

# Journal Policy and Reproducible Computational Research

Victoria Stodden  
(with Peixuan Guo and Zhaokun Ma)  
Department of Statistics  
Columbia University

International Association for the Study of the Commons (IASC)  
1st Thematic Conference on “The Knowledge Commons”  
Thomas More College, Université catholique de Louvain,  
Brussels, Belgium  
September 12, 2012

# Setting

- Scientific discoveries now pervasively *computational*, but standard publication practices do not include the associated data and code.
  - ➔ *A credibility crisis* in computational science: most published results are not *reproducible*.
- Journal publishing requirements are part of the solution.
- *Question: How are journal policies today addressing this issue?*

# Experimental Setup

- Sample selection, computational research:
  - Select all journals from ISI classifications “Statistics & Probability,” “Mathematical & Computational Biology,” and “Multidisciplinary Sciences” (this includes Science and Nature).
  - Delete all journals that have ceased publication (5),
  - $N = 170$ .
- Create dataset with ISI information (impact factor, citations, publisher) and supplement with publication policies as listed on journal

# Data Sharing Policy

	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

	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

	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	64	58	-7

# Review/Hosting Policies, 2012

## Data Sharing Policy (n=64)

---

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=64)

---

Reviewed	11	9.8%
Hosted	69	61.6%

---

# Publishing Houses

	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, 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%
Other Not-for-Profit Non-Society Publishers	22	12.9%



# Predicting Open Data and Code Policies by Publisher and Impact Factor

Variable	Coefficient Estimate	Std Error	p-value
Impact Factor	<i>0.5271</i>	0.1719	0.0022
Elsevier	<i>2.0601</i>	0.8342	0.0135
Taylor & Francis	0.2721	1.0225	0.7902
Macmillan	9.0718	980.736	0.9926
Springer	0.3760	0.8046	0.6403
Wiley	<i>1.9021</i>	0.8011	0.0176
Scientific Society Publisher	<i>1.6794</i>	0.7529	0.0257
Other Not-for-Profit Publisher	1.2880	0.7594	0.0899

# Impact of Open Access Policy

	Data or Code Policy	No Mention
Open Access	42	60
Subscription	24	44

Open Access status doesn't imply a greater likelihood of open data and open code policies.

# The Leaders

# Journals mentioning reproducibility, 2011

Proceedings of the National Academy of Sciences	authors must make materials, data, and associated protocols available to readers.
Biometrical Journal	results reported in the manuscript coincide with results produced by the software code submitted.
Biostatistics	kite-marking
International Journal of Physical Sciences	Materials and methods should be complete enough to allow experiments to be reproduced
Scientific Research and Essays	Materials and methods should be complete enough to allow experiments to be reproduced
Journal of Computational and Graphical Statistics	If an accepted manuscript describes software, authors are expected to submit that software as online supplements
PLoS Computational Biology	results described in the paper must be reproducible when peer reviewers, editors, or readers run the software on the deposited dataset and with the provided control parameters.
Nature	
Nature Genetics	An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims
Nature Physics	
Econometrica	all empirical, experimental and simulation results must be replicable

# Journals with required data policies 2011

Nature	Proceedings of the Japan Academy Series B - Physical and Biological Sciences
Nature Genetics	Lancet
Cell	Science
Nature Physics	Bioinformatics
Proceedings of the National Academy of Sciences	BMC Systems Biology
PLoS Computational Biology	Econometrica
Journal of the Royal Statistical Society Series B - Statistical Methodology	BMC Bioinformatics
Journal of the American Statistical Association	Biostatistics
Journal of Molecular Graphics & Modelling	Stata Journal
Evolutionary Bioinformatics	Algorithms for Molecular Biology
Journal of Computational Biology	Journal of the Royal Statistical Society Series A - Statistics in Society
Journal of Business & Economic Statistics	Journal of the Royal Statistical Society Series C - Applied Statistics

# Journals with required code policy 2011

Proceedings of the National Academy of Sciences

Biostatistics

PLoS Computational Biology

Journal of Computational Neuroscience

Stata Journal

Journal of the Royal Statistical Society Series B - Statistical Methodology

Science

Journal of the Royal Statistical Society Series A - Statistics in Society

Bioinformatics

Journal of Computational and Graphical Statistics

Econometrica

Journal of the Royal Statistical Society Series C - Applied Statistics

# Findings

1. Changemakers are journals with high impact factors.
2. Progressive policies are not widespread, but being adopted rapidly.
3. Close relationship between the existence of a supplemental materials policy and a data policy.
4. Data and supplemental material policies appear to lead software policy.