

Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research (PSTP request dated 4 Nov 2011 to implement Section 103 of America Competes Reauthorization Act of 2010)

submit electronically to: [digitaldata@ostp.gov](mailto:digitaldata@ostp.gov)

2 January 2012

Comments of Paul Beier, [paul.beier@nau.edu](mailto:paul.beier@nau.edu), School of Forestry, Northern Arizona University, Flagstaff AZ 86011-5018, 928 523 9341

**My qualifications:** I am a wildlife ecologist who has engaged on research related to conservation of mammals, birds, reptiles, and amphibians for over 30 years. I am currently President of the Society for Conservation Biology, the largest professional society of conservation scientists. I am submitting these comments as an individual, but I believe they reflect the informed opinions of many other conservation biologists.

**Scope of my comments:** I confine my remarks entirely to one very simple type of biological data – namely the locations of species occurrences. These data consist solely of species name, date, x-coordinate, y-coordinate, and metadata that will allow users to infer the spatial resolution (precision) of the method used to determine the locations. Henceforth I refer to these as **species occurrence data**. These data (when used in conjunction with other freely available data on land cover, soils, topography, and the like) can be used to map species distributions and identify areas of suitable habitat that have not been surveyed for the species of interest. The data and subsequent analyses would inform management decisions by federal agencies, landowners, and others wishing to promote conservation of these species. Such data will be particularly useful for modeling shifts of species' ranges in response to climate change.

I address this one type of data because this data type is simultaneously important and simple. If the federal government cannot promote sharing of species location data, it will utterly fail to promote sharing of more complex types of data and metadata collected with federal funding.

**My experience related to lack of availability of species occurrence data:** Several experiences lead me to believe that sharing species occurrence data is crucial to advancing the science of conservation biology. I have served on 3 recovery teams for species listed under the Endangered Species Act, and on several panels attempting to make recommendations for other species at risk. In every case, our work was severely constrained by lack of access to existing species occurrence data. In every case most of these data had been acquired with federal funds, and in some cases the data had been acquired by federal agencies. With access to these data, we could have built reliable, empirical models of suitable habitat for the species of concern. Although many researchers voluntarily shared data in response to our requests, some of them no longer had access to the raw location data (e.g., because they had left their previous institution) and others never found the time to comply with our request. The process of request was cumbersome and slow because it required us to find out THAT the data existed before we could request it. In no case were we able to gather all of the relevant data we needed. This situation is inexcusable, and is depriving the public of the benefit of information that has been collected at taxpayer expense.

The following paragraphs are organized around specific questions in the request for information.

**(1) What specific Federal policies would encourage public access to and the preservation of broadly valuable digital data resulting from federally funded scientific research, to grow the U.S. economy and improve the productivity of the American scientific enterprise?”**

I strongly believe that all federal agencies funding research related to biodiversity should require PIs on federally funded research projects to deposit species occurrence data with NBII as a condition of funding. PI's who have not complied with this requirement on a past grant or contract should not be eligible for future federal grants and contracts.

Arizona Game & Fish Department has funded several of my research projects, and has required me (and all funded PIs) to provide location data to their database management system before providing the final payment. Compliance has been almost painless for me as a researcher. The Department fuzzes the data to prevent abuse, and contacts me before sharing the data outside the regulatory agencies.

I have submitted one proposal to NSF under their new guidelines, which require a Data Management Plan as part of the proposal. In its current form, the NSF requirement for a data management plan is a feeble step forward that will do very little to improve data sharing. The data management plan, for instance, does nothing to alert the community of potential users that the data exist, nor does it provide a single access point for data from multiple PIs. It also cannot be monitored and enforced.

**(2) What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders, with respect to any existing or proposed policies for encouraging public access to and preservation of digital data resulting from federally funded scientific research?**

The NBII should allow PI to specify how species occurrence data may be used during an initial period of, say, three to five years. For instance, the PIs could specify that no user can use the data to investigate the particular issues that the PIs designed their project to address. PIs could also specify that no peer-reviewed publications using the data can be submitted during the period of restricted use.

**(3) How could Federal agencies take into account inherent differences between scientific disciplines and different types of digital data when developing policies on the management of data?**

Some species occurrence data are sensitive, in that unscrupulous persons could use them to kill, capture, or harm individual animals or plants. Many agencies (Arizona's Heritage Database Management System, for instance, and other state databases) have already solved this problem using two simple measures: (1) The publicly available data consist only of low-resolution maps with locations "fuzzed" by up to a few km. This provides enough preliminary information for a potential user to determine if the data cover the area of interest to the user. (2) Precise location data are provided only to legitimate requestors who agree to specific terms on use of the data, including agreements not to depict or share precise locations in any way.

**(4) How could agency policies consider differences in the relative costs and benefits of long-term stewardship and dissemination of different types of data resulting from federally funded research?**

For species occurrence data, the costs are miniscule and the benefits are large. I suggest that OSTP might for now require data sharing only for similar types of low-cost high-benefit data. OSTP could use the experience to start to produce reliable estimates of long term costs and benefits that could be used to guide future decisions.

**(5) How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?**

The Society for Conservation Biology (of which I am president) publishes several scientific publications. SCB could work with our publisher to require authors to archive their species location data with NBII. However, if only SCB took this step, some authors would submit elsewhere to avoid this extra responsibility. But a broad consortium of professional societies in ecology (SCB, Ecological Society of America, The Wildlife Society) and a handful of dominant publishers (e.g., Wiley-Blackwell, Elsevier, Springer-Verlag) could create a new culture in which data-sharing is viewed as a responsibility of publishing. I have appointed a Task Force in SCB to investigate how SCB could start a dialogue with our sister professional societies and the publishers of their journals to start to create this culture. It will take years, and there will be strong resistance from some academic PIs, but I believe this is an achievable long-term goal. Again, I think it makes sense to start with low-hanging fruit (e.g., species occurrence data); once the new culture of sharing has been in place for a few years, I think it will become obvious which other types of data to share, and how to share them.

**(6) How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?**

The federal government should fund the NBII at a level that would bear virtually all the costs of preserving and sharing digital data on species occurrences.

**(7) What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?**

For some types of data, ensuring compliance will be difficult, but it should be relatively easy for species occurrence data. Federal funders of biodiversity-related research (NSF, USDA, DOD SERDP, EPA) could require the Data Management Plan in each proposal to list the species for which occurrence data will be collected. Funders should convey this information to NBII, who would need staff persons to track compliance and report non-compliance to all federal funders.

One more drastic measure is worthy of consideration: NBII could identify institutions with a pattern of non-compliant PIs and bar such institutions from future federal grants and contracts. This would motivate universities and other research institutions to monitor compliance.

**(9) What mechanisms could be developed to assure that those who produced the data are given appropriate attribution and credit when secondary results are reported?**

Most scientists are reasonable people and realize that acknowledgments will facilitate future collaboration with other investigators. If asked to acknowledge data providers, most scientists will gladly do so.

**(12) How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?**

I believe NBII already collaborates with GBIF. Such collaboration for species occurrence data should be simple.

Thank you for the opportunity to comment on this important initiative. The Policy Committee and Science & Publications Committee of SCB are also considering your request for information, and will probably provide comments, which may or may not agree with mine. In any event, I would be happy to discuss with OSTP how to move this important initiative forward. I truly believe it is very important to conservation biology..

/s/ Paul Beier