Evaluating Data Citation and Sharing Policies in the Environmental Sciences

Nicholas M. Weber
University of Illinois Urbana-Champaign
501 E. Daniel Street,
Champaign, IL 61820
nmweber@illinois.edu

Heather A. Piwowar
National Evolutionary Synthesis Center
2024 W. Main Street, Suite A200
Durham, NC 27705
hpiwowar@nescent.org

Todd J. Vision
National Evolutionary Synthesis Center
2024 W. Main Street, Suite A200
Durham, NC 27705
tjv@bio.unc.edu

ABSTRACT
The need to share and cite data is central to a scientific method that depends on verifiable results. Recent events in the field of environmental science underscore the need to hold researchers accountable for their claims and a desire amongst domain practitioners to make findings more widely accessible. The report that follows is a preliminary analysis of the data sharing and citation policies of three types of stakeholders in environmental science research: organizations funding work, journals publishing findings, and repositories archiving primary data. Our aim is to acquire a holistic view of the data sharing policies affecting environmental science researchers, in order to inform studies of the influence of these policies on scientists' data sharing behavior, and ultimately guide development of best practices. Our initial analysis found that an overwhelming majority of funding agencies, repositories and journals fail to provide explicit directions for sharing and citing data. Many policies are vague in their directions as to how data should be shared or archived, and how attribution should be noted for secondary data use. These results point to major gaps in data policy in the environmental sciences.

Keywords
Data Sharing, Data Citation, Policy Perspectives

INTRODUCTION
Journals in ecology and evolutionary biology have recently recognized the need to require sharing, archiving and preserving of raw research data, or risk losing the primary record of their given discipline (Whitlock et al., 2010). Such policies play an important role in affecting the conduct of scientific research, especially in the environmental sciences (Zimmerman, 2003).

Yet, researchers operate in a complex environment where multiple stakeholders may set policy, including home institutions, funding agencies, scientific societies and journals. Relevant policies may change from one study to the next, and there is usually little coordination of policy among stakeholders. Thus, it is of interest to understand the range of policies that researchers are subject to within a field of study. Previous studies have surveyed institutional policies on digital research data practices (Wouters, 2002), investigated data sharing policies in journals (McCain, 1995), and compiled reference lists of funder's open data policies[1].

Our present work supplements these focused investigations with a holistic analysis of relevant data policies including funding agencies, data repositories and journals. We use this data to examine variability in policy among stakeholders, to identify policy gaps, and to identify cases that may suggest best practices for future policy development. We focus on policies relevant to the environmental sciences because of the diversity of stakeholders in the field and the obvious practical benefits of data sharing for the betterment of public policy and social welfare.

METHODS

Funding Agency Policies
Our initial collection of funding agency policies focused on those currently in the Sherpa-Juliet database[1]. We then

ASIST 2010, October 22–27, 2010, Pittsburgh, PA, USA.

[1] Sherpa/ Juliet: http://www.sherpa.ac.uk/juliet/index.php

expanded our survey to include the top 10-25 most acknowledged funding sources[2] for articles found in key ecology, environmental science and evolutionary biology journals. For each of the 53 funding agencies, we recorded the following: *Agency Name, Alias, Country or Territory Represented*, *Date Established*, *Existence of a Data Sharing/ Citation Policy, Directions for Deposit, Archiving Destination, Citation Format, Time Data Must Be Available,* and *Accessibility Expectation*.

Repository Policies

We identified repositories named specifically in journal policies and expanded this list to include consortia partners and related domain archives. For each of the 26 repositories, we recorded the following: *Repository Name, Alias, Affiliated Institution, Main Funding Sources, Date of Establishment, Size of Holdings, Type of Data Expected for Deposit, Journal Affiliation, Restrictions on Deposit, Mutability of Deposit, Directions for Citing Data Held by Rep, Requires Associated Publication, DOI / URI / Accession Number issued.*

Journal Policies

We examined the "guidelines to authors" statement of each journal listed in the Reuters Journal Citation Reports[3] for the following categories: Ecology, Evolutionary Biology and Environmental Sciences. For each of the 307 journals, we recorded the following: *Publisher, Journal Alias, ISSN, Society Affiliation, Impact factor, Subscription Model, Total Citations, Publication Rate, Sharing/Citation Policy, Requested /Required Archiving, Data Required for Peer Review, Place of Deposit Specified, Direction for Citing Datasets.*

FINDINGS

Funding Agencies (53 total evaluated): 44% (23) funding agencies require data sharing in some way; 8% (4) specify the duration for which data must be preserved by the primary investigator; 25% (13) give directions on the type of repository to which data should be deposited 8% (4) provide supplemental funds available for deposition of data; only one gives direction on how such data should be cited.

Repositories (26 total evaluated): 12% (3) data repositories have direct journal affiliations; 12% (3) require associated publication for deposit; 31% (8) give directions how to cite their holdings.

Journals (307 total evaluated): 10% (31) journals request and 4% (11) require data to be archived; 10% (32) give explicit directions where to archive 6% (20) give directions on how to share data; 6% (17) give directions on how to cite data.

[2]As aggregated by ISI Web of Knowledge: http://isiknowledge.com/

[3]Reuters Journal Citation Report/: http://thomsonreuters.com/products

DISCUSSION

Very few stakeholders have data sharing, archiving or citation policies, including less than half of funding agencies, and only about 1 of every 8 journals. Instructions for how to cite were particularly rare.

It would be valuable to measure the effect of applicable policies on the behavior of researchers. Ideally, such a study would also address areas not covered by our current analysis: unwritten policies, policy awareness, and policy enforcement. Institutional policies may also prove informative for future investigations.

Investigators rarely volunteer to share their data (Eysenbach, 2001) but they may do so if required by a funding agency or journal policy, or if made aware of possible benefits such as increased citation rates. Our analysis suggests that key stakeholders could do much more to encourage best practices.

This study exemplifies how content analyses can reveal gaps in policy focus, conflicting constraints on researchers and existing best practices.

This project also represents an experiment in open science. Interested readers are invited to follow this project's future iterations at http://openwetware.org/wiki/DataONE:Notebook/Data_Citation_and_Sharing_Policy to reuse data or share ideas.

ACKNOWLEDGMENTS

This work was sponsored by the DataONE Summer 2010 internship program, funded by INTEROP: Creation of an International Virtual Data Center for the Biodiversity, Ecological and Environmental Sciences, US NSF grant #0753138, and the Data Observation Network for Earth (DataONE) NSF cooperative agreement #0830944. The authors are grateful to Valerie Enriquez, Sarah Walker Judson, Suzie Allard, Maribeth Manoff, Robert Sandusky and Bruce Wilson for their input into this work.

REFERENCES

Eysenbach G, Sa ER. (2001) Code of conduct is needed for publishing raw data**.** BMJ. 21;323(7305):166.

McCain KW (1995) Mandating sharing: Journal policies in the natural sciences. Science Communication 16: 403–431.

Whitlock MC, McPeek MA, Rausher MD, Rieseberg L, Moore AJ. (2010). Data archiving. American Naturalist 175: 145–146.

Wouters P (2002). Policies on digital research data: An international survey. The Public Domain of Digital Research Data. Amsterdam: NIWI-KNAW.

Zimmerman AS (2003). Data sharing and secondary use of scientific data: Experiences of ecologists. PhD Dissertation, University of Michigan, Ann Arbor, MI. http://hdl.handle.net/2027.42/39373