

Data Sharing Policies in Social Sciences Academic Journals: Evolving Expectations of Data Sharing as a Form of Scholarly Communication

Joel Herndon, Ph.D. and Robert O'Reilly, Ph.D.

DISCUSSIONS ABOUT DATA sharing and data availability are widespread in the public and private sectors and within academia. For academic researchers, these discussions often focus on how data sharing and transparency can increase the rigor and credibility of scholarship. In response to these discussions, many academics have argued that academic journals should adopt sharing/replication policies to address these concerns. In this chapter, we analyze whether academic journals in the social sciences have been adopting such policies over time and whether such adoption varies by field.[†] Our conclusion is that while journals are indeed adopting such policies, they are doing so in an incomplete and varied manner. Moreover, the adoption of such policies and the debates that surround them are matters with which libraries should engage themselves, as doing so has the potential to strengthen connections between libraries and researchers with regard to data services and support.

Background and Context

It is now quite common to see stories and commentary about growth in the availability of digital data and about the various implications of that growth. In 2012,

[†] Data and syntax that produce the tables in this chapter are available at <http://dx.doi.org/10.15139/S3/12157>.

the *New York Times* announced that we have entered “The Age of Big Data,” one in which the amount of digital information being accumulated is growing at a non-linear rate.¹ A few years earlier, *The Economist* magazine made similar points in its cover story on what it called the “Data Deluge”: the rapidly-increasing quantities of data that are being created, the implications of those data for both the public and private sectors, and the challenges involved in storing, managing, and making effective use of those data.²

While there are many facets to the “Data Deluge,” one of the more salient issues in academia is the concept of “open data”: the idea, and even expectation, that data gathered and assembled for research purposes should be readily-accessible to the public, especially when such undertakings receive support from the public sector. In the United States, the “Shelby Amendment” to the Omnibus Appropriations Act for FY1999 required the Office of Management and Budget to ensure that all data gathered via grants from federal agencies would be available through the Freedom of Information Act.³ Since 2003, the National Institutes of Health has required grant applications with costs greater than \$500,000 in any single year to include data sharing plans as part of those applications.⁴ Since 2011, the various directorates within the National Science Foundation have required all grant applications to include data management plans that spell out what data will be generated and how those data will be managed and shared both during and after the duration of any grant.⁵ In a similar vein, the Office of Science and Technology Policy announced in 2013 that agencies with research and development budgets greater than \$100,000,000 are required to develop plans for greater public sharing of funded research, including data collected from or for that research.⁶

Nor is this trend confined to the United States. In the United Kingdom, governments have been increasing the availability of government-gathered data.⁷ There are now numerous “open data sites”⁸ operated by national statistical offices. At an international level, the OECD Committee for Scientific and Technical Policy issued a communiqué in 2004 encouraging increased public access to research data, and the OECD followed that statement up by publishing a set of “principles and guides for access to research data from public funding.”⁹ The European Union has similarly endorsed principles of open access to research data to encourage greater use and re-use of such resources.¹⁰

These trends have not gone unnoticed by academics, nor have they gone unnoticed by academic libraries. Gary King has written extensively about the opportunities and challenges presented by the growth of potentially-available data, noting that the social sciences “appear to be in the midst of a massive collision between unprecedented increases in data production and availability about individuals and the privacy rights of human beings worldwide, most of whom are also effectively research subjects.”¹¹ Academic libraries, meanwhile, have discussed and debated how to respond to such developments, with many of them now offering

or contemplating offering research data management as a new service, such as by assisting grant applicants in crafting language for inclusion in data management plans.¹²

However, it should also be noted that calls and demands for greater access to data and related discussions of the potential impacts and benefits of such access are not anything new. Long before the media was awash with references to “big data” and long before libraries were deliberating over how to address funder mandates for data sharing plans, academics in the social sciences were deliberating over whether scholars should be more open and sharing with data used in their research. Likewise, academics have engaged in related debates for several decades over whether academic journals should have policies that require scholars to make data available as part of, or even as a condition for, publication.

Data Sharing, Transparency, and Replication

The Benefits of Reuse

Why do various academics value and encourage data sharing as a behavior in which researchers should engage? Likewise, why do they advocate for policies that require authors to make data available? In a nutshell, many of the offered reasons are variations on a theme stated by the OECD: “the value of data lies in their use,” as greater availability of research data provides various positive externalities in the form of benefits that accrue to parties other than those who did the original data collection.¹³ Shared data provide opportunities for reuse outside of the context in which they were collected and allows other researchers opportunities to combine data collections to address novel questions that would be otherwise be difficult to empirically evaluate. For example, data initially collected to study ethnic conflict could later be repurposed to test for the effects of property rights on economic growth.¹⁴ Greater reuse of research data also reduces unnecessary duplication of effort, as scholars do not need to expend time collecting data that others have already gathered if they can access those data. Gleditsch and Metelits and others argue that sharing research data offers benefits in the form of greater visibility and more citations to those scholars who do so.¹⁵ Jeremy Freese, meanwhile, notes that the common practice of selectively sharing data amongst colleagues and within informal networks creates asymmetries of access that disadvantage graduate students and other less-established researchers. Greater sharing of data thus provides for more egalitarian levels of data access.¹⁶

The Benefits of Transparency and Replication

The main reasons for and benefits from greater sharing of research data emphasized by its advocates, however, relate to matters of transparency in research practices and replication of research findings. In a nutshell, data sharing promotes transparency in research and makes it easier for others to evaluate and assess empirical scholarship. When data used in a given work are readily available, other scholars can see how the data were assembled and used, the choices and assumptions that went into their usage, and how the analyses were conducted. Ideally, shared data allow the research community to determine and assess the entire sequence of steps from the collection of the raw/original data to the empirical findings of the analysis. In being able to do so, researchers can uncover any flaws or weaknesses in the analysis, thus improving the likelihood that errors are both identified and shared with the larger academic community while simultaneously offering new opportunities to extend and expand promising research. By sharing data and conducting research in a more transparent manner, academics make their work more accessible and their findings more credible—their work is easier to evaluate and engage with, and their findings have additional legitimacy precisely because it is easier to determine how those findings were produced.¹⁷

Greater sharing of data also enhances the scientific rigor of research. For research to be “scientific,” it must, in principle, be *replicable*. Other scholars should be able to take the data used in a publication and reproduce the results using the methods described by the authors. If the social sciences are to be truly scientific in terms of applying the scientific method to the explanation of social phenomena, empirical research in the social sciences must likewise be reproducible by others. King argues that replication of existing findings is “the most common and scientifically productive method of building on existing research”¹⁸ while Anderson et al. note that, “in theory, economic research is scientific research—that is, economic theory suggests empirically falsifiable hypotheses and investigators seek to confront those hypotheses with data.”¹⁹ Similarly, they argue that “empirical economics, by actively discouraging replication, does not incorporate the self-correcting mechanism of the scientific method.”²⁰ Data sharing, then, is a means of promoting replication and thus making research more scientific. As Lupia and Elman put it, “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.”²¹ Or, as Ben Bernanke (yes, **that** Ben Bernanke) notes, “Replicability is essential if empirical findings are to be credible and usable as a starting point for other researchers.”²²

We should note here that these discussions and debates are not merely theoretical. Instead, they have been inspired in part by recurring questions about the extent to which empirical research in the social sciences can be replicated and

about the research practices of academics in various fields. Within political science, for instance, the Data Access and Research Transparency initiative (DA-RT)[†] grew in part out of a “growing concern that scholars could not replicate a significant number of empirical claims that were being made in the discipline’s top journals.”²³ Dafoe similarly notes that “it is not uncommon for the key results of scientific research to be non-reproducible or to arise from errors.”²⁴ Eubank undertook an attempt to replicate twenty-four empirical papers submitted to the *Quarterly Journal of Political Science* and found that “13 (54 percent) had results in the paper that differed from those generated by the author’s own code.”²⁵ Numerous attempts to replicate articles in various top-ranked economics journals have resulted in similar conclusions: the empirical findings in such works are often—even usually—difficult to successfully reproduce.²⁶ Within the field of psychology, debates about replication have been motivated by cases of outright fabrication of data.²⁷ More generally, some academics have suggested that most empirical findings in academic research are effectively false.²⁸ Not just an esoteric debate amongst academics, this matter has even received attention from the mass media.²⁹

