/project/TVyXZ/wiki/home/). This endeavor allows organizations to award scholars "badges" for "open data," "open materials" (e.g., for providing information about case-selection procedures and "do-files"), and "preregistration." Given the increased attention to such matters in recent years, it seems likely that many scholars will want to attach such labels to their work. Such practices can make research transparency more rewarding for individual investigators while also offering credibility benefits to research communities.

CONCLUSION

This article details the value of increased data sharing and research transparency to quantitative social science and factors that affect incentives for quantitative researchers to engage in such practices. An increasing population of scholars is recognizing the link between sharing, transparency, and the ability to evaluate scholarly claims. Regardless of the exact process by which change occurs, scholarly decisions to share data and information about the procedures that produced their conclusions are critical to the future of quantitative social science. Social scientists who commit to sharing their data and code give broader populations a basis for treating their work as an endeavor that is valuable, credible, and worth supporting. ■

Social scientists who commit to sharing their data and code give broader populations a basis for treating their work as an endeavor that is valuable and worth supporting.

NOTES

1. See, for example, Larsen and von Ins (2010) for a comprehensive review of trends in scientific publication and citation broken down by scientific area. See Appendix B-1 of Chiswick, Larsen, and Pieper (2010) for a report on steady growth in the number of social science PhDs granted in the United States and Canada from 1966 to 2006, Ware and Mabe (2009, 5) for statistics on the steady growth of scientific journals and output and (2009, 56) for a chart on journal article use over time.
2. In rare cases an editor may request data during the review process. Data associated with a manuscript under review should be covered by the same expectation of confidentiality that applies to the manuscript itself. If the manuscript is not accepted for publication, the review should not compromise the author's exclusive access to the data.
3. NIH advises: "In preparing and submitting a data-sharing plan during the application process, investigators should avoid developing or relying on consent processes that promise research participants not to share data with other researchers. Such promises should not be made routinely or without adequate justification described in the data-sharing plan" (Office of Extramural Research. National Institutes of Health 2004; see also Inter-university Consortium for Political and Social Research 2012, 13).
4. http://politicalsciencereplication.wordpress.com/. Accessed on September 20, 2013.

REFERENCES

Acord, Sophia Krzys, and Diane Harley. 2012. "Credit, Time, and Personality: The Human Challenges to Sharing Scholarly Work Using Web 2.0." *New Media and Society* 15: 379–97.

Adam, Nabil R., and John C. Worthmann. 1989. "Security Control Methods for Statistical Databases." *Association for Computing Machinery: Computing Surveys* 21 (4): 515–56.

Aldrich, John H., and Kathleen McGraw. 2011. *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*. Princeton, NJ: Princeton University Press.

Altman, Micah, and Gary King. 2007. "A Proposed Standard for the Scholarly Citation of Quantitative Data." *D-lib Magazine* 13 (3/4). Cited on September 20, 2013 from http://www.dlib.org/dlib/march07/altman/03altman.html.

American Economic Association. 2013. "American Economic Review: Data Availability Policy 2013." Cited June 18, 2013. Available from http://www.aeaweb.org/aer/data.php.

American Political Science Association. 2012. "Proposed Changes to Ethics Guide." Cited September 20, 2013 from http://www.apsanet.org/content_2483.cfm.

Berent, Matthew, Jon A. Krosnick, and Arthur Lupia. 2013. *Coding Open-ended Answers to Office Recognition Questions from the 2008 ANES Time Series Interviews*. Cited September 20, 2013 from http://www.electionstudies.org/studypages/2008prepost/ANES2008TS_CodingProject.htm.

Chiswick, Barry R., Nicholas Larsen, and Paul Pieper. 2010. "The Production of PhDs in the US and Canada." Institute for the Study of Labor (IZA) Discussion Paper 5367. Cited September 20, 2013 from http://ftp.iza.org/dp5367.pdf.

DataCite. 2013. "Why Cite Data?" Cited June 15, 2013 from http://www.datacite.org/whycitedata.

Data Preservation Alliance for the Social Sciences. 2013. "Data Citations 2013." Cited June 15, 2013 from http://www.data-pass.org/citations.html.

DeBell, Matthew, and Jon A. Krosnick. 2009. *Computing Weights for American National Election Study Survey Data*. ANES Technical Report Series, No. nes012427.

Dorch, Bertil. 2012. "On the Citation Advantage of Linking to Data." Cited on September 21, 2013 from http://hprints.org/docs/00/71/47/34/PDF/Dorch_2012a.pdf.

