Name/Email: OSU Libraries Scholarly Communication Working Group
Authors: Bonnie Avery, Michael Boock, Sue Kunda, Terry Reese, Janet Webster, Andrea Wirth

Affiliation/Organization: Oregon State University Libraries

City, State: Corvallis, OR

Comment 1 We propose that the federal government require data management plans to accompany funding proposals. Data management plans should include a standard way to estimate the cost for specific stages of data management including description, preservation and archiving, and dissemination. The federal government should provide researchers with a standard way of estimating these costs.

We propose a policy that requires grantees to explain how their data will be described and disseminated. Guidelines should be in place for what types of data can be embargoed or excluded from public dissemination.

Comment 2 It is our understanding that data are facts and not subject to copyright. A dataset is automatically in the public domain unless a contract is put on it; however, an original selection, arrangement or manipulation of data is subject to copyright law. In principle, we believe that federally funded data, whether it is subject to copyright and patent law or not, should be in the public domain, or, at the least made available under a Creation Commons Attribution license (CC-BY). Making digital data available in the public domain or under this type of license allows private and public interests to access and reuse data for commercial and non-commercial purposes as efficiently as possible. We recognize that there need to be exceptions to this rule, in cases where the research data is defined as “sensitive” or where research is on-going for example. The federal government should establish guidelines for these exceptions and require researchers and funding agencies to explain why their data should be an exception in their data management plans. Many restrictions on access to research data can be avoided if the data life cycle is well defined.

We also support efforts to make the code required to recreate the data available in the public domain or under an unrestricted license.

Comment 3 First, federal agencies could make a statement about one area where there is no inherent difference in the type of “data”: publications. A definitive statement that all publications resulting from publicly funded research data must be freely available to the public via a persistent link on the Internet is needed. A federal standard could be set for a reasonable lag between publication and availability in an open access repository which researcher could then quote to publishers when assigning their copyrights, rather than allowing the publisher to make this choice against the public interest -- which is now the case.

Second, all agencies should adopt the NSF precedent of requiring “data management

plans" (DMPs) as part of the application process. Differences between disciplines and types of data should be captured in these DMPs. Further if statements about adherence to or evaluation of the DMP become part of the research reporting process, agencies might improve DMP guidelines to enhance data sharing among disciplines.

Finally, federal funders might provide researchers with a standard way of estimating the cost for specific stages of data/preservation/archiving/curation by dataset type/format in their DMPs. (see comment #6)

Comment 5 Libraries at research institutions such as Oregon State University have a long-term interest in supporting their faculty. If not serving as the repository for institutional research results in their published form, academic libraries serve as the pointers to where these results can be found long after the authors cease to be part of the institution. As concern shifts from disseminating information in publications to dissemination of information about the data behind those publications, the library is uniquely placed to provide leadership in longer term research data management where the end goals are greater sharing and better management of research data as an institutional resource. Library professionals can both provide expertise concerning specific aspects of DMPs (how, when and which metadata standards to apply to research data throughout its life cycle; data repository options, linkage between datasets and published research, etc.).

Comment 6 When applying for federal grants, researchers should be required to submit plans similar to the NSF Data Management Plan. Plans should include projected costs for archiving and making digital data accessible, something that the NSF Data Management does not currently require. Awards should then include funding for the curation and distribution of data.

The federal government may want to adopt a central or regional approach for data preservation similar to what Australia is adopting with their National Data Service or the U.K. with their Digital Curation Center. Both recognize that a centralized approach protects against spending money on redundant infrastructures and inconsistent standards. A national model such as DataOne or a regional model of regional data warehouses makes more sense.

Comment 7 Grant recipients should be required to provide a link to their open data at the completion of their research. The government should track conformance with this requirement. Past compliance would be a tool for agencies to use in determining future awardees. Institutions would encourage their faculty to conform in order to ensure future funding.

Comment 8 Putting a national or regional data repository system in place, adopting metadata standards for organization and description of the data, and requiring dissemination are first steps toward broader and more innovative commercial uses of data. Make it easy to find, understand, and use data that is funded by taxpayers.

Comment 9 The federal government may wish to recommend a standard citation format for data produced with public funds. Data that is made more widely available in long-term digital repositories and that is easily cited according to standards supported by the federal government may encourage publishers to allow citation of data.

Comment 10 Digital data standards for specific scientific data will be largely domain specific. Within domains, there are many examples of community-driven data standards. More generally, the library and archival community has developed and participated in a number of standards based efforts to make data sharing easier. Standards like OAI-PMH (Open Archive Initiative) and RDF profiles for the library represent like efforts geared towards the dissemination and findability of data.

Comment 11 Within the library community, a number of successful standards efforts have effectively produced differing data standards. The primarily characteristics of these groups that have made them successful has been their inclusiveness. Almost universally, the standards processes that have been the most successful have utilized a transparent process that has allowed lots of feedback from the communities. But more importantly, those that work on the standards bodies tend to make up many different stakeholder groups allowing for a larger set of concerns to be addressed.

Comment 12 For starters, Federal agencies could seek to participate in the international community, rather than simply designing their own digital data standards. Within the international community, many countries have been wrestling with digital scientific data for a number of years. What's more, many governments are moving to open their own data to the general public (for example, the UK's recent decision to open public health information). Federal agencies need to resist the need to develop a uniquely U.S. solution and look at the work currently be done within the international community and become an active participant.