# HOW JOURNALS ARE ADOPTING OPEN DATA AND CODE POLICIES[1]

Victoria Stodden[2]
Peixuan Guo[3]
Zhaokun Ma[4]

*~ DRAFT ~*

*ABSTRACT*

*The move toward journal policy on data and code publication is indisputable. We show that increasing numbers of journals are using their publication policies to support reproducible research, facilitate data and code re-use, and enable broader communication of the output of digital scientific research. In this paper we take stock of data and code policies for 170 journals, and document how their policies have changed from June of 2011 to June 2012. We hypothesize that open data and code policies are in the process of being adopted more widely, that data policy adoption leads code policy adoption, and that open access journals would be more likely to have policies making data and code open as well. We found evidence to support our first two hypotheses, and little evidence to support the third. Of the journals in this study, 38% had a data policy, 22% had a code policy, and 66% had a supplemental materials policy as of June of 2012. This reflects a striking one year increase of 16% in the number of data policies, a 30% increase in code policies, and a 7% increase in the number of supplemental materials policies. We find that of journals with open data and code policies, they tend to adopt open data policies first and then adopt open code policies later, independently of the existence of supplemental materials policies. Data and code sharing is no longer an obscure or unconventional part of the publication process, but has entered the mainstream.*

**Keywords:** reproducible research, journal policy, open data, open code, data policy, code policy, data sharing, code sharing, open science, supplemental materials policy, access to knowledge

[1] Presented at the International Association for the Study of the Commons, 1st Thematic Conference on the "The Knowledge Commons," September 12, 2012.
[2] Assistant Professor, Department of Statistics, Columbia University, and communicating author: victoria@stodden.net.
[3] Student, Department of Statistics, Columbia University.
[4] Student, Department of Industrial Engineering and Operations Research, Columbia University.

## INTRODUCTION

Digitization is changing the practice of science in pervasive and important ways. First, we live in a world in which data is being digitally recorded on a massive scale. Our DNA, for example, is encoded as genomic sequence data, scans of brain activity exist in functional magnetic resonance image datasets, and records of our personal interactions are stored as social network datasets. Readings from our environment and climate are stored in myriad time series datasets, business and government transactions are dutifully recorded in digital databases, and scans of published text make available vast corpuses of our thinking and communications.

Equally as importantly, reasoning about these data is now recorded digitally in the software, scripts, and codes that analyze our digitally recorded world. The result is a deep digitization of the process of scientific knowledge generation toward computational science, bringing new ways of understanding our world. The parallel development of the Internet as a pervasive communication mechanism for digital content has created an unprecedented opportunity for the acceleration of scientific knowledge generation, giving us the ability to share scientific data and methods in an open way.

As yet this opportunity still escapes us. Communication practices that were established in the pre-digital scientific age prevail and are now engendering a deep credibility crisis in computational science. Without access to the data and computer code that underlie scientific discoveries, published findings are all but impossible to verify. Computational results are frequently of a complexity that makes a complete recitation of the steps taken to arrive at a result prohibitive in typical scientific publications today. As one of us noted in 2009,

> At conferences and in publications, it's now completely acceptable for a researcher to simply say, "here is what I did, and here are my results." Presenters devote almost no time to explaining why the audience should believe that they found and corrected errors in their computations. The presentation's core isn't about the struggle to root out error — as it would be in mature fields — but is instead a sales pitch: an enthusiastic presentation of ideas and a breezy demo of an implementation. Computational science has nothing like the elaborate mechanisms of formal proof in mathematics or meta-analysis in empirical science. Many users of scientific computing aren't even trying to follow a systematic, rigorous discipline that would in principle allow others to verify the claims they make. How dare we imagine that computational science, as routinely practiced, is reliable! [1]

A necessary response to this crisis is the adoption of reproducible computational research, in which all details of the computations — the underlying data and the code that generated the results — are made conveniently available to others. This raises communication standards for computational science to permit the longstanding scientific requirement of verification for published results, as well as facilitating re-use of the code and data.

There have been numerous calls for data and code release from across the computational sciences. Most recently the July/August 2012 issue of Computing in Science and Engineering was dedicated to Reproducible Research [2] and called for "changing the culture" [3]. In 2009 a Roundtable at Yale Law School addressed the issue of reproducibility by convening computational scientists from across a variety of disciplines [4,5]. The mantra is the same: making available the data upon which the published results are based, and making available the complete set of computer instructions which generated the published results. Numerous editorials and commentaries have made the same appeals [6,7,8,9].

Reproducible computational research can be facilitated through a variety of possible mechanisms. Individual scientists can adopt data and code sharing practices of their own accord, and many have [10,11,12,13,14]. Scientists are also subject to funding agency guidelines, tenure and promotion committee standards, and journal publication requirements. If funds are awarded to a scientist and the funding body enforces a requirement that datasets and code created in the course of the research are made available, those actions can have a profound impact on the dissemination of the research. Such guidelines exist for several U.S. Federal Funding Agencies. For example National Science Foundation (NSF) and the National Institutes of Health (NIH) require dataset disclosure and encourage software availability, as seen in the following excerpts from their grant guidelines,

> "NSF ... expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable." (2005 and earlier) [15], and
>
> NIH (2003): "The NIH endorses the sharing of final research data to serve these and other important scientific goals. The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers." (for grants over $500,000, also include data sharing plan) [16]

A common criticism of funding agency guidelines is their general lack of enforcement. In fact, the recent implementation of a Data Management Plan at the NSF could be seen as a first step toward enforcement of these grant guidelines [17]. Since January 2011 the NSF has required the submission of a 2-page data management plan with every new grant application that outlines plans for the stewardship of the data arising from the funding opportunity. The requirement is agnostic as far as data release, but nonetheless the plan must be submitted and it is subject to peer review during the application's evaluation. Recently tentative steps toward a similar Software Management Plan have been made at the federal agency level. A recent proposal initiated by the NSF called for a Software Sharing Plan [18] as have a number of NIH grants [19]. Enforcing data and software dissemination strategies for sponsored research would be an important step toward reproducible computational science.

This paper investigates a third lever affecting scientific communication, the standards imposed by journal requirements for published computational articles. Both the prevalence and nature of journal data and code sharing policies is thought to have a direct effect on data and code availability, and the reproducibility of published computational results. Beyond funding agency or institutional requirements, journals exert a tremendous amount of pressure on communication standards for scientific knowledge dissemination. We follow on one principle and one recommendation made in a National Academies of Science 2003 report, called Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences stating:

> **Principle 1. (Chapter 3) Authors should include in their publications the data, algorithms, or other information that is central or integral to the publication—that is, whatever is necessary to support the major claims of the paper and would enable one skilled in the art to verify or replicate the claims.** This is a *quid pro quo*—in exchange for the credit and acknowledgement that come with publishing in a peer-reviewed journal, authors are expected to provide the information essential to their published findings. (p. 5)
>
> **Recommendation 6. (Chapter 6) Scientific journals should clearly and prominently state (in the instructions for authors and on their Web sites) their policies for distribution of publication-related materials, data, and other information. Policies for sharing materials should include requirements for depositing materials in an appropriate repository. Policies for data sharing should include requirements for deposition of complex datasets in appropriate databases and for the sharing of software and algorithms integral to the findings being reported. The policies should also clearly state the consequences for authors who do not adhere to the policies and the procedure for registering complaints about noncompliance.** Many journals do not specify policies about sharing data and materials in their instructions to authors. By incorporating transparent standards into their official policies (including a statement of consequences for authors who do not comply), journals can encourage compliance. It is not known how many instances of noncompliance are ever brought to the attention of journal editors or other external authorities; however, a letter from

> the editor-in-chief or managing editor is often sufficient to resolve problems. Although some journal editors would consider denying a noncomplying author further rights to publish in their journals, on rare occasions, public opinion might be the most influential way to obtain an author's compliance. A journal might choose to declare an author's noncompliance (after all honest attempts were exhausted) in a specific section dedicated to this purpose. (p. 10). [20]

In summary, Principle 1 calls for the dissemination of data, software, and all information necessary for a researcher to "verify or replicate the claims" made in the publication. Recommendation 6 is a call for journals to clarify and explicitly state their policies regarding data and code release requirements, and to state consequences for authors who do not comply with these requirements. Appendix 1 lists the Principles and Recommendations given by the National Academies task force in their entirety.