The increased usage of data harvested from social media has further raised questions about replication, given how highly asymmetrical access to such data can be.³⁰ A particularly vivid instance of how replication can uncover non-trivial issues with research involves the case of an article by Harvard economists Carmen Reinhart and Kenneth Rogoff on the relationship between public-sector debt and economic growth. The article attracted considerable attention because its findings suggested that public-sector debt has a marked, negative effect on economic growth when debt levels rise above 90% of GDP. A replication of their study by graduate students at the University of Massachusetts, however, concluded that their findings were a function of decisions on how to aggregate observations within and across countries and of spreadsheet errors in the underlying data. The macroeconomic policy implications of their original findings were such that this specific case generated a great deal of commentary amongst academics and the media about replication and transparency in general, and about the usage of Microsoft Excel in particular and its suitability (or lack thereof) for producing robust, replicable analyses.³¹

Implications for Scholarly Communication —The Role of Journal Policies

Given the potential benefits of greater sharing of research data, the question then becomes one of how such sharing can be encouraged. Toward this end, many advocates of data sharing argue that one effective way to promote greater data availability is for academic journals to adopt policies that expect or require authors

† See Elman and Lupia, “The (DA-RT) Data Access and Research Transparency Joint Statement” for additional information about this initiative and its origins.

of submitted manuscripts to provide data used to produce the results in those manuscripts as part of the publication process. As Stodden, Guo, and Ma put it, “the journal publication process is a key lever shaping the nature of scholarly communication and promoting the integrity of the scholarly record.”³² Consequently, the presence of journal policies that require data sharing as a condition of publication provide a strong incentive for authors who wish to publish to make their data publicly available. Anderson, et al, succinctly summarize the role that journals can play here:

As a form of collective action, journals can assure that published research is sufficiently documented—including data and program archives—and that published articles’ results are replicable. Absent such a journal structure, both economic models of rational choice—and the evidence—suggests that most authors will not voluntarily choose to incur the costs of creating archival files and documentation.³³

Journal policies, then, are a “key lever” that can be used to pry open data resources that might otherwise remain closed, and to enhance the methodological soundness of quantitative research by encouraging academics to be more thorough in their work.³⁴

Data Sharing—What Is It?

Such discussions of data sharing and its benefits and how journals might encourage it in turn raise the questions of what data sharing actually entails and what forms data-sharing policies should take. Early statements on this topic from researchers such as Robert Hauser and Gary King mainly focus on providing the data themselves and are not otherwise that detailed in specifying what policies should require. King’s proposed “replication standard,” for instance, stated that “sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party could replicate the results without any additional information from the author.”³⁵

With time, however, proposals for data sharing and transparency have become much more detailed. Certainly, the data themselves and/or detailed lists of data sources used are important, as is detailed documentation of those data that explains units of measure, response categories, and methods used for data collection. But if the points of data sharing are to permit replication and to provide greater transparency in terms of how the data produced a given collection of empirical findings, then the data and documentation alone are not enough. Researchers must also provide the code that implemented the analyses that produced those results so that others can see how data were weighted, how observations were fil-

tered, what specific variables were used, and what specific estimators and options were employed. Ideally, the code used for the analyses should provide enough information for an independent observer to understand the data sources used in the study and the steps used to process those resources for the final analysis. A researcher may even need to provide or identify the software used, as different programs or versions of a program may produce different results.³⁶ As empirical analyses increase in complexity, space limitations in articles often preclude providing the level of detail necessary to allow for replication and to meet standards of transparency. This makes it all the more important that policies for data sharing and replication include requirements to provide code in addition to the data themselves.³⁷ Indeed, the code may still be useful even if the data themselves cannot be provided due to their proprietary or sensitive nature.³⁸

Such considerations about the importance of providing both data and code have in turn worked their way into specific proposals for policies that journals should adopt. The data-sharing policy suggested by Gleditsch et al. is an illustrative example:

Authors of quantitative empirical articles must make their data available for replication purposes. A statement of how that is done should appear in the first footnote of the article. Required material would include all data, specialized computer programs, program recodes, and an explanatory file describing what is included and how to reproduce the published results.³⁹

Freese's proposed policy is of a similar nature:

Authors of accepted articles of empirical quantitative research are expected to use online archives to deposit maximum possible information pertinent to the verification of presented results at the time of publication. Ideally, data, code and other materials would be provided that would allow others to duplicate the analysis procedures that lead from original data to presented results without the need for any additional information about what was done from authors.⁴⁰

Hoffler and Kneib, commenting on the aftermath of efforts to replicate Reinhart and Rogoff's work on debt and economic growth, make a similar point:

In the narrow sense, replicability means that the raw data for an analysis can be accessed, that the transformation from the raw data to the final data set is well documented, and that software code is available for producing the final data set and the empirical results. Basically, this comes down to a question

of data and code availability, but nonetheless it is a necessary prerequisite for replication.⁴¹

Finally, there is the “Data Access and Research Transparency Joint Statement,” proposed by a group of political scientists and data archivists as a standard for data-sharing policies used by political science journals. This statement includes the following language, as part of the requirements journals should have for authors of submitted manuscripts:

Require authors to delineate clearly the analytic procedures upon which their published claims rely, and where possible to provide access to all relevant analytic materials. If such materials are not published with the article, they must be shared to the greatest extent possible through a trusted digital repository.⁴²

In other words, there is a common (if not unanimous) view among academics that if replication of empirical findings and transparency in research practices are to be more than abstract notions, then journals should require more than just provision of the data (or a listing of the data sources) used in a given work. Authors must also provide a detailed explanation of how the data were assembled and analyzed so as to clearly show how a given data resource or collection of data resources were used to produce a given set of empirical findings. The need is for transparency in terms of data access, data production, and data analysis.⁴³

Measuring Journals’ Data-Sharing Policies

Given the growing specificity of proposed standards for data sharing, transparency, and replication, how have social sciences journals responded? Have journals been adopting such policies? If so, what is the content of those policies? Is there any variation over time, or across academic fields?

These are questions that have not been systematically addressed in an empirical manner. The Association of Learned and Professional Society Publishers, for instance, produced a report in 2008 on scholarly publishing practices, based on a survey of publishers of academic journals. The survey included a question about availability of datasets associated with journal articles, and presented the results in summary form: 46.63% of publishers had such a policy. However, the report did not make the underlying data from the survey available for others to examine.⁴⁴ Therefore, the report does not necessarily offer much insight into our questions.

There has been some field-specific work on social sciences journals’ data and replication policies. Gleditsch and Metelits look at the policies of a collection of

27 journals in Political Science and International Relations and found that 18% of them had a stated replication policy.⁴⁵ Gherghina and Katsanidou examine 120 journals in Political Science that publish studies which make intensive use of data and found that 16% (19/120) of those journals had a policy about data availability, with the likelihood of a journal having such a possibility being a function of characteristics such as the age of the journal and its ISI Impact Factor.⁴⁶ In a study of data-sharing behavior of individual economists, Andreoli-Versbach and Mueller-Langer discuss the data sharing policies of 147 Economics journals and Business journals and identified 20% (30/147) of them with a data-availability policy.⁴⁷ However, none of these works is a comparative analysis that looks for differences across a comparable set of journals in different fields within the social sciences. Nor does any one of them look at changes over time to see whether more journals are adopting such policies.