Esterling, Kevin. 2013. "Transparency-Inducing Institutions and Legitimacy." Berkeley Initiative for Transparency in the Social Sciences. Cited on June 7, 2013 from http://cegablog.org/2013/03/20/tss_esterling/.

Fung, Benjamin C. M., Ke Wang, Rui Chen, and Philip S. Yu. 2010. "Privacy-Preserving Data Publishing: A Survey of Recent Developments." *Association for Computing Machinery: Computing Surveys* 42 (4), Article 14 53 pages. DOI=10.1145/1749603.1749605 http://doi.acm.org/10.1145/1749603.1749605.

Gerber, Alan S., Donald P. Green, and David Nickerson. 2001. "Testing for Publication Bias in Political Science." *Political Analysis* 9: 385–92.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21: 1–20.

International Association for Social Science Information Services and Technology. 2013. "Quick Guide to Data Citation." Cited June 15, 2013 from http://iassistdata.org/sites/default/files/iassist-data-citation-quick-guide.pdf.

International Council for Science: Committee on Data for Science and Technology. 2013. "CODATA Data Citation Standards and Practices Task Group." Cited June 15, 2013 from http://www.codata.org/taskgroups/TGdatacitation/.

Inter-university Consortium for Political and Social Research. 2012. "Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle." Ann Arbor, MI.

Inter-university Consortium for Political and Social Research. 2013. "Data Citations." Cited June 15, 2013 from http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp.

King, Gary. 1986. "How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30: 666–87.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Larsen, Peder Olesen, and Markus von Ins. 2010. "The Rate of Growth in Scientific Publication and the Decline in Coverage Provided by Science Citation Index." *Scientometrics* 84: 575–603.

Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.

Lupia, Arthur. 2008. "Procedural Transparency and the Credibility of Election Surveys." *Electoral Studies* 27: 732–39.

Lupia, Arthur, and Colin Elman. 2010. "Memorandum on Increasing Data Access and Research Transparency (DA-RT)." Submitted to the Council of the American Political Science Association, September.

Miller, Warren. 1962. Letter to Rensis Likert about the Establishment of the Interuniversity Consortium for Political and Social Research. Downloaded on June 6, 2013 from http://www.icpsr.umich.edu/files/ICPSR/fifty/Miller-Likert.pdf.

Mooney, H. 2011. "Citing Data Sources in the Social Sciences: Do Authors Do It?" *Learned Publishing* 24 (2): 99–108.

National Research Council. 2003. *Protecting Participants and Facilitating Social and Behavioral Sciences Research*. Washington, DC: National Academies Press.

National Research Council. 2005. *Expanding Access to Research Data: Reconciling Risks and Opportunities*. Washington, DC: National Academies Press.

Office of Extramural Research. National Institutes of Health. 2004. "Frequently Asked Questions: Data Sharing, February 16, 2004." Cited June 18, 2013 from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm#907.

Pew Research Center. 2013. "Public Split over Impact of NSA Leak, But Most Want Snowden Prosecuted." Cited on September 20, 2013 from http://www.people-press.org/2013/06/17/public-split-over-impact-of-nsa-leak-but-most-want-snowden-prosecuted/.

Piwowar, Heather, and Todd J. Vision. 2013. "Data Reuse and the Open Data Citation Advantage." *PeerJ PrePrints*. Published: 4 Apr 2013, doi: 10.7287/peerj.preprints.1.

Prewitt, Kenneth. 2013. "Is Any Science Safe?" *Science* 340 (6132): 525.

Rubin, Donald B. 1993. "Comment on "Statistical Disclosure Limitation."" *Journal of Official Statistics* 9: 461–68.

Rubin, Donald B. 2007. "The Design *versus* the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials." *Statistics in Medicine* 26: 20–36.

Ware, Mark, and Michael Mabe. 2009. "The STM Report: An Overview of Scientific and Scholarly Journal Publishing." International Association of Scientific, Technical, and Medical Publishers (Oxford, UK). Cited on September 20, 2013 from http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf.

Xiao, Xiaokui, and Yufei Tao. 2006. "Personalized privacy preservation." In *Proceedings of the ACM SIGMOD Conference*. Association for Computing Machinery, New York.

Zhang, Qing, Nick Koudas, Divesh Srivastava, and Ting Yu. 2007. "Aggregate Query Answering on Anonymized Tables. In *Proceedings of the 23rd IEEE International Conference on Data Engineering*.