Based on these exigencies we undertake an examination of modern journal policy in computational fields. We are interested in the nature of the journal's data sharing policy, the nature of their code sharing policy, and the nature of their supplemental materials policy. We expect policy changes to move in the direction of greater disclosure requirements, with code sharing requirements lagging behind data sharing requirements.

A number of journals have enacted data, code, and/or supplemental materials requirements. The two highest ranked journals in scientific publication, *Nature* and *Science*, both now require authors to make available the data underlying their published results upon request, and in February of 2011 *Science* extended this policy to include code and software [21]. One fundamental research question we seek to address in this paper is the role of leadership in journal policy setting, specifically whether this action on the part of flagship journals could be expected to create a "trickle down" effect to other journals. Another policy implemented by the journal *Biostatistics* in 2009 offered the authors of accepted articles the option to present their code for independent replication by the Associate Editor for Reproducibility. If the Associate Editor was able to replicate the results in the article this would be broadcast by kitemark on the first page of the published article. Authors can choose to release the code, the data, or both data and code (making the paper fully reproducible), indicated by kitemarks of C, D, or R, respectively, on the title page. In its first year, three issues of *Biostatistics* with a total of 43 papers have been published and of those, four papers have been marked with code availability, two with data availability, one with both, and two as fully reproducible.

In this paper we seek a unified understanding of the evolution of journal policy and its role in encouraging reproducible computational science. We study the data and code sharing policies of a group of computational journals in June of 2011 and then again in June of 2012.

## METHODS

From the ISI Web of Knowledge journal categorizations, we chose "Mathematical & Computational Biology," "Statistics & Probability," and "Multidisciplinary Sciences" since they are likely to contain computational results and therefore most likely to contain journals engaging in the development of data and code sharing policies. We choose to include computational biology because of the strides made toward data sharing in this field over the last decade or so, increasing our chances of studying journals with more advanced policy implementations regarding data and code sharing. We then chose to add 5 additional journals, Nature Genetics, Cell, Lancet, Nature Physics, and Materials Science and Engineering Reports, due to their high impact factors and likelihood of publishing computational results. After removing the small number of journals that are no longer published, this effort selected 170 journal titles as shown in Table 1 (Appendix 2 gives the complete alphabetical list of journal titles included in this study).

**Table 1: ISI Classifications Represented in the Journal Titles**

| **ISI Classification** | **Count** |
|---|---|
| Statistics & Probability | 98 |
| Multidisciplinary Science | 45 |
| Mathematical & Computational Biology | 30 |
| Genetics & Heredity (Nature Genetics) | 1 |
| Biochemistry & Molecular Biology; Cell Biology (Cell) | 1 |
| Medicine, General & Internal (Lancet) | 1 |
| Physics, Multidisciplinary (Nature Physics) | 1 |
| Materials Science, Multidisciplinary; Physics, Applied (Materials Science & Engineering R - Reports) | 1 |
| In both "Statistics & Probability" and "Mathematical & Computational Biology" | -8 |
| **Adjusted Total** | **170** |

These journals have a wide variety of impact factors and publishers. Tables 2 and 3 summarize this information, showing the distribution of impact factors for journal titles included in the study and the publishing houses represented. Unsurprisingly, given the distribution of impact factor rankings across journals in general, a majority of the titles in this study have an impact factor of 1 or less (110 of 170) and 15 titles had an impact factor of 5 or greater. As shown in Table 3, the Springer Publishing House publishes the greatest percent of the journal titles in this study, at 17.1%, with Wiley, Elsevier, and Taylor & Francis rounding out the largest proportions of the journals in this study.

**Table 2: Distribution of Impact Factors for Journal Titles**

| ISI Impact Factor (inclusive) | Count |
|---|---|
| 30-35 | 5 |
| 10-29 | 2 |
| 8-9 | 1 |
| 6-7 | 1 |
| 4-5 | 6 |
| 2-3 | 45 |
| 0-1 | 110 |
| **Total** | **170** |

**Table 3: Publishing Houses for Journal Titles**

| Publishing House | Count | Percent |
|---|---|---|
| Springer (incl. Springer Heidelberg, Springer/Plenum Publishers, MAIK Nauka Interperiodica Springer, BioMed Central) | 29 | 17.1% |
| Wiley (incl. John Wiley & Sons, Wiley-Blackwell Publishing, and Wiley-VCH Verlag GmbH) | 20 | 11.8% |
| Reed Elsevier (incl. Elsevier Science BV, Academic Press LTD – Elsevier Science, and Pergamon-Elsevier Science LTD) | 19 | 11.2% |
| Taylor & Francis (incl. Lawrence Erlbaum Associates Inc. and Routledge Journals) | 13 | 7.6% |
| Macmillan (Nature Publishing Group) | 3 | 1.8% |
| Scientific Societies | 31 | 18.2% |
| Other For-Profit Publishers | 33 | 19.4% |
| Not-For-Profit Non-Society Publishers | 22 | 12.9% |
| **Total** | **170** | **100%** |

For all of these journal titles we inspected their websites, once in June of 2011 and a second time in June of 2012, to ascertain their policies on data sharing, code sharing, and supplementary materials. Each of these policies was evaluated on a 5-point scale, as shown in Table 4. We included supplemental materials policy as a proxy for openness as data or code and as a possible bellweather for changes in data and code policies. Supplemental materials, however, tend to include figures and explanations that were not included in the main article rather than data or code.

**Table 4: Classification of Journal Policies**

**Data Sharing Policy**

1. Required as condition of publication, barring exceptions
2. Required but may not affect editorial decisions
3. Explicitly encouraged/addressed, may be reviewed and/or hosted
4. Implied
5. No mention

**Code Sharing Policy**

1. Required as condition of publication, barring exceptions
2. Required but may not affect editorial decisions
3. Explicitly encouraged/addressed, may be reviewed and/or hosted
4. Implied
5. No mention

**Supplemental Materials Policy**

1. Required as condition of publication, barring exceptions
2. Required but may not affect editorial decisions
3. Explicitly encouraged/addressed, may be reviewed and/or hosted
4. Implied
5. No mention

We also collected information to supplement these rankings to help illuminate and contrast policies. Each data, code, and supplemental materials policy ranking was augmented depending on whether they were specified to be shared via submission to the journal, upon request from readers, or whether this was left unspecified. The policy ranking was further augmented to indicate whether the journal specified that the author was intended to share with colleagues and other researchers in the field, or with the general public (unspecified). The code sharing policy classification was augmented with an additional parameter that signaled whether the journal has restricted the code policy to apply only to articles with “substantial” code or software. We also recorded whether the journal explicitly permitted either the posting of the final version or a draft of the published version of the article on authors’ website. The final factor we recorded was whether the journal indicated it would review data, code, or supplemental materials submissions, and whether these would be hosted by the journals. These additional policy classifications documented the vast majority of journal policy variation.

## RESULTS

Classifying journal policies according to the ranking system given in Table 4 yielded a snapshot of the journal publication standards on the availability of the data and code associated with published computational findings. Table 5 gives counts for each classification by year, with the change from 2011 to 2012.