To fill this gap, we conduct a comparative analysis in which we look at the policies of journals in different fields at different points in time and in which we do the following: (1) identify major academic journals in several disciplines in the social sciences; (2) determine whether they have policies about sharing data associated with articles published in them; and (3) evaluate the nature of those policies in terms of what they expect or require authors to do. Our starting point here is the aforementioned Gleditsch and Metelits study, which is one of the earliest empirical examinations of journals' data-sharing policies. In that study, the authors identify the top 15 journals in both International Relations and Political Science based on the number of citations reported for journals in the *Journal of Citation Reports*. We expand upon that approach by looking at a wider range of fields in the social sciences that make extensive use of quantitative data: Business/Finance, Economics, International Relations, Political Science, and Sociology. In addition, we identify the top 20 journals in each field, but based on each journal's 2003-vintage ISI Impact Factor score rather than the raw number of citations. This initial cut gives us 92 unique journal titles, as some journals are duplicated across two fields. In instances when a title is duplicated, we then assign it to one field and select the next-highest-ranked title in the other field until we are left with 100 unique titles overall that are evenly distributed across our 5 fields (i.e. 20 journals/field). We then code the data-sharing policies of each of these journals at two points in time, 2003 and 2015, so that we have a panel dataset of journals that will allow us to compare variation in journal policies both across fields and over time.[†]

In other words, we analyze the policies of top-ranked journals from 2003 to see how or whether the policies of that group of journals have changed over time. Since the publication of Gleditsch and Metelits' study, academics in the social sciences have continued to discuss and debate data sharing and transparency, to

† The final list of journals is available on the companion website (<https://data-brarianship.wordpress.com/>). The list indicates which journals were duplicated across fields and to which field a duplicated journal was assigned.

propose that journals adopt data-sharing policies, and to flesh out what those policies should expect of authors. Our approach allows us to see to what extent these discussions have subsequently affected the behavior of journals that Gleditsch and Metelits examined, as well as allowing us to see whether such discussions have influenced the policies of journals in other fields that were likewise top-ranked at the time of their study.

To code journal policies, we develop a 10-point ordinal index for these policies, which we have used in earlier presentations on this topic.⁴⁸ This policy consists of the following 5 elements:[†]

1. whether the journal has a stated policy about data availability/replication
2. the terms/language of policy, such as whether authors are “encouraged” to make data available or whether they are “expected” or “required” to make data available
3. the materials that authors have to provide as part of the policy, such as whether they have to provide data only or whether they also have to provide code/syntax to reproduce the analyses in the submission
4. the time-frame for providing data/replication materials, such as whether they have to be available upon request from other researchers post-publication versus whether they have to be provided prior to publication
5. the location at which the data must be available, such as whether authors are required to make the data available via a journal’s website or Dataverse

To collect data for journal policies from 2003 and construct index values, we consult issues of journals available via either print or microfilm. When print issues are not readily available, we rely on any digitized front or back matter available for issues from 2003 accessible in J-STOR. When print issues or issues in J-STOR referred to any URLs for a given journal, we also check any 2003-vintage copies of those sites that are accessible via the Internet Archive. For data on policies from 2015, we consult websites for individual journals and look specifically for submission instructions, any instructions for “Supplemental Materials,” editorial policies, and guidelines from publishers.

Journal Policies—Findings

Our results, presented in Table 14.1 below, indicate that a considerably greater fraction of our journals of interest has data-sharing policies now relative to 2003. In 2003, 10% (10/100 journals) had policies, while 39% (39/100) have such policies in 2015. While many of the journals that were top-ranked in 2003 adopted

† The index components in their entirety are available on the companion website (<https://databrarianship.wordpress.com/>).

policies between 2003 and 2015, most of them have not. In addition, the results also clearly indicate that the adoption of such policies has varied greatly across fields within the social sciences. Most of our Economics journals (13/20, or 65%) have adopted policies since 2003. For our remaining fields, however, half or fewer of the journals now have policies: 50% (10/20) for Business/Finance, 35% (7/20) for Political Science, 30% (6/20) for International Relations, and 15% (3/20) for Sociology. Even with all the ensuing discussion since 2003 about data sharing and replication, most of the journals that were top-ranked in 2003 did not subsequently adopt data-sharing policies.

Table 14.1. Summary Statistics by Academic Field and Year for 2003 Top-Ranked Journals in 2003 and in 2015

Field	Year	Number of Journals with Policies	Mean Policy Index Value	Maximum Policy Index Value	Minimum Policy Index Value
Business/Finance	2003	3	0.70	7.0	0
	2015	10	2.33	7.0	0
Economics	2003	1	0.15	3.0	0
	2015	13	4.45	9.0	0
International Relations	2003	4	1.00	6.5	0
	2015	6	2.23	9.0	0
Political Science	2003	1	0.20	4.0	0
	2015	7	2.30	10.0	0
Sociology	2003	0	0.00	0.0	0
	2015	3	0.90	8.0	0
Total	2003	9	0.41	7.0	0
	2015	39	2.44	10.0	0

As for the values for our 0–10 policy index score, we see similar patterns over time and across fields. Overall, the mean value for our index increased from 0.41 in 2003 to 2.44 in 2015. The mean value for the index has increased across all fields, with Economics again being the field with the largest mean value in 2015 (4.45), followed by Business (2.33), Political Science (2.30), International Relations (2.23), and Sociology (0.90). The minimum and maximum values also suggest considerable variability across journals within fields. While Sociology has not adopted journal policies to the same degree as have our other fields, for instance, the 2015 policy for *Sociological Methods & Research*, which has a value of 8, is amongst the most exacting policies in our list.

Table 14.2. Summary Statistics by Academic Field and Year for 2003 Top-Ranked Journals v. Current Top-Ranked Journals

Field	Year	Number of Journals with Policies	Mean Policy Index Value	Maximum Policy Index Value	Minimum Policy Index Value
Business/ Finance	2003	3	0.70	7.0	0
	2015	12	2.67	8.0	0
Economics	2003	1	0.15	3.0	0
	2015	17	5.63	9.0	0
International Relations	2003	4	1.00	6.5	0
	2015	8	2.40	9.0	0
Political Science	2003	1	0.20	4.0	0
	2015	11	3.83	10.0	0
Sociology	2003	0	0.00	0.0	0
	2015	7	1.52	8.0	0
Total	2003	9	0.41	7.0	0
	2015	55	3.21	10.0	0

As an alternate approach to looking at our questions of interest, we also compare top-ranked journals from 2003 to current top-ranked journals. Our basic methodology is the same as above, except that we are now also looking at the policies of top journals based on their most recent (2014) JCR impact rankings. Instead of using a panel dataset to compare past and present policies for journals that were top-ranked in 2003, we are comparing the policies of journals that were top-ranked in 2003 to the policies of journals that are currently top-ranked.[†] The results from this analysis, shown in Table 14.2, are quite similar in terms of variation over time and across fields. For journals that are currently top-ranked, a majority of them now have data-sharing policies. Economics and Business/Finance are once again the fields with the most journals with policies, followed by Political Science, International Relations, and Sociology. Economics and Political Science, meanwhile, are the fields with the highest mean policy index values, followed by Business/Finance, International Relations, and Sociology. In comparison to journals that were top-ranked in 2003, then, a small majority of journals that are currently top-ranked do have data-availability policies.

