Virginia Tech's University Libraries responds to the Office of Science and Technology Policy Request for Information: Public Access to Digital Data Resulting From Federally Funded Research (http://www.gpo.gov/fdsys/pkg/FR-2011-11-04/html/2011-28621.htm)

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?*

   All digital data resulting from taxpayer-funded research should be freely available to the taxpayer. This will minimize barriers to information that can be used to drive develop and commercialize new products and services by entrepreneurs either working alone, in small businesses, or in large corporations.

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?*

   An agency's requirements for data curation and open access should not undermine any patent rights under current or enacted laws or treaties for funded investigators or intuitions.

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?*

   Individual researchers may not have expertise in the establishment of efficient databases, while information management professionals may not have expertise in the use or form of data across multiple disciplines. Therefore, it would be helpful to develop a core set of baseline standards that are applicable to all disciplines, then develop further discipline-specific standards and guidelines where necessary.

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?*

   Funding agencies should consider these issues when they review research proposals. If the research is likely to result in data that is costly to manage, but offers little value to the research community, then it may be a weak proposal in the first place.

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

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

   Agency research sponsorship should provide adequate funding for research and archival activities (such as that provided by university libraries) to accommodate all data curation and archival requirements of federal funding agencies. Consideration should be given towards the funding requirements for acquiring, processing, and preserving access to research data, including consideration for the cost of bandwidth and disaster plans.

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?*

   Federal agencies (NSF, NIH, NARA, NOAA, DOE, EPA, LOC, NARA, etc.) and other stakeholders need to identify best practices. Requirements are currently too vague to help researchers comply. Once requirements are clear and measureable, it will be easier to verify compliance. Furthermore, it will be easier to verify compliance through systematic approaches. Technical infrastructure components such as persistent identifiers and appropriate licenses represent critical mechanisms through which compliance and verification can be automated thereby reducing costs.

8. *What additional steps could agencies take to stimulate innovative use of publicly accessible research data in new and existing markets and industries to create jobs and grow the economy?*

   Provide free tools and services to visualize this data in ways that are useful to researchers, consumers, voters, business owners, investors, etc. These tools and services could be developed through grants, or by federal institutions.

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?*

   At the institutional level, and organizational level, attribution of data sources should be as important as attribution of other sources. Librarians, funders, researchers, publishers, professional organizations and other stakeholders need to identify citation requirements for repurposed data.

10. *What digital data standards would enable interoperability, reuse, and repurposing of digital scientific data? For example, MIAME (minimum information about a microarray experiment; see Brazma et al., 2001, Nature Genetics 29, 371) is an example of a community-driven data standards effort.*

XML is a good standard for storing the metadata that corresponds with the research data, including instructive information for how to access the data if it has particular platform requirements.

11. *What are other examples of standards development processes that were successful in producing effective standards and what characteristics of the process made these efforts successful?*

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

    Federal agencies should coordinate international digital data standards by working with existing standards bodies such as NISO and ISO.

13. *What policies, practices, and standards are needed to support linking between publications and associated data?*

    One of the most important considerations from a policy, practices and standards is a requirement to use persistent, unique identifiers for publications, data, authors, etc. These identifiers not only bolster the linking of publications and data, but also help foster the re-use and development of new services by people and machines. While there are multiple identifier schemes, at this point, perhaps the most important policy decision would be to require using persistent identifiers instead of relying upon existing mechanisms such as website URLs.