**Table 5: Net Changes in Data, Code, and Supplementary Materials Policy Classifications from 2011 to 2012**

| Data Sharing Policy (n=170) | 2011 | 2012 | Change |
|---|---|---|---|
| Required as condition of publication, barring exceptions | 18 | 19 | 1 |
| Required but may not affect editorial decisions | 3 | 10 | 7 |
| Explicitly encouraged/addressed, may be reviewed and/or hosted | 35 | 30 | -5 |
| Implied | 0 | 5 | 5 |
| No mention | 114 | 106 | -8 |

| Code Sharing Policy (n=170) | 2011 | 2012 | Change |
|---|---|---|---|
| Required as condition of publication, barring exceptions | 6 | 6 | 0 |
| Required but may not affect editorial decisions | 6 | 6 | 0 |
| Explicitly encouraged/addressed, may be reviewed and/or hosted | 17 | 21 | 4 |
| Implied | 0 | 3 | 3 |
| No mention | 141 | 134 | -7 |

| Supplemental Materials Policy (n=170) | 2011 | 2012 | Change |
|---|---|---|---|
| Required as condition of publication, barring exceptions | 8 | 6 | -2 |
| Required but may not affect editorial decisions | 7 | 10 | 3 |
| Explicitly encouraged/addressed, may be reviewed and/or hosted | 86 | 93 | 7 |
| Implied | 4 | 3 | -1 |
| No mention | 65 | 58 | -7 |

Table 5 can be interpreted as follows. The majority of journal titles included in our study have not followed the recommendations of the National Academies Committee mentioned in the introduction by describing their data and code sharing policies on their websites. In June of 2012, 62% of the journals in this study make no mention of a data policy and 79% make no mention of a software policy. However, in 2012 66% have a supplemental materials policy. Of the remaining journals that mention data or software policies on their website, the majority encourage the practice of sharing but do not require it: 47% of journals with a data sharing policy encourage sharing and 45% require it (including the 16% who state compliance with this requirement will not affect publication decisions). Similarly, 56% of journals with a software policy choose encouragement of code sharing, and 32% require code disclosure (including the 16% who indicate that noncompliance with this requirement will not affect publication decisions).

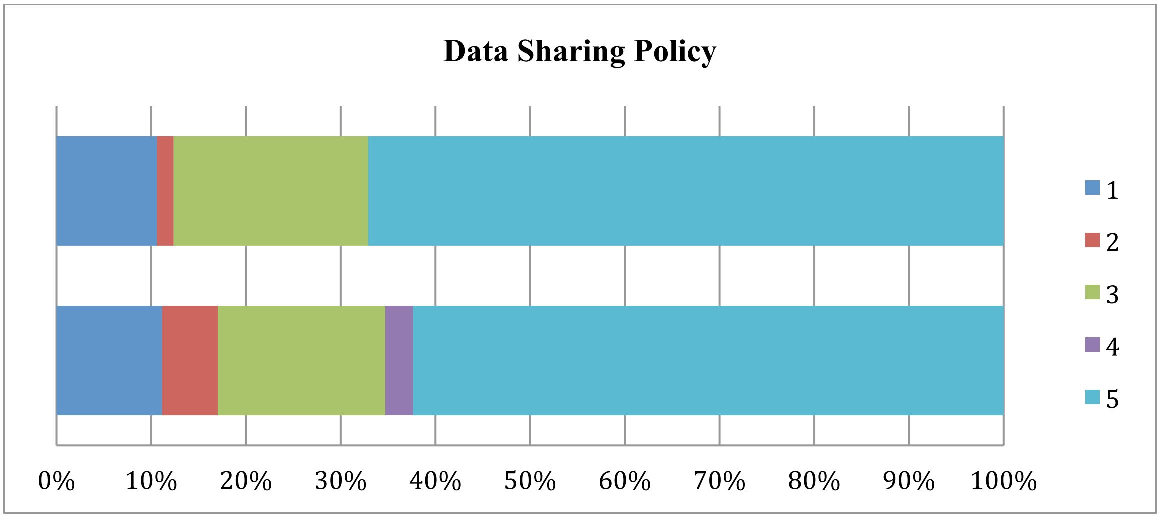
Supplement materials exhibited a different pattern. There was a net reduction of two journals requiring supplemental materials, and there was a net gain of seven journals adopting explicit encouragement of supplemental materials inclusion with publication.

Table 5 displays aggregate results and not individual policy change numbers, and analysis of these changes from 2011 to 2012 requires digging deeper into the results. Overall, thirty journals made a data policy change from 2011 to 2012, 12 made a change in their software policy, and 36 made a change to their supplemental data policy. The net change numbers in Table 5 indicate this change is markedly in the direction of openness.

There are a total of 104 journals with neither an open data nor open code policy in 2012, down from 110 in 2011. Finally, 39% of journals had some form of open data or open code policy in 2012, up from 35% in 2011.

Figures 1 through 3 display the data from Table 5 graphically. The top bar in each shows the distribution of policy classifications in 2011, and the bottom bar shows the distribution of policy classifications in 2012. In Figure 1, the proportion of journals with no data policy shrank (the teal blue portion), whereas data sharing requirements without impact on publishing rose, and a new section representing an implied data sharing policy appeared.

**Figure 1: Breakdown of Changes in Data Sharing Policy, 2011-2012**




In Figure 2, similar results are apparent for the changes in code sharing policy. The proportion of journals in the study with no code sharing policy shrank from 83% to 79%, and those that encouraged code sharing grew from 10% to 15%.

**Figure 2: Breakdown of Changes in Code Sharing Policy, 2011-2012**

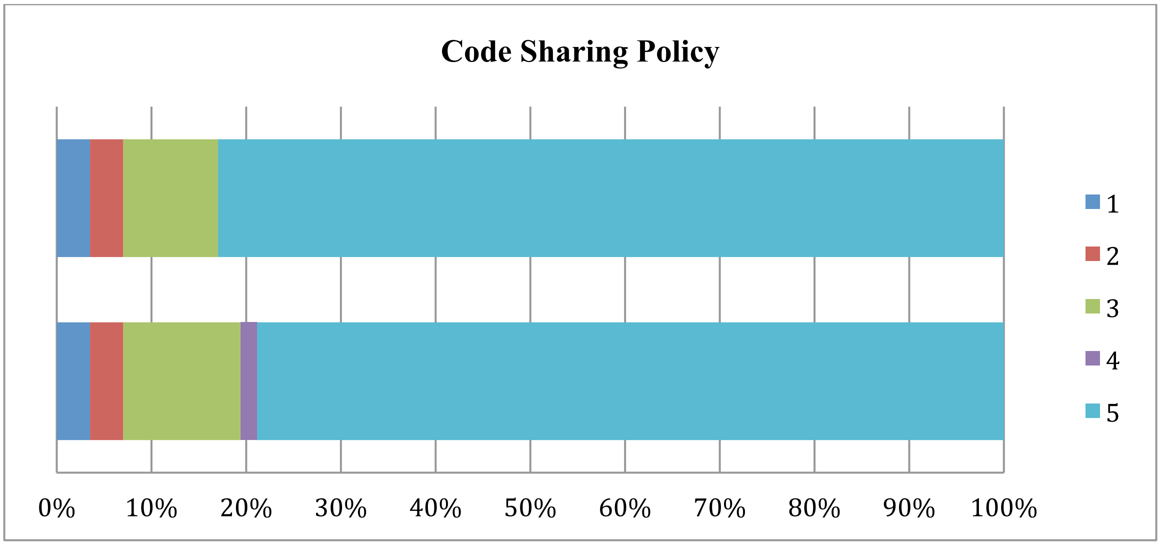
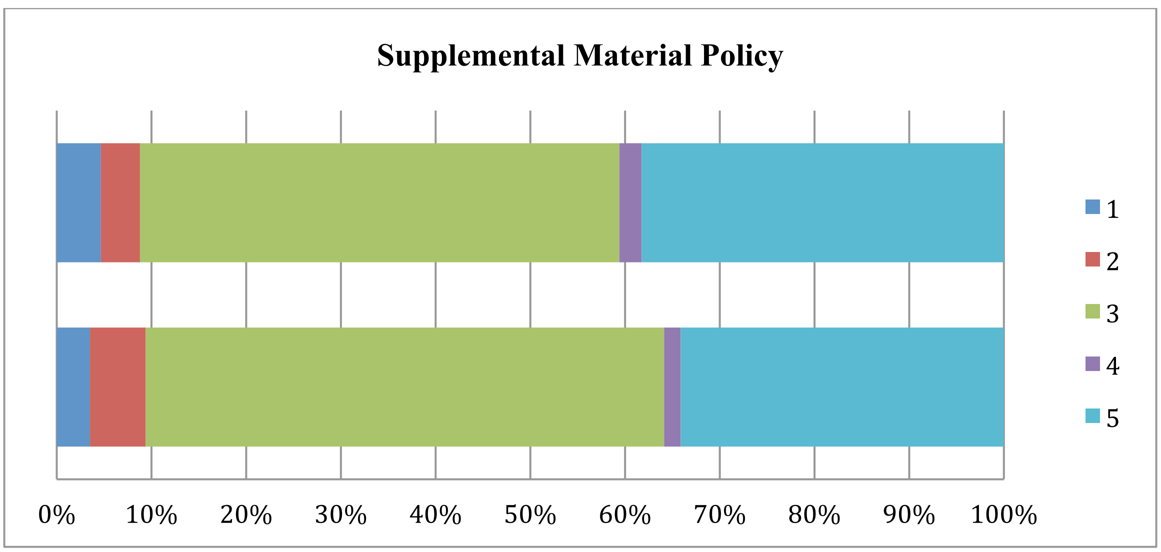