Journal Policies—Discussion

As both of our sets of results make clear, there is still considerable variation in data-availability policies, even as more journals have adopted such policies over time. Some examples taken from different journals in different fields provide illustrative examples of how policies vary. The *Review of Financial Studies*, for instance, does not have any formal policy on data availability. Instead, it “encourages our authors to post on the RFS Oxford University Press web page any material that they think will help others productively use their published research” and lists datasets as one example of such supplemental materials.⁴⁹

Public Opinion Quarterly has an explicit data policy, but it is relatively limited in terms of what authors are expected to do:

To permit competent professionals to confirm the results and analyses, authors are expected to retain raw data for a minimum of 5 years after publication of the research. Other information related to the research (e.g., instructions, treatment manuals, software, details of procedures) should be kept for the same period. This information is necessary if others are to attempt replication. Authors are expected to comply promptly and in a spirit of cooperation with such requests.⁵⁰

† The list of journals by field are available on <https://databrarianship.wordpress.com/>.

Authors for *Public Opinion Quarterly* are expected to retain and provide data, but only for a specific period of time. There are no additional requirements to deposit the data at a designated archive or that the data be available prior to publication, nor is there any explicit requirement to provide code explaining how the original data were transformed into whatever dataset was analyzed.

The *American Journal of Political Science*, meanwhile, has a very exacting policy. In a very lengthy statement on its new policies, the *AJPS* notes that authors have to deposit data in the journal's Dataverse site, have to provide code or instructions for how the analysis dataset was produced from the source data, and have to provide these materials prior to publication. Moreover, the *AJPS* makes final publication conditional upon the successful replication of the analysis in any submission; until the results can be successfully reproduced, there will be no publication of the manuscript.⁵¹ While the data policies for *Comparative Political Studies*⁵² and the *Journal of Conflict Resolution*⁵³ are not as strict as that of the *AJPS*, they do similarly go beyond stating that provision of data are "required" to explicitly state that even articles that have otherwise cleared the review process will not be published in the absence of replication files.

The *AJPS* language about replication of results being part of the review process is, it should be noted, not exclusive to that journal. While they are not in our list of journals under study, the *Quarterly Journal of Political Science*⁵⁴ and *Political Science Research and Methods*⁵⁵ have similar policies of conducting in-house reviews of submissions for replication purposes. While it is likewise outside of the scope of our study, we should also note that *Science* employs a similar approach.⁵⁶ This is a striking development, as it is an approach that very much goes beyond what proponents of data-sharing policies for journals have generally advocated.

Another noteworthy development involves whether journals designate specific locations for where data might reside. In earlier debates about data-sharing policies, various participants often did not get into the specifics of where replication data and materials should be deposited, and some even expressed skepticism about the notion of placing replication materials in the likes of data archives such as the ICPSR.⁵⁷ More recent proposals for replication policies, however, are more explicit about this matter and less inclined to endorse private communications or individual websites as appropriate means to share data. The recent DA-RT statement mentioned earlier, for instance, explicitly calls for the requirement that authors deposit replication materials in a "trusted digital repository."⁵⁸ Similarly, various scholars of inter-state and intra-state conflict argue that researchers are not trained archivists and that individual websites do not constitute data archives. Consequently, researchers should instead deposit data at "an organization that has professional staff and resources dedicated to that effort," with the likes of the ICPSR or the Dataverse Network given as examples.⁵⁹ While our focus is journals in the social sciences, we also note in passing that both *Nature* and *Science* likewise recommend that authors deposit data at bona fide archives such as Dryad.⁶⁰ This particular thread of the discussions about data sharing is also reflected in the policies of journals such as the *World Bank Economic*

Review, the *Journal of Money, Credit, and Banking*, the *American Journal of Political Science*, and *Sociological Methods & Research*, which now expect authors to make data available via journal websites and/or journal Dataverses.

Continued Barriers and Obstacles to Sharing

Our results indicate that a growing number of academic journals in different fields within the social sciences have been adopting data-sharing/replication policies since 2003. Our results also indicate that such adoption is not uniform across disciplines, and that many journals still do not have such policies. While our methods and list of journals differ from those of Gherghina and Katsanidou's study of political science journals and those of Andreoli-Verbach and Mueller-Langer's study of economics journals, our findings are nonetheless similar in that they likewise show that data-sharing/replication policies are not yet a standard practice for journals in the social sciences. Despite decades-long deliberations about this matter, it is still only partially the norm for academic journals to require authors to submit data and code as part of the publication process. Moreover, even when journals have policies, they do not always produce the desired results. McCullough and McCullough, et al. undertook various attempts to replicate findings in articles published in economics journals with replication policies and had only limited success in doing so, in part because the policies of the journals they looked at were not necessarily actively monitored and enforced at the time of writing.⁶¹ Likewise, Eubank's assessment of the in-house replication policy of the *Quarterly Journal of Political Science* noted that there were issues replicating findings in 13 out of 24 submissions under review from 2012-2014.⁶² More generally, data sharing is still not a common practice amongst academics, as evinced by Andreoli-Versbach and Mueller-Langer's study of data-sharing behavior amongst economists, which concluded that roughly 80% of their sample of scholars did not share data and code from their research.⁶³ Wicherts et al. similarly found that psychologists were not in the habit of sharing research data used in publications, even when requested to do so.⁶⁴ Why is this still the case, even when the potential benefits from data sharing seem so compelling?

We should note here that there are legitimate reasons why research data may not be readily sharable. Proponents of data sharing readily concede that data resources that are proprietary often cannot be shared. Likewise, the potential for disclosure risk is a sound and defensible reason to not share data that contain sensitive information about potentially-identifiable individuals.[†] Furthermore, we

† Questions about protecting the identity of individuals have been particularly salient in debates among political scientists over how or whether to apply DART to qualitative research. See the many commentaries at <http://dialogueondart.org/> for an overview of this particular debate.

should note that the debates about data sharing and replication policies are precisely that—debates. That these discussions have been ongoing for decades in and of itself suggests that there is not unanimity on this matter. While their focus is not replication specifically, Herrera and Kapur note that repeated reuse of pre-existing data can result in datasets having legitimacy in the eyes of researchers regardless of their quality in terms of measurement accuracy and the like.⁶⁵ Herrnson makes a related point that data-availability policies may have the effect of dissuading researchers from collecting their own data, thereby stunting the development of valuable skills that can result from such efforts.⁶⁶ Various participants in and observers of debates on data sharing and replication policies have also raised concerns about the de facto outcomes of such policies, such as focusing on minutiae and minor errors that then get undue levels of attention and have the potential to damage reputations because of relatively small mistakes and good-faith errors.⁶⁷ This in turn leads to questions such as how much replication is “enough” for a finding to be deemed robust and credible, how serious an error or discrepancy in an attempted replication has to be for the results to be called into question, and how to ensure that the replications themselves are conducted in a transparent and sound manner.⁶⁸ Even proponents of data sharing and transparency recognize that there may be practical difficulties in implementing replication policies. As Carsey notes, there are nuts-and-bolts considerations such as when replication materials should be submitted, how to address concerns about keeping replication materials secure during the review process, and whether providing replication materials may effectively identify submitters of manuscripts.⁶⁹

We should also note here that the incentive structures that researchers often face are not necessarily conducive to encouraging data sharing or replication. Preparing data and code and documentation so that others can readily attempt to replicate one’s work is a time-consuming process that becomes especially difficult as time passes and recollections fade. The rewards for such work may not be readily forthcoming. Data sharing is not necessarily something that will result in favorable tenure decisions or other such professional benefits. Indeed, from the perspective of individual researchers, sharing data may instead result in costs such as being “scooped,” losing the ability to maximize the benefits of one’s work, having one’s results be undermined via the replication efforts of others, and losing one’s competitive advantages in the context of a tight academic job market.⁷⁰ Moreover, replication work itself is not necessarily rewarded or even encouraged in the first place. While Ishiyama is very much in favor of data sharing and replication and transparency, he is also of the view that replication studies should not necessarily be published in major journals, where the focus should be more on “original” research. But, as Ishiyama acknowledges, such a position raises the questions of where replication studies should be published and why scholars should undertake efforts to replicate findings if such efforts are unlikely to be published in major journals.⁷¹ And as Andreoli-Versbach and Mueller-Langer point out, the lack of

reward for undertaking replication studies further reduces the incentives of individual researchers to make the effort necessary to ensure that their work can be replicated: why undertake such efforts when the likelihood of having one's work undergo replication is minor due to the lack of incentive for others to do so?⁷²

The incentives being what they are, it is thus not surprising that Hamermesh drily notes that “Economists treat replication the way teenagers treat chastity—as an ideal to be professed but not to be practiced.”⁷³ And while such sentiments may cause the heart of many a librarian to skip a beat or two, it should likewise not be a surprise for Gleditsch and Metelits to note that for many academics, “documentation, for instance, is often thought of as a waste of time.”⁷⁴ In the abstract, greater sharing of research data and transparency in their use can constitute a genuine public good for academics, and journal policies that encourage such behavior thus offer many potential collective benefits to the social sciences. However, the philosophical and practical questions about such policies and the incentive structures that weigh against such behavior still constitute considerable barriers to data sharing and the adoption of replication policies.