Figure 3 shows very little change in supplemental material policy, excepting an increase in policies that encourage supplemental materials, and a decrease in those that require supplemental materials. The proportion of journals in this study with a supplemental materials policy increased from 62% to 66%.

**Figure 3: Breakdown in Changes in Supplemental Materials Policy, 2011-2012**




Figures 1 through 3 above show that the proportion of journals with a code sharing policy is lowest and those with a supplemental materials policy is highest for both years studied. Just under half the journals in the study have a data sharing policy. The data sharing policy also has the greatest proportion of journals that require sharing in their instantiation of the policy, at over 10%.

Only three journals in 2012 included a qualifier in their code release policy by stating it applied only to articles that employed substantial software. In 2012 eleven journals planned to review supplemental materials submissions, and 69 were explicitly willing to host supplemental materials. In the same year five journals would review data submissions and 10 were willing to host such submissions. Also in 2012, two journals would review code, and two would host code (these were not the same two journals). Table 6 summarizes these findings.

**Table 6: Journal Review and Hosting Policies, 2012**

| Data Sharing Policy (n=64) | 2012 Count | Percent of Total |
|---|---|---|
| Reviewed | 5 | 7.8% |
| Hosted | 10 | 15.6% |
| **Code Sharing Policy (n=36)** | | |
| Reviewed | 2 | 5.6% |
| Hosted | 2 | 5.6% |
| **Supplemental Materials Policy (n=112)** | | |
| Reviewed | 11 | 9.8% |
| Hosted | 69 | 61.6% |

Most journals with a Supplemental Materials Policy were willing to host submissions, and of those with a data sharing policy nearly 16% were willing to host submitted data. Very low proportions, less than 10% in all cases, were willing to review any of data, code, or supplemental materials.

The following three plots, Figures 4 through 6, show the changes from 2011 to 2012 in how journal policy intends the data, code, or supplemental materials to be accessed. The journal can allow readers to contact the author to request the materials or provide them through the journal's website or external repositories. The journals can specify in policy whether the author is intended to share materials with the public or only with other researchers and colleagues. The most important finding is that there has been a shift away from journals accepting data to policy that provides for reader access upon request, with the opposite trend for Supplemental Materials. A potential explanation is that journals requiring or encouraging access to data and code are becoming less willing to host these files, since these can be of widely varying size and complexity. Supplemental materials, however, often do not contain data or node, but are more similar to traditional publications in that they often contain figures and text and are often a traditional .pdf file. A very small percentage of journals, 0.6%, shifted to requiring code upon acceptance, otherwise there was no change to type of code access in journal policy from 2011 to 2012.

This is consistent with the notion of greater adoption of data and code disclosure policies by journals in general. The potential difficulty of journal provision of large datasets and complex code bases, and its relative ease for supplemental material, may underlie the opposing trends of increasing access by request for data and code and decreasing for supplemental materials (Figure 4). This may also explain increases in data disclosure requirements being applied to colleagues only coupled with a decrease for supplemental materials (Figure 6).

**Figure 4: Journal Policy Request Changes by Type, 2011-2012**

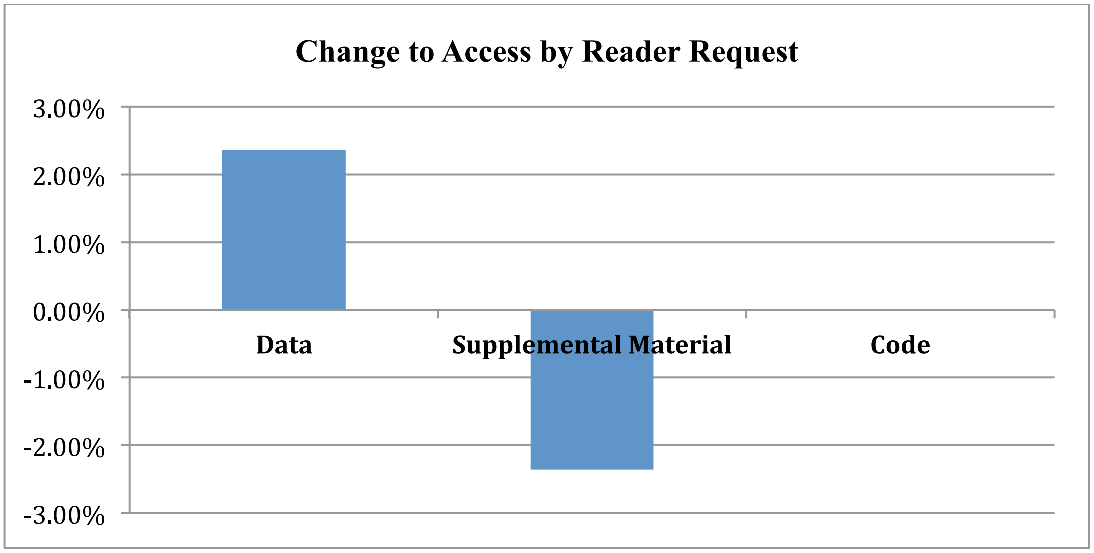



**Figure 5: Journal Policy Submission Requirement Changes by Type, 2011-2012**

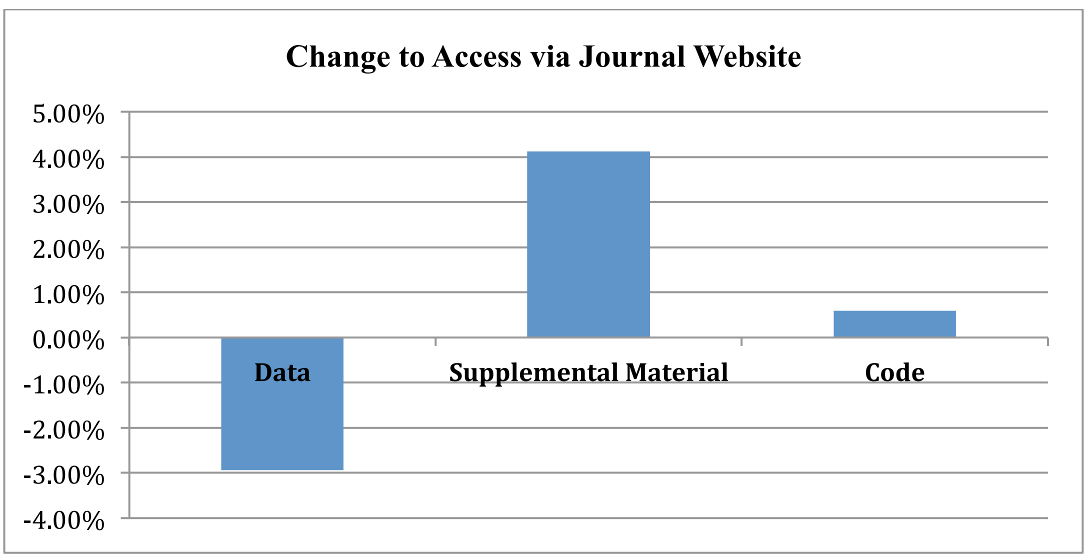
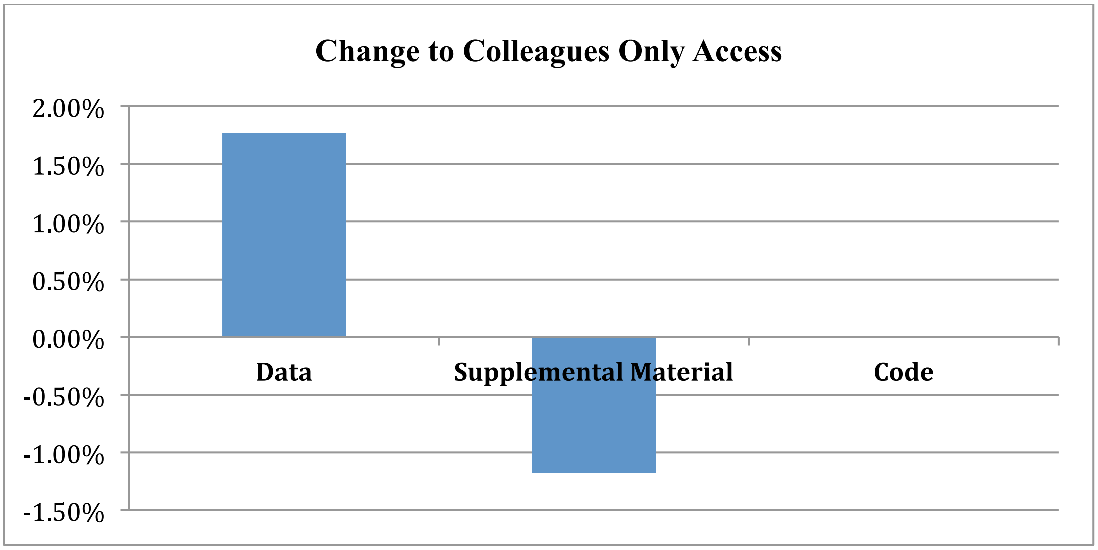


**Figure 6: Journal Policy Sharing Restrictions to Colleague and Researcher Only, 2011-2012**




The higher impact journals appear to be adopting open data and code requirements more readily than lower impact journals. It is possible that higher impact journals are more comfortable increasing the demands on contributing authors, making them more likely to institute open data and code policies. One way to quantify this would be to regress each journal's impact factor, as given by the ISI in 2012, on a binary variable indicating whether or not the journal had a data or code policy. Supplemental Materials policy, since not directly geared toward the sharing of data and code, was not included in the creation of the dependent variable. The binary dependent variable was set to be whether the journal had either a data or code policy since this is a measure of openness, and very few journals that had code policies did not have data policies. Fitting a logistic model revealed a strong linear relationship between the two: a coefficient on impact factor of 0.5496 (p-value = 0.0017), meaning that higher impact factor significantly increases the odds that a journal will have a data or code sharing policy.

It is also illuminating to consider whether journal ownership has an influence on whether or not a particular journal has a data or code sharing policy. One way to measure this would be to regress publisher data on a binary variable indicating the presence or absence of a data or code sharing policy, as was done in the case of impact factor. We constructed a model predicting whether or not a journal has a data or code sharing policy from its publishing house, using the variables described in Table 2 (leaving the "Not-For-Profit Publisher" variable out since all journals are classified by publisher). We included the Impact Factor variable and fit a logistic regression model, and give the coefficient estimates in Table 7, where several variables are significant. At the 5% level all of Elsevier, Wiley, and Scientific Society Publisher are significant, indicating that having one of these publishers increases the likelihood of having a data or code sharing policy. Impact Factor remains a highly significant determinant of the existence of a journal data or code sharing policy.

**Table 7: Regression Coefficients from Predicting Open Data and Code Policies by Publisher and Impact Factor**

| Variable | Coefficient Estimate | Std Error | p-value |
|---|---|---|---|
| Intercept | -2.4600 | 0.7207 | 0.0006 |
| Impact Factor | 0.5271 | 0.1719 | 0.0022 |
| Elsevier | 2.0601 | 0.8342 | 0.0135 |
| Taylor & Francis | 0.2721 | 1.0225 | 0.7902 |
| Macmillan | 9.0718 | 980.7362 | 0.9926 |
| Springer | 1.9021 | 0.8046 | 0.6403 |
| Wiley | 0.3760 | 0.8011 | 0.0176 |
| Scientific Society Publisher | 1.6794 | 0.7529 | 0.0257 |
| Other For-Profit Publisher | 1.2880 | 0.7594 | 0.0899 |

Since Macmillan has three observations in this dataset the associated large standard error is unsurprising. The data and code from this study is available at http://www.stodden.net/JournalPolicies2012/ .

DISCUSSION

These results show a move toward journal involvement in data and code disclosure. Many journals are using their communication policies to support reproducible research, facilitate data and code re-use, and enable broader communication of the output of digital scientific research. Of the journals in this study, 38% have a data policy, 21% have a code policy, and 66% have a supplemental materials policy as of June of 2012. These figures are up from 33%, 17%, and 62% respectively in June of 2011. Data and code sharing is no longer an obscure or eccentric part of the publication process, but has entered the mainstream.

Aside from increasing awareness of data and code disclosure as an integral part of reproducible computational science, an adoption pattern may be discerned from these data. Of the 13 journals that had a change to their code policy, seven were new policies instantiated sometime between June 2011 and June 2012. Of these seven, all were journals that had had pre-existing data policies in place in 2011. It seems code disclosure policies follow on the heels of data disclosure policies in this sample of journals. For the seven journals that adopted new code disclosure policies, four of them also shifted their data disclosure requirements from encouraged to required (the others had no change to their existing data policies). Such a "follow-on" hypothesis is also supported by the greater proportion of journals with a data access policy versus those with a code access policy in 2012, 38% and 21% respectively – data policy appears to be the gateway toward more open policies generally.

Of the journals that had a change in their data sharing policy from 2011 to 2012, eleven adopted a data policy for the first time, nine shifted from encouraging data sharing to requiring it, four shifted from required to encouraged, and four dropped their data access policy (the two remaining journals with a change, shifted to data sharing being implied and to explicitly stating a failure to share data will not affect editorial decisions such as publication). This shows a marked shift toward more stringent data sharing policies, in the course of only one year.

Changes in the policies regarding supplemental materials followed a similar pattern. Of the 36 that had a change in their supplemental materials policy from 2011 to 2012, fourteen journals had instantiated a new policy, six had shifted from encouraged to required and six had shifted from required to encouraged, and seven dropped their supplemental materials policy (two shifted to stated that supplemental materials will not affect editorial decisions and one shifted from an implied policy to one explicitly encouraging the provision on supplemental materials). Supplemental materials policies seem to lead or be in tandem with data policies. Of the 11 new data policies implemented from 2011 to 2012, 5 of those journals had no supplemental materials policy in 2011, but by 2012 nine of the 11 journals with new data sharing policies had supplemental materials policies.

We also ranked the 170 journals in our study regarding their policies about open access to the published paper itself. We were especially interested in whether there was a correlation between journals that are amenable to open access and the availability of data and code through journal policy. We were also interested in whether a "follow-on" effect existed for scholarly object policies, i.e. data, code, and supplementary materials, from open access policy. We saw very little change however in open access policies for our 170 journals from 2011 to 2012. Table 8 details these differences: the net change was an increase in one journal requiring open access publication, one journal shifting away from delayed open access or membership requirements, a net increase of one journal explicitly permitting the posting of the draft of final versions of the paper on the web, and finally a net decrease of one journal requiring subscription.