Implications for Data Sharing and Data Management

Our findings here suggest that the ongoing efforts of proponents of data-sharing policies for academic journals in the social sciences have not been entirely in vain. Our reviews of the policies of top-ranked journals to see how many have adopted policies over time shows that many of them have, and that the policies they are adopting are increasingly demanding. However, our results also show that such adoption has not been uniform across the social sciences. In our collection of journals, many more of them have such policies now than then. But many still do not. This suggests that there are continued issues for replication of empirical findings and transparency in research practices. If one agrees that replication is a crucial part of the scientific method and that transparency is key if academic research is to be credible, then one may still be concerned about the scientific rigor and credibility of much empirical research in the social sciences. Due to reasons such as practical considerations and problematic incentive structures, the level of data sharing is one that is still not socially-optimal within the social sciences.

Having said that, we want to emphasize that our findings show that the number of journals in different fields with data policies is on the increase, and that those policies are increasingly rigorous. And, there is some evidence in the social sciences (and in the sciences) that such policies have a positive effect on data sharing. Andreoli-Versbach and Mueller-Langer are, in general, downbeat about the willingness of economists to share research data. But they also find that economists who publish in journals with replication policies are more likely to make

data available, as the sunk costs of prepping data to meet those policies are such that the additional marginal costs of making data available more generally are minor.⁷⁵ Dafoe compares data availability for articles published in the *American Political Science Review* to those published in the *American Journal of Political Science* and concludes that the more exacting replication policy of the latter journal, while not perfect, “seems to have substantially increased the actual availability of replication files” when compared to the less exacting policy of the former.⁷⁶ In the sciences, both Piwowar and Piwowar and Chapman find that replication policies have a positive effect on availability of data associated with journal articles.⁷⁷ As previously noted, there is a tension between the social benefits for academic disciplines from data sharing and the potential costs for individual researchers in meeting such expectations. These findings, however, suggest that the adoption of data-sharing policies by academic journals provides an institutional means for resolving that tension in favor of greater data availability. This suggests that as initiatives such as DA-RT gain traction in fields such as political science, we should witness an even greater level of data sharing over time. At a minimum, we have a hypothesis which we will be able to test in the future.

Implications for Libraries

Our findings and reviews are also relevant for considerations of how librarians and libraries might interact with researchers with regard to data management and data sharing. While academic discussions have focused on issues of replication and incentives for sharing, libraries and librarians have focused on the rise of e-science and e-research and how libraries can and should respond to the rise of data-driven, networked research.⁷⁸ Environmental scans of the current data management landscape in research libraries reveal a fairly diverse set of institutional responses to campus needs for research data support, ranging from more traditional library services in the area of identifying data to highly engaged consulting services providing data management and sharing support for research teams.⁷⁹ In general, libraries supporting Ph.D. programs and research intensive programs tend to demonstrate a deeper set of services around data management plan support, data curation, and training. Additionally, some libraries have actively explored expanding librarian roles in working on research teams and providing data and information literacy training to teams charged with managing and sharing research.⁸⁰

Despite increased emphasis on research data at many university libraries, however, there is something of a gap between faculty and librarian perspectives of the availability and form of research services.⁸¹ The language used by academics in discussions about data sharing is often strikingly different than that used by librarians. For libraries providing research data management services, the focus is often on funder mandates and statements from entities such as the Office of Science and Technology Policy and how libraries should respond to such develop-

ments, for example, by assisting researchers with writing data management plans for grants or by developing institutional repositories as a means to store and preserve and disseminate data produced by grants. For academics, however, the focus is more often on replication and methodological considerations such as scientific rigor and transparency and credibility of findings. The 2014 symposium on data access and research transparency published in *PS: Political Science & Politics*, for instance, mentions funder mandates for data management plans in passing.⁸² The main focus is instead on how data sharing promotes research transparency and how such transparency might be further encouraged. As for where research data might reside, we have also noted that the tendency amongst academics and journal policies is to recommend usage of established archives such as the ICPSR or the Dataverse Network (or, in the sciences, the likes of Dryad or GenBank).

While data management plans for grant applications are certainly important for various fields and in certain contexts, they are not what academics tend to focus on when they talk about data management and data sharing. Instead, academics focus much more on matters of replication and transparency. As an example, a proposed open letter from a Director of Graduate Studies in the Political Science department at the University of Missouri says nothing about NSF mandates, but does mention transparency and the need to satisfy journals' replication policies.⁸³ Indeed, funder mandates are likely not applicable to many, if not most, researchers in the social sciences—whereas assistance in meeting the data availability policy of a journal as a step in the publication process is a service that likely will resonate with many researchers whether they are graduate students or faculty.⁸⁴ Likewise, data management support might also be better focused on helping researchers identify appropriate external options for depositing and archiving research data and on helping them with how to best prepare data for submission to such archives, as academics and journals are increasingly recommending pre-existing archives as suitable locations to place research data. Once more, the point here is not so much that libraries' efforts in data management support are misplaced, *per se*. It is instead that such efforts will be more effective if their scope is adjusted to incorporate the concerns and considerations expressed by academics in their own debates and deliberations about how they should best handle and use data. To the extent that librarians discuss such issues with the same language as do academics, they are more likely to make effective connections with them.

Given how academics are increasingly emphasizing evidence, transparency, and sharing, we suggest that librarians should incorporate such arguments and concerns into their deliberations about data management support and their interactions with researchers about data services, so that their proposed services and conversations address and inform faculty concerns about data as a scholarly object. This shift would largely be a matter of placing more emphasis, services, and instruction on many of the scholarly communication themes and workflows surrounding research data and data driven research. Indeed, much of the early

planning for librarian roles in e-science initiatives noted this connection such as Soehner et al., who specifically note that “the libraries may not see these scholarly communication issues as being connected to e-science, when, in fact, the connection is closer than is realized.”⁸⁵

Writing in *PS: Political Science & Politics* in 1995, Kenneth J. Meier predicted that “in 2010 A.D. every journal in political science will have some form of replication policy.”⁸⁶ Suffice it to say that this prediction has not proved to be entirely accurate. The findings of our analysis in this chapter show that top-ranked journals in political science and other fields within the social sciences have indeed been adopting such policies over time, but that such adoption has been incomplete and uneven across the disciplines. However, the apparent effectiveness of data-availability policies in promoting data sharing suggests that the increased adoption of such policies bodes well for the future promotion of data services. The academic literature and commentaries on matters of data sharing also offer suggestions for how librarians can discuss data management with academics in ways that more directly relate to researchers’ concerns and interests. Data sharing and data management already are and will likely continue to be central topics in larger discussions of scholarly communications and open access. Such discussions will only benefit from a greater awareness of the debates amongst academics about the benefits of data sharing and the means for best providing it.