**Table 8: Changes in Open Access Policy 2011-2012**

| Open Access Policy (n=170) | 2011 | 2012 | Change |
|---|---|---|---|
| Open access | 29 | 30 | 1 |
| Open access with delay and/or journal membership requirement | 73 | 72 | -1 |
| Subscription but authors explicitly allowed to post draft or final | 13 | 14 | 1 |
| Subscription | 55 | 54 | -1 |

With such a small amount of policy shift to 2012 it is difficult to believe that recent or concurrent changes in open access policy are driving changes in data and code sharing policies, but examining the correlation can be instructive as to whether previous open access policy changes might impact a journal's likelihood of adopting open data and code policies. We divided the journals into open access (2012 classifications 1 and 2) and subscription (2012 classifications 3 and 4) and looked at the differences in 2012 data and code policies. If open access journals are more likely to later adopt open data and code policies, we should see significant differences between the two groups in 2012. We classified any mention of data or code policy in 2012 as having a policy. Table 9 shows these differences.

**Table 9: Open Access and Open Data/Code Policies 2012**

| Publication Access | Data or Code Policy | No Mention | Total |
|---|---|---|---|
| Open Access | 42 | 60 | 102 |
| Subscription | 24 | 44 | 66 |
| **Total** | 66 | 104 | 170 |

Although the proportion of open access journals with a data or code policy is greater than the proportion for subscription journals, a chi-square test of independence is not significant for these data ($p=0.44$). These data provide no evidence that an open access policy indicates a greater likelihood of an open data or code policy. As present, access to published papers appears to be a separate issue to reproducible research.

We did find, however, that in June 2012 a full 22 journals or 13% of our sample explicitly permitted the posting of the draft version of the published paper (in practice this is almost always permitted), and five journals explicitly permitted the posting of the final published version of the published article on the authors' websites. These five journals are predominately mathematical and carried no data or code sharing policy.

With two years of data it is difficult to speculate as to the reasons for the substantial increases in journals with code and data access policies, but there are some exogenous policy changes that may be affecting journal policy creation. In January of 2011 the National Science Foundation begin requiring all grant application to include a two-page Data Management Plan, describing the intended availability and archiving of any dataset produced in the course of the research [17]. This may have alerted journal editors to the importance of data and code sharing, although because of the biology focus in this paper many of the articles published in the 170 journals included in this study are more likely to be affected by National Institutes of Health policies rather than National Science Foundation. The genomics world was rocked by flawed cancer research emerging from researchers at Duke University [22]. This culminated in the Institute of Medicine report entitled "Evolution of Translational Omics: Lessons Learned and the Path Forward," that recommended, among other things, code and data disclosure for biomarker tests wishing FDA approval to proceed to clinical trial [23]. It is possible journal data and code disclosure policy may have been affected by this incident.

One measure that may help shed light on the rationale underlying the shift to data and code disclosure policies is the wording in the policy statement. We examined the frequency of the use of the term "reproducibility" or similar terms such as "replication," in journal policy statements regarding data and code. We found that eleven journals of 66 with either a data sharing or code sharing policy specifically referenced these terms when explaining their publication policies in 2012, whereas only four journals did so in 2011 (the eleven include those four, no journal that mentioned reproducibility as a rationale for its policies stopped doing so in 2012). This seems to indicate the importance of reproducibility as a rationale underlying data and code access policies. Of the eleven mentioning reproducible research terms in their policies, all except two had open data policies (six required and two encouraged), and five had required code disclosure policies (one encouraged code disclosure). Of course if a journal does not explicitly mention these terms, it does not preclude reproducibility from being their underlying rationale for implementing data and code access policies. The average impact factor for this group of eleven was 10, much higher than the average impact factor for all journals in this study of 1.82. In fact, the two journals in this group of eleven that do not have data policies have the lowest impacts factors of 0.324 and 0.554.

## CONCLUSION

Pervasive and large-scale computation is changing our practice of the scientific method. Without conveniently available data and code, we risk providing insufficient information for others to replicate published computational results. In this study we sought to understand the role journals are playing in the move toward the open availability of data and code accompanying publishing results. This paper seeks to provide some understanding of the current state of journal policy toward data and code, and how this policy is evolving. We documented 170 journal policies from computational areas in June of 2011 and June of 2012, classifying their data, code, and supplemental materials policies. We hypothesized that open data and code policies are in the process of being adopted more widely, that data policies would lead code policies, and that open access journals would be more likely to have policies making data and code open as well.

We found evidence to support our first two hypotheses, and little evidence to support the third. In June of 2012 38% of the journals in this study had a data policy, 22% had a code policy, and 66% had a supplemental materials policy. This is an increase from June of 2011 when the proportions were 33%, 17%, and 62% respectively. The greatest increase has been in the adoption of new code policies. Of journals that have open data and code policies, they tend to adopt open data policies first followed by open code policies. Surprisingly, supplemental materials policies do not seem to lead data or code polices in a similar way, nor do they appear to crowd out or displace data and code policies. There seems to be no difference in open data and code policy adoption rates for open access versus subscription journals.

The majority of journal titles included in our study have not followed the recommendations of the National Academies Committee that journals publishing articles containing computational findings implement explicit policies that make the underlying data and code available. These recommendations are reproduced in Appendix 1. We find sources of encouragement though due to the increase in the number of data and code sharing policies from 2011 to 2012, even if the proportion is still small. We are also encouraged by the marked increase in code sharing policies over the same period, particularly as code policies seem to follow on where data policies exist.

This study was limited to the journals listed in three classifications by the ISI Web of Knowledge: “Mathematical & Computational Biology,” “Statistics & Probability,” and “Multidisciplinary Sciences” and a handful of additional journals. This selection has a bias toward bioinformatics and life sciences research due to the inclusion of computational biology journals. This was deliberate as data sharing has a strong and long established foothold in bioinformatics areas and greater insights would be gained by studying areas grapping with open science issues for longer periods of time. The Protein Data Bank (PDB) for example was established in 1971 and deposit within PDB is now required for papers describing three-dimensional structures of biological macromolecules. A concerted effort by editors of leading journals in 1998 brought about the pre-publication registration of structure coordinates in PDB [24]. The Human Genome Project promulgated widely agreed upon community standards of data sharing as early as 1996 that established data openness as a norm in the field of genomics [25, 26]. This long history is the exception in data intensive empirical science and was part of the rationale behind including computational biology journals in the study, in order to understand a more mature and more pervasive response to the question of open data. Future work however should expand the sample under study to include other computational fields. Such an expansion would reduce potential bias due to the inclusion of computational biology journals and verify whether the same patterns of policy adoption persist in other areas.

An open question in this study is why several journals reduced or eliminated their data and code sharing requirements from 2011 to 2012. It would be instructive to learn on a case by case basis why this occurred. This would provide information on which policies seem to work best for which fields, and which do not. Understanding the process of deliberation and journal policy adoption could also be illuminating for journals seeking to instantiate policies of their own and researchers studying changes in academic journal policies.

A deeper study on the reasons that journals have adopted or not adopted policies on data and code sharing. This study makes a first cut at modeling this relationship, using impact factor and publishing house as explanatory variables, but research should be carried out using a more extensive set of co-founding variables such as field characteristics, journals size, journal age, frequency of publication, proportion of computational results published in the journal, proportion of computational results publishing in the field, and others. In this research we seek to introduce a novel dataset on journal policy changes and some preliminary analysis and results.

This study does not take into account the enforcement and effectiveness of data and code sharing policies enacted by journals. It documents the state of such policies on journal websites in 2011 and 2012, but does not extends the analysis to effectiveness. An important question is whether the existence of such policies as detailed in this study materially affects the ability to access the data and code that underlies published computational results. This would verify policy enforcement by journals and could illuminate differences in sharing policy effectiveness by field.