-
1. Steve Lohr, “The Age of Big Data.” *New York Times*, February 11, 2012, <http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>.
 2. “The Data Deluge,” *The Economist*, February 25, 2010, <http://www.economist.com/node/15579717>.
 3. National Research Council, *Access to Research Data in the 21st Century: An Ongoing Dialogue Among Interest Parties: Report of a Workshop*, (Washington, D.C.: The National Academies Press, 2002), viii.
 4. National Institutes of Health, “NIH Data Sharing Policy,” last modified April 17, 2007, http://grants1.nih.gov/grants/policy/data_sharing/index.htm.
 5. National Science Foundation, “NSF Data Management Plans,” accessed April 20, 2015, <https://www.nsf.gov/eng/general/dmp.jsp>.
 6. Office of Science and Technology Policy, “Increasing Access to the Results of Federally Funded Scientific Research,” last modified February 22, 2013, https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
 7. Chris Thorpe and Simon Rogers, “Ordnance Survey Opendata Maps: What Does It Actually Include?” *The Guardian*, April 2, 2010, <http://www.theguardian.com/news/datablog/2010/apr/02/ordnance-survey-open-data>.
 8. Office of Citizen Services and Innovation Technologies, “Open Government,” accessed April 20, 2015. <https://www.data.gov/open-gov/>.
 9. Organization for Economic Co-Operation and Development, “Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29–30 January 2004—Final Communiqué,” accessed April 19, 2015. <http://www.oecd.org/science/sci-tech/sciencetechnologyandinnovationforthe21stcentury-meetingoftheoecdcommitteeforscientificandtechnologicalpolicyatministeriallevel29-30january2004-finalcommunique.htm>; Organization for Economic Co-Operation and Development, “OECD Principles and Guidelines for Access to Research Data from Public Funding,” accessed

- April 19, 2015. <http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>.
10. European Commission, "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020," last modified December 11 2013, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.
 11. Gary King, "Ensuring the Data-Rich Future of the Social Sciences." *Science* 331 (2011): 719, doi: <http://dx.doi.org/10.1126/science.1197872>.
 12. Anne R. Diekema, Andrew Wesolek, and Cheryl D. Walters, "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories," *Journal of Academic Librarianship* 40 (2014): 322–331, <http://dx.doi.org/10.1016/j.acalib.2014.04.010>; Regina Raboin, Rebecca Reznik-Zellen, and Dorothea Salo, "Forging New Service Paths: Institutional Approaches to Providing Research Data Management Services." *Journal of eScience Librarianship* 1 (2012): 134–147, doi: <http://dx.doi.org/10.7191/jeslib.2012.1021>.
 13. Organization for Economic Co-Operation and Development, "OECD Principles and Guidelines for Access to Research Data From Public Funding," 11.
 14. Terra Lawson-Remer, "Property Insecurity," *Brooklyn Journal of International Law* 38 (2012): 145–191. http://www.terralawsonremer.com/wp-content/uploads/2012/12/13_LawsonRemer_145_191-1_Property_Insecurity_Brooklyn_Law.pdf. More generally, see Kenneth J. Meier, "Replication: A View from the Streets," *PS: Political Science & Politics* 28 (1995): 456–459, <http://www.jstor.org/stable/420303>.
 15. Nils Petter Gleditsch, Claire Metelits, and Havard Strand, "Posting Your Data: Will You Be Scooped or Will You Be Famous?" *International Studies Perspectives* 4 (2003): 89–97, doi: <http://dx.doi.org/10.1111/1528-3577.04105>; see also Amy Pienta, George C. Alter, and Jared A. Lyle, *The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data* 2010, <http://hdl.handle.net/2027.42/78307>; Jeremy J. Albright and Jared A. Lyle, "Data Preservation Through Data Archives," *PS: Political Science & Politics* 43 (2010): 17–21, doi: <http://dx.doi.org/10.1017/S1049096510990768>; and Heather A. Piwowar, Roger S. Day, and Douglas B. Fridsma, "Sharing Detailed Research Data Is Associated with Increased Citation Rate," *PLoS ONE* 2 (2007): e308, doi: <http://dx.doi.org/10.1371/journal.pone.0000308>.
 16. Jeremy Freese, "Replication Standards for Quantitative Social Science: Why Not Sociology?" *Sociological Methods & Research* 36 (2007): 153–172, doi: <http://dx.doi.org/10.1177/0049124107306659>.
 17. Freese, "Replication Standards for Quantitative Social Science;" Arthur Lupia and George Alter, "Data Access and Research Transparency in the Quantitative Tradition," *PS: Political Science & Politics* 47 (2014): 54–59, <http://dx.doi.org/10.1017/S1049096513001728>; Arthur Lupia and Colin Elman, "Openness in Political Science," *PS: Political Science & Politics* 47 (2014): 19–42, doi: <http://dx.doi.org/10.1017/S1049096513001716>; Marcia McNutt, "Raising The Bar," *Science* 345 (2014): 9, doi: <http://dx.doi.org/10.1126/science.1257891>; Richard G. Anderson, William H. Greene, B. D. McCullough, and H. D. Vinod, "The Role of Data & Program Code Archives in the Future of Economic Research," *Federal Reserve Bank of St. Louis Working Paper 2005–014*, (St. Louis, MO. 2005), <http://research.stlouisfed.org/wp/2005/2005-014.pdf>; Nils Petter Gleditsch and Claire Metelits, "The Replication Debate," *International Studies Perspectives* 4 (2003): 89–97, doi: <http://dx.doi.org/10.1111/1528-3577.04105>; James Lee Ray and Brandon Valeriano, "Barriers to Replication in Systematic Empirical Research on World Politics," *International Studies Perspectives* 4 (2003): 79–85, doi: <http://dx.doi.org/10.1111/1528-3577.04105>; William G. Dewald, Jerry G. Thursby, and Richard G. Anderson, "Replication in Empirical Economics: The Journal of Money, Credit and Banking Project," *American Economic Review* 76 (1986): 587–603, <http://www.jstor.org/stable/1806061>; King, "Replication, Replication," *PS: Political Science & Politics* 28 (1995): 444–452, <http://www.jstor.org/stable/420301>; Meier, "Replication: A View from the Streets."
 18. Gary King, "Replication, Replication," 445.

19. Richard G. Anderson, William H. Greene, B.D. McCullough, and H.D. Vinod, "The Role of Data & Program Code Archives in the Future of Economic Research," 2.
20. *Ibid.*, 3.
21. Lupia and Elman, "Openness in Political Science: Data Access and Research Transparency," *PS: Political Science & Politics* 47 (2014): 20.
22. Ben Bernanke, "Editorial Statement," *American Economic Review* 94 (2004): 404, <http://www.jstor.org/stable/3592790>.
23. Luipa and Elman, "Openness in Political Science," 19.
24. Allen Dafoe, "The Imperative to Share Complete Replication Files," *Lux et Data* (blog), September 2013, <http://isps.yale.edu/news/blog/2013/09/the-imperative-to-share-complete-replication-files>.
25. Nicholas Eubank, "A Decade of Replications: Lessons from the Quarterly Journal of Political Science," *The Political Methodologist* (blog), December 9, 2014, <http://thepoliticalmethodologist.com/2014/12/09/a-decade-of-replications-lessons-from-the-quarterly-journal-of-political-science/>.
26. Dewald, Thursby, and Anderson, "Replication in Empirical Economics;" Anderson, Greene, McCullough, and Vinod, "The Role of Data & Program Code Archives;" B.D. McCullough, "Got Replicability: The Journal of Money, Credit, and Banking Archive," *Econ Journal Watch* 4 (2007): 326–337, http://econjwatch.org/file_download/170/2007-09-mccullough-econ_practice.pdf; B.D. McCullough, Kerry Anne McGeary, and Teresa D. Harrison, "Do Economics Journal Archives Promote Replicable Research?" *Canadian Journal of Economics / Revue canadienne d'Economie* 41 (2008): 1406–1420, <http://www.jstor.org/stable/25478330>; B.D. McCullough and H.D. Vinod, "Verifying the Solution from a Nonlinear Solver," *American Economic Review* 93 (2003): 873–892. doi: <http://dx.doi.org/10.1257/000282803322157133>.
27. David Funder, "Does 'Failure to Replicate' Mean Failed Science?" *LiveScience*, May 15, 2013, <http://www.livescience.com/32041-revisiting-science-studies.html>; Henry L. Roedigger, III, "Psychology's Woes and a Partial Cure: The Value of Replication," *Observer* 25 (2012), <http://www.psychologicalscience.org/index.php/publications/observer/2012/february-12/psychologys-woes-and-a-partial-cure-the-value-of-replication.html>; Ed Yong, "Replication Studies: Bad Copy," *Nature* 485 (2012): 298–300, <http://www.nature.com/news/replication-studies-bad-copy-1.10634>.
28. John Ioannidis, "Why Most Published Research Findings Are False," *PLoS Medicine* 2 (2005): e124, doi: <http://dx.doi.org/10.1371/journal.pmed.0020124>.
29. George Johnson, "New Truths That Only One Can See," *New York Times*, January 20, 2014, <http://www.nytimes.com/2014/01/21/science/new-truths-that-only-one-can-see.html>; "Trouble in the Lab," *The Economist*, October 19th, 2013, <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarms-degree-it-not-trouble>.
30. Bernardo A. Huberman, "Sociology of Science: Big Data Deserve A Bigger Audience," *Nature* 482 (2012): 308, doi: <http://dx.doi.org/10.1038/482308d>; Nathaniel Beck, "Research Replication in Social Science," *OUPblog* (blog), August 24 2014, <http://blog.oup.com/2014/08/research-replication-nathaniel-beck/>; Nathaniel Beck, "Replication Redux and Facebook Data," *OUPblog* (blog), January 19, 2015, <http://blog.oup.com/2015/01/replication-redux-facebook-data/>; John Markoff, "Trove of Personal Data, Forbidden to Researchers," *New York Times* May 21, 2012, <http://www.nytimes.com/2012/05/22/science/big-data-trove-stay-forbidden-to-social-scientists.html>.
31. Jan Hoffler and Thomas Kneib, "Economics Needs Replication," *The Institute Blog* (blog) April 23, 2013, <http://ineteconomics.org/blog/inet/economics-needs-replication>; James Kwak, "The Importance of Excel," *The Baseline Scenario* (blog) February 9, 2013, <http://baselinescenario.com/2013/02/09/the-importance-of-excel/>; James Kwak, "More Bad Excel," *The Baseline Scenario* (blog), April 18, 2013, <http://baselinescenario.com/2013/04/18/more-bad-excel/>; Paul Krugman, "The Excel Depression," *New York Times* April 18, 2013, <http://www.nytimes.com/2013/04/19/opinion/krugman-the-excel-depression.html>.

32. Victoria Stodden, Peixuan Guo, and Zhaokun Ma, "Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals," *PlosONE* 8 (2013): e67111, doi: <http://dx.doi.org/10.1371/journal.pone.0067111>.
33. Anderson, Greene, McCullough, and Vinod, "The Role of Data & Program Code Archives in the Future of Economic Research," 13.
34. Daniel S. Hamermesh, "Viewpoint: Replication in Economics." *Canadian Journal of Economics / Revue canadienne d'Economie* 40 (2007): 715–733, doi: <http://dx.doi.org/10.1111/j.1365-2966.2007.00428.x>.
35. King, "Replication, Replication," 444; Robert M. Hauser, "Sharing Data: It's Time for ASA Journals to Follow the Folkways of a Scientific Sociology," *American Sociological Review* 52 (1987): vi–viii, <http://www.jstor.org/stable/2095829>.
36. Nils Petter Gleditsch, Patrick James, James Lee Ray, and Bruce Russett, "Editors' Joint Statement: Minimum Replication Standards for International Relations Journals," *International Studies Perspectives* 4 (2003): 105, doi: <http://dx.doi.org/10.1111/1528-3577.04105>; Anderson, Greene, McCullough, and Vinod, "The Role of Data & Program Code Archives;" Freese, "Replication Standards for Quantitative Social Science;" Lupia and Alter, "Data Access and Research Transparency in the Quantitative Tradition." For discussion of the effects of software choices on replication specifically, see Anderson, Greene, McCullough, and Vinod, pp. 16–17.
37. Anderson, Greene, McCullough, and Vinod, "The Role of Data & Program Code Archives;" Freese, "Replication Standards for Quantitative Social Science."
38. Freese, "Replication Standards for Quantitative Social Science."
39. Gleditsch, James, Ray, and Russett, "Editors' Joint Statement," 105.
40. Freese, "Replication Standards for Quantitative Social Science," 167.
41. Hoffer and Kneib, "Economics Needs Replication."
42. Colin Elman and Arthur Lupia, "The (DA-RT) Data Access and Research Transparency Joint Statement," accessed April 1, 2015, <http://www.dartstatement.org/>.
43. Lupia and Elman, "Openness in Political Science."
44. John Cox and Laura Cox, *Scholarly Publishing Practice: Academic Journal Publishers' Policies and Practices in Online Publishing* (Brighton: Association of Learned and Professional Society Publishers, 2008), 74–75, 112.
45. Gleditsch and Metelits, "The Replication Debate."
46. Serigu Gherghina and Alexia Katsanidou, "Data Availability in Political Science Journals." *European Political Science* 12 (2013): 333–347, doi: <http://dx.doi.org/10.1057/eps.2013.8>.
47. Patrik Andreoli-Versbach and Frank Mueller-Langer, "Open Access to Data: An Ideal Professed but not Practiced," *Research Policy* 43 (2014) 1621–1633, doi: <http://dx.doi.org/10.1016/j.respol.2014.04.008>.
48. Joel Herndon and Robert O'Reilly, "Data Sharing Across the Disciplines: An Empirical Study," presented at the Annual Meeting of the International Association for Social Science Information Services and Technology, Tampere, Finland, May 2009, http://www.iassistdata.org/downloads/2009/c4_herndon_oreilly.pdf; Robert O'Reilly, "Data Sharing Across the Disciplines, Revisited: Academic Journals and Replication Policies," presented at the Annual Meeting of the International Association for Social Science Information Services and Technology, Washington, D.C., June 2012, http://www.iassistdata.org/downloads/2012/2012_a3_oreilly.pdf.
49. "Paper Addenda," *Review of Financial Studies*, accessed April 6, 2015, <http://rfsfs.org/paper-addenda/>.
50. "Note to Contributors," *Public Opinion Quarterly*, accessed April 20, 2015, http://www.oxford-journals.org/our_journals/poq/for_authors/general.html.
51. William G. Jacoby, "The AJPS Replication Policy: Innovations and Revisions," *AJPS Editor Blog* (blog), March 26, 2015, <http://ajps.org/2015/03/26/the-ajps-replication-policy-innovations-and-revisions/>.
52. "Submission Instructions," *Comparative Political Studies*, accessed April 14, 2015, <http://www.sagepub.com/journals/Journal200828/manuscriptSubmission#tabview=manuscriptSubmission>.