The reproducibility of computational results is necessary in our efforts, as scientists, to ensure the publication of reliable findings and better facilitate downstream research. Access to data and code is an essential part of this transition, and this paper makes a contribution to understanding how the role of journal policies evolve and contribute to data and code availability. Data and code sharing is no longer an obscure or unconventional part of the publication process, but has entered the mainstream. Because of its facilitating role in reproducible research, it is of vital importance to understand how journal policy influences data and code availability.

LITERATURE CITED

[1] Donoho, D., A. Maleki, I. Ur Rahman, M. Shahram, V. Stodden, "Reproducible Research in Computational Harmonic Analysis," Computing in Science and Engineering, Vol 11, Issue 1, 2009, p. 8-18. http://www.computer.org/csdl/mags/cs/2009/01/mcs2009010008-abs.html

[2] "Reproducible Research," Special Issue, Computing in Science and Engineering, Vol 14, Issue 4, 2012, p. 11-56. http://www.computer.org/portal/web/csdl/abs/mags/cs/2012/04/mcs201204toc.htm

[3] Stodden, V., I Mitchell, R. LeVeque, "Reproducible Research for Scientific Computing: Tools and Strategies for Changing the Culture," Computing in Science and Engineering, Vol 14, Issue 4, p. 13-17, 2012. http://www.computer.org/csdl/mags/cs/2012/04/mcs2012040013-abs.html

[4] See Yale Roundtable on Reproducible Research, 2009. http://www.stanford.edu/~vcs/Conferences/RoundtableNov212009/

[5] Yale Roundtable Participants, "Reproducible Research," Vol 12, Issue 5, 2010. http://www.computer.org/portal/web/csdl/doi/10.1109/MCSE.2010.113

[6] "Devil in the details," Nature editorial, Vol 470, p. 305-306, Feb. 17, 2011. http://www.nature.com/nature/journal/v470/n7334/full/470305b.html

[7] Hanson, B., A. Sugden, B. Alberts, "Making Data Maximally Available," Science Magazine editorial, Vol 331, Issue 6018, Feb. 11, 2011. http://www.sciencemag.org/content/331/6018/649.short

[8] Merali, Z., "Computational Science:... Error," Nature, Vol 467, Oct. 13, 2010. http://www.nature.com/news/2010/101013/full/467775a.html

[9] Barnes, N., "Publish your computer code: it is good enough," Nature, Vol 467, Oct. 13, 2010. http://www.nature.com/news/2010/101013/full/467753a.html

[10] Buckheit, J., S. Chen, D. Donoho, I. Johnstone, J. Scargle, "About Wavelab," Dept of Statistics Tech. Report EFS-NSF-491, 1995. http://statistics.stanford.edu/~ckirby/techreports/NSF/EFS NSF 491.pdf

[11] R. J. LeVeque, M. J. Berger, et. al., Clawpack Software, www.clawpack.org.

[12] S. Fomel, J. Claerbout, Guest Editors' Introduction: Reproducible Research: Computing in Science and Engineering, vol. 11, Issue 1, p. 5–7, 2009. doi:10.1109/MCSE.2009.14

[13] Hero, A., http://tbayes.eecs.umich.edu/ and http://tbayes.eecs.umich.edu/reproducible_research

[14] King, G., "Replication, Replication," Political Science and Politics, Vol 28, p. 443-499, 1995. http://gking.harvard.edu/files/abs/replication-abs.shtml

[15] National Science Foundation Grant Guidelines, http://www.nsf.gov/cise/cise_dmp.jsp

[16] National Institutes of Health Grant Guidelines, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

[17] NSF Data Management Plan, Jan. 2011. http://www.nsf.gov/bfa/dias/policy/dmp.jsp

[18] National Science Foundation Grant Solicitation, "Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)," http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm

[19] See http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-12-002.html and http://grants.nih.gov/grants/guide/pa-files/PAR-11-203.html for example.

[20] Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council, Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences, National Academies Press, Washington, D.C., U. S. A., 2003. http://www.nap.edu/openbook.php?record_id=10613&page=27

[21] Hanson, B., A. Sugden, B. Alberts, "Making Data Maximally Available," Science Magazine editorial, Vol 331, Issue 6018, p. 649, Feb. 11, 2011. http://www.sciencemag.org/content/331/6018.toc#Editorial

[22] "An Array of Errors," The Economist, Sep. 10, 2011. http://www.economist.com/node/21528593

[23] Institute of Medicine Committee Report, "Evolution of Translational Omics: Lessons Learned and the Path Forward," National Academies Press, Washington, D.C., U. S. A., March 23, 2012. http://iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx

[24] "The Gatekeepers" Editorial, Nature Structural Biology, Vol 5, Issue 3, p. 164, 1998. http://www.nature.com/nsmb/wilma/v5n3.892130820.html

[25] "Policies on Release of Human Genomic Sequence Data" Human Genome Project Information, Oak Ridge National Laboratory. http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml

[26] Ferry, G., J. Sulston, The Common Thread: A Story of Science, Politics, Ethics and the Human Genome, Joseph Henry Press, Washington, D. C., U. S. A., 2002. http://www.nap.edu/catalog.php?record_id=10373

## APPENDIX 1

**Excerpted Principles and Recommendations from the National Academies of Science 2003 book, "Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences," pages 4-14. Reprinted with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, D.C.**
**See http://www.nap.edu/catalog.php?record_id=10613 for the full exposition.**

UPSIDE
“the uniform principle for sharing integral data and materials expeditiously (UPSIDE):”
**Community standards for sharing publication-related data and materials should flow from the general principle that the publication of scientific information is intended to move science forward. More specifically, the act of publishing is a *quid pro quo* in which authors receive credit and acknowledgment in exchange for disclosure of their scientific findings. An author’s obligation is not only to release data and materials to enable others to verify or replicate published findings (as journals already implicitly or explicitly require) but also to provide them in a form on which other scientists can build with further research. All members of the scientific community—whether working in academia, government, or a commercial enterprise—have equal responsibility for upholding community standards as participants in the publication system, and all should be equally able to derive benefits from it.**

DATA AND SOFTWARE
**Principle 1. (Chapter 3) Authors should include in their publications the data, algorithms, or other information that is central or integral to the publication—that is, whatever is necessary to support the major claims of the paper and would enable one skilled in the art to verify or replicate the claims.**

**Principle 2. (Chapter 3) If central or integral information cannot be included in the publication for practical reasons (for example, because a dataset is too large), it should be made freely (without restriction on its use for research purposes and at no cost) and readily accessible through other means (for example, on-line). Moreover, when necessary to enable further research, integral information should be made available in a form that enables it to be manipulated, analyzed, and combined with other scientific data.**

**Principle 3. (Chapter 3) If publicly accessible repositories for data have been agreed on by a community of researchers and are in general use, the relevant data should be deposited in one of these repositories by the time of publication.**

MATERIALS
**Principle 4. (Chapter 4) Authors of scientific publications should anticipate which materials integral to their publications are likely to be requested and should state in**

**the "Materials and Methods" section or elsewhere how to obtain them.**

**Principle 5. (Chapter 4) If a material integral to a publication is patented, the provider of the material should make the material available under a license for research use.**

RECOMMENDATIONS

**Recommendation 1. (Chapter 3) The scientific community should continue to be involved in crafting appropriate terms of any legislation that provides additional database protection.**

**Recommendation 2. (Chapter 4) It is appropriate for scientific reviewers of a paper submitted for publication to help identify materials that are integral to the publication and likely to be requested by others and to point out cases in which authors need to provide additional instructions on obtaining them.**

**Recommendation 3. (Chapter 4) It is not acceptable for the provider of a publication-related material to demand an exclusive license to commercialize a new substance that a recipient makes with the provider's material or to require collaboration or coauthorship of future publications.**

**Recommendation 4. (Chapter 4) The merits of adopting a standard MTA should be examined closely by all institutions engaged in technology transfer, and efforts to streamline the process should be championed at the highest levels of universities, private research centers, and commercial enterprises.**

**Recommendation 5. (Chapter 4) As a best practice, participants in the publication process should commit to a limit of 60 days to complete the negotiation of publication-related MTAs and transmit the requested materials or data.**

**Recommendation 6. (Chapter 6) Scientific journals should clearly and prominently state (in the instructions for authors and on their Web sites) their policies for distribution of publication-related materials, data, and other information. Policies for sharing materials should include requirements for depositing materials in an appropriate repository. Policies for data sharing should include requirements for deposition of complex datasets in appropriate databases and for the sharing of software and algorithms integral to the findings being reported. The policies should also clearly state the consequences for authors who do not adhere to the policies and the procedure for registering complaints about noncompliance.**

**Recommendation 7. (Chapter 6) Sponsors of research and research institutions should clearly and prominently state their policies for distribution of publication-related materials and data by their grant or contract recipients or employees.**

**Recommendation 8. (Chapter 6) If an author does not comply with a request for data or materials in a reasonable time period (60 days) and the requestor has**

**contacted the author to determine if extenuating circumstances (travel, sabbatical, or other reasons) may have caused the delay, it is acceptable for the requestor to contact the journal in which the paper was published. If that course of action is not successful in due course (another 30 days), the requestor may reasonably contact the author's university or other institution or the funder of the research in question for assistance. Those entities should have a policy and process in place for responding to such requests for assistance in obtaining publication-related data or materials.**

**Recommendation 9. (Chapter 6) Funding organizations should provide the recipients of research grants and contracts with the financial resources needed to support dissemination of publication-related data and materials.**

**Recommendation 10. (Chapter 6) Authors who have received data or materials from other investigators should acknowledge such contributions appropriately.**

FINAL STATEMENT

**Universal adherence, without exception, to a principle of full disclosure and unrestricted access to data and materials that are central or integral to published findings will promote cooperation and prevent divisiveness in the scientific community, maintain the value and prestige of publication, and promote the progress of science.**

## APPENDIX 2

These journal titles were obtained from the ISI Web of Science 2011 classifications of "Mathematical & Computational Biology," "Statistics & Probability," and "Multidisciplinary Sciences," including some additional titles.

**Journal Title**

Acta Scientiarum - Technology
Advances in Applied Probability
Advances in Complex Systems
Algorithms for Molecular Biology
American Scientist
American Statistician
Anais da Academia Brasileira de Ciências
Annals of Applied Probability
Annals of Applied Statistics
Annales de l'Institut Henri Poincaré, Probabilités et Statistiques
Annals of the Institue of Statistical Mathematics
Annals of the New York Academy of Sciences
Annals of Probability
Annals of Statistics
Applied Stochastic Models in Business and Industry
Arab Gulf Journal of Scientific Research
Arabian Journal for Science and Engineering
Advances in Statistical Analysis
ASTIN Bulletin
Australian & New Zealand Journal of Statistics
Bulletin of Mathematical Biology
Bayesian Analysis
Bernoulli
Bioinformatics
Biometrical Journal
Biometrics
Biometrika
Biostatistics
BMC Bioinformatics
BMC Systems Biology
British Journal of Mathematical & Statistical Psychology
Canadian Journal of Statistics
Cell
Chemometrics and Intelligent Laboratory Systems
Chinese Science Bulletin
Combinatorics, Probability & Computing
Communications in Statistics - Simulation and Computation
Communications in Statistics - Theory and Methods

Complexity
Computers in Biology and Medicine
Computational Statistics & Data Analysis
Computational Statistics
Comptes rendus de l'Academie bulgare des Sciences
Current Bioinformatics
Current Science
Defence Science Journal
Discrete Dynamics in Nature and Society
Econometrics Journal
Econometric Reviews
Econometrica
Electronic Communications in Probability
Electronic Journal of Probability
Endeavour
Environmental and Ecological Statistics
Environmetrics
Evolutionary Bioinformatics
Finance and Stochastics
Fractals - Complex Geometry Patterns and Scaling in Nature and Society
Fuzzy Sets and Systems
Hacettepe Journal of Mathematics and Statistics
Herald of the Russian Academy of Sciences
IBM Journal of Research and Development
IEEE-ACM Transactions on Computational Biology and Bioinformatics
IEEE Transactions on Information Technology in Biomedicine
IET Systems Biology
Infinite Dimensional Analysis, Quantum Probability, and Related Topics
Insurance Mathematics & Economics
International Journal of Agricultural and Statistical Sciences
International Journal of Bifurcation and Chaos
International Journal of Data Mining and Bioinformatics
International Journal of Game Theory
International Journal of Physical Sciences
International Statistical Review
Interdisciplinary Science Reviews
Iranian Journal of Science and Technology Transcation A - Science
Issues in Science and Technology
Journal of Agricultural, Biological, and Environmental Statistics
Journal of the American Statistical Association
Journal of Applied Probability
Journal of Applied Statistics
Journal of Biological Systems
Journal of Biopharmaceutical Statistics
Journal of Business & Economic Statistics
Journal of Chemometrics

Journal of Computational Biology
Journal of Computational and Graphical Statistics
Journal of Computational Neuroscience
Johns Hopkins APL Technical Digest
Journal of the Korean Statistical Society
Journal of Mathematical Biology
Journal of Molecular Graphics & Modelling
Journal of Multivariate Analysis
Journal of Nonparametric Statistics
Journal of Quality Technology
Journal of the Royal Society Interface
Journal of the Royal Statistical Society Series A - Statistics in Society
Journal of the Royal Statistical Society Series B - Statistical Methodology
Journal of the Royal Statistical Society Series C - Applied Statistics
Journal of the Royal Society of New Zealand
Journal of Statistical Computation and Simulation
Journal of Statistical Planning and Inference
Journal of Statistical Software
Journal of Theoretical Biology
Journal of Theoretical Probability
Journal of Time Series Analysis
Kuwait Journal of Science & Engineering
Lancet
Lifetime Data Analysis
Maejo International Journal of Science and Technology
Materials Science & Engineering R - Reports
Mathematical Biosciences
Mathematical Medicine and Biology - a Journal of the IMA
Mathematical Population Studies
Medical & Biological Engineerinng & Computing
Methodology and Computing in Applied Probability
Metrika
Multivariate Behavioral Research
Nature Genetics
Nature Physics
Nature
Naturwissenschaften
Open Systems & Information Dynamics
Oxford Bulletin of Economics and Statistics
Proceedings of the Estonian Academy of Sciences
Proceedings of the Japan Academy Series B - Physical and Biological Sciences
Proceedings of the National Academy of Sciences of India Section A - Physical Sciences
Proceedings of the National Academy of Sciences of the United States of America
Proceedings of the Romanian Academy Series A - Mathematics, Physics, Technical Sciences, Information Science
Proceedings of the Royal Society A - Mathematical, Physical, and Engineering Sciences

Pharmaceutical Statistics
Philosophical Transactions of the Royal Society A - Mathematical, Physical, and Engineering Sciences
PLoS Computational Biology
Probability in the Engineering and Informational Sciences
Probability Theory and Related Fields
Probabilistic Engineering Mechanics
Progress in Natural Science
Quality & Quantity
R&D Magazine
South African Journal of Science
SAR & QSAR in Environmental Research
Scandinavian Journal of Statistics
Scientific American
Science and Engineering Ethics
Scientific Research and Essays
Science
Statistics and Operations Research Transactions
Journal: Statistical Applications in Genetics and Molecular Biology
Statistics and Computing
Statistics in Medicine
Statistical Methods and Applications
Statistical Methods in Medical Research
Statistical Modelling
Statistica Neerlandica
Statistical Papers
Statistics & Probability Letters
Statistical Science
Statistica Sinica
Stata Journal
Statistics
Stochastic Analysis and Applications
Stochastics and Dynamics
Stochastic Environmental Research and Risk Assessment
Stochastic Models
Stochastic Processes and Their Applications
Survey Methodology
Transactions of the Royal Society of South Australia
Technometrics
TEST
Theory of Probability and its Applications
The Scientific World Journal