53. "Submission Guidelines," *Journal of Conflict Resolution*, accessed April 14, 2015, <http://www.sagepub.com/journals/Journal200764/manuscriptSubmission>.
54. "Author Instructions," *Quarterly Journal of Political Science*, accessed December 19, 2014, <http://www.nowpublishers.com/Journal/AuthorInstructions/QJPS>.
55. "Editorial Statement," *Political Science Research Methods*, accessed December 19, 2014, http://journals.cambridge.org/images/fileUpload/documents/RAM_ifc.pdf.
56. McNutt, "Raising the Bar."
57. Ray and Valeriano, "Barriers to Replication."
58. Elman and Lupia, "The (DA-RT) Data Access and Research Transparency Joint Statement."
59. Christian Davenport and Will H. Moore, "Conflict Consortium Standards & Best Practices for Observational Data," last updated April 7, 2015, <http://conflictconsortium.weebly.com/uploads/1/8/3/5/18359923/cc-datastandardspractices7apr2015.pdf>; Idean Salehyan and Henrik Urdal, "Data Challenges in Conflict Research," *Political Violence @ a Glance* (blog), February 6, 2015, <http://politicalviolenceataglance.org/2015/02/06/data-challenges-in-conflict-research/>.
60. "Availability of Data, Materials, and Methods," *Nature* accessed December 19, 2014, <http://www.nature.com/authors/policies/availability.html>; "Data and Materials Availability," *Science*, accessed February 17, 2015, http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail.
61. McCullough, "Got Replicability?"; McCullough, McGeary, and Harrison, "Do Economics Journals Promote Replicable Research?"
62. Nicholas Eubank, "A Decade of Replications: Lessons from the *Quarterly Journal of Political Science*."
63. Andreoli-Versbach and Mueller-Langer, "Open Access to Data."
64. Jelte M. Wicherts, Denny Borsboom, Judith Kats, and Dylan Molenaar, "The Poor Availability of Psychological Research Data for Reanalysis," *American Psychologist* 61 (2006): 726–728, doi: <http://dx.doi.org/10.1037/0003-066X.61.7.726>.
65. Yoshiko M. Herrera and Devesh Kapur, "Improving Data Quality: Actors, Incentives, and Capabilities," *Political Analysis* 15 (2007): 365–386, <http://www.jstor.org/stable/25791902>.
66. Paul Herrnson, "Replication, Verification, Secondary Analysis and Data Collection in Political Science." *PS: Political Science & Politics* 28 (1995): 452–455, <http://www.jstor.org/stable/420302>.
67. James L. Gibson, "Cautious Reflections on a Data-Archiving Policy," *PS: Political Science & Politics* 28 (1995): 473–476, <http://www.jstor.org/stable/420310>; Nicole Janz, "Replication Bullying: Who Replicates the Replicators?" *Political Science Replication* (blog), May 25, 2014, <https://politicalsciencereplication.wordpress.com/2014/05/25/replication-bullying-who-replicates-the-replicators/>.
68. Thomas Carsey, "Making DA-RT A Reality," *PS: Political Science & Politics* 47 (2014): 72–77, doi: <http://dx.doi.org/10.1017/S1049096513001753>; Janz, "Replication Bullying." John Ishiyama, "Replication, Research Transparency, and Journal Publications: Individualism, Community Models, and the Future of Replication Studies," *PS: Political Science & Politics* 47 (2014): 78–83, doi: <http://dx.doi.org/10.1017/S1049096513001765>.
69. Carsey, "Making DA-RT A Reality."
70. Andreoli-Versbach and Mueller-Langer, "Open Access to Data;" Anderson, Greene, McCullough, and Vinod, "The Role of Data & Program Code Archives;" Herrnson, "Replication, Verification, Secondary Analysis;" Piwowar, Day, and Fridsa, "Sharing Detailed Research Data."
71. Ishiyama, "Replication, Research Transparency, and Journal Publications."
72. Andreoli-Versbach and Mueller-Langer, "Open Access to Data."
73. Daniel Hamermesh, "Viewpoint: Replication in Economics."
74. Gleditsch and Metelits, "The Replication Debate."
75. Andreoli-Versbach and Mueller-Langer, "Open Access to Data."
76. Allen Dafoe, "Science Deserves Better: The Imperative to Share Complete Replication Files," *PS: Political Science & Politics* 47 (2014): 61, doi: <http://dx.doi.org/10.1017/S104909651300173X>.

77. Heather A. Piwowar, "Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data," *PLoS ONE* 6 (2011): e18657, doi: <http://dx.doi.org/10.1371/journal.pone.0018657>; Heather A. Piwowar and Wendy W. Chapman, "A Review of Journal Policies for Sharing Research Data," In *Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0—Proceedings of the 12th International Conference on Electronic Publishing (EL-PUB)* June 25–27 2008, Toronto Canada, http://www.researchremix.org/wordpress/wp-content/uploads/2009/09/ELPUB_2008_Piwowar.pdf.
78. Amy Friedlander and Prudence Adler, "To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering," *A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe* (Washington, D.C.: Association of Research Libraries, 2006), <http://www.arl.org/storage/documents/publications/digital-data-report-2006.pdf>; Anna Gold, "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians," *D-Lib Magazine* 13 (2007), <http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>; Anna Gold, "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge," *D-Lib Magazine* 13 (2007), <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>; Catherine Soehner, Catherine Steeves, and Jennifer Ward, *E-Science and Data Support Services: A Study of ARL Member Institutions* (Washington, D.C.: Association of Research Libraries, 2010), <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>; Amanda L. Whitmire, "Thoughts on 'eResearch': A Scientist's Perspective," *Journal of eScience Librarianship* 2(2013): 68–72, doi: <http://dx.doi.org/10.7191/jeslib.2013.1045>.
79. Carol Tenopir, Ben Birch, and Suzie Allard, *Academic Libraries and Research Data Services: Current Practices and Plans for the Future* (Washington, D.C.: Association of Research Libraries, 2012) http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf; David Fearon, Betsy Gunia, Barbara E. Pralle, Sherry Lake, and Andrew L. Sallans, *ARL Spec Kit 334: Research Data Management Services*. (Washington, D.C.: Association of Research Libraries, 2013) <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>.
80. Jacob Carlson, Michael Fosmire, CC Miller, and Megan Sapp Nelson, "Determining Data Information Literacy Needs: A Study of Students and Research Faculty," *Portal: Libraries and the Academy* 11 (2011): 629–57, doi: <http://dx.doi.org/10.1353/pla.2011.0022>; Jennifer Muilenburg, Mahria Lebow, and Joanne Rich, "Lessons Learned From a Research Data Management Pilot Course at an Academic Library," *Journal of eScience Librarianship* 3 (2014): 67–73, doi: <http://dx.doi.org/10.7191/jeslib.2014.1058>; Mary Piorun, Donna Kafel, Tracey Leger-Hornby, Siamak Najafi, Elaine Martin, Paul Colombo, and Nancy LaPelle, "Teaching Research Data Management: An Undergraduate/Graduate Curriculum," *Journal of eScience Librarianship* 1 (2012): 46–50, doi: <http://dx.doi.org/10.7191/jeslib.2012.1003>; Donna Kafel, Andrew Creamer, and Elaine Martin, "Building the New England Collaborative Data Management Curriculum," *Journal of eScience Librarianship* 3 (2014): 60–66, doi: <http://dx.doi.org/10.7191/jeslib.2014.1066>.
81. Anne R. Diekema, Andrew Wesolek, and Cheryl D. Walters, "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories," *Journal of Academic Librarianship* 40 (2014): 322–331.
82. See Dafoe, "Science Deserves Better," and Carsey, "Making DA-RT A Reality."
83. Amanda Murdie, "An Open Letter From the New DGS," *Duck of Minerva* (blog) July 2, 2014, <http://duckofminerva.com/2014/07/an-open-letter-from-the-new-dgs.html>.
84. Katherine G. Akers, "Going Beyond Data Management Planning: Comprehensive Research Data Services," *College & Research Libraries News* 75 (2014): 435–436, <http://crln.acrl.org/content/75/8/435.full>.
85. Soehner, Steeves, and Ward, "E-Science and Data Support Services."
86. Meier, "Replication: A View from the Streets," 458.