OFFICE OF SCIENCE AND TECHNOLOGY POLICY
Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research

From: C. Carl Jaffe, MD, Professor of Radiology Boston University, carl.jaffe@bmc.org

(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?
Any such effort must be framed by a strategic plan that is step-wise since not-sharing is ingrained in current science behavior and encumbered by legal precedent (e.g. Supreme Court re: William Catalona vs WUSTL decision 2008) With regard to human biologic and tissue resources/data, federal policy should promulgate and encourage pre-procedural human consent documents that enables and encourages, on the model of a health-care directive, the healthcare entity to develop sharing processes (allowing for cost-recovery) open to applications from qualified research entities. Except for organ donation, current legal precedent and policies applicable to healthcare tissues removed from patients lack a cohesive modern basis and treat human tissue as potential marketable property.

(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?
For activities whose funding is more than 50% contributed by public funds, provide guidelines for a uniform period, say 6 months (as is the case with publications), during which the data can be privately withheld by the investigators for exploitation of analysis. With databases that are continuously being assembled without clear endpoint, the principal investigator at the time of initial receipt of public funds should define specific plans for an acceptable release of partial data-sets. Although NIH has stated 'data sharing' policies for extramural research projects that meet a certain cost threshold, little implementation occurs often because no meaningful pathway to ease the burden of data conditioning pre-sharing is funded. Moreover, the policy seems not to be applicable to intramural investigations at NIH. It is not clear whether DOD sponsored healthcare research has defined a research data-sharing policy.

(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?
Any new policy must recognize that data-conditioning (such as de-identifying human data and mapping data fields to common-data elements before sharing) require effort and add cost. Local process development needs to be encouraged like IRBs which should consider whether the data is amenable and worthy of sharing (not all data is worth preserving if it doesn't meet quality standards or have public funding)

(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?
Alternative pathways with different cost requirements need development. Storage costs keep falling and the main cost burden may be pre-conditioning the data and auditing it for quality control. Researchers should be aware that some organizations, federal and private, offer common vocabularies and common data elements that when used at the data acquisition planning greatly ease query retrieval

(5) How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?
Federal and professional encouragement resting on ethical and open-science principals that encourage scientists to recognize their self-interest in participating in self-forming cross-disciplinary teams that speeds confirmation of science postulates when data is accessible. This becomes self-evident in certain branches of science such as genetic 'signatures' that are based on statistical clusters since those conclusions are best verified by challenging them with different tools and models

(6) How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?
Providing different adaptable/adoptable pathways with alternative effort and cost bases that are case appropriate that investigators can choose from. Then publicizing successful model case examples

that can be emulated.

(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?
Compliance that benefits the investigator such as a two-step bonus award such as 1) placement of data in a publically internet available shared container 2) evidence of use-case impact of the data on other researchers

(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?
Recognize the power of Program Announced public-private partnerships and Cooperative Agreements (such as the NIH U01 mechanism). Set goals for numbers of such classes of funding with the funding agencies (including DOD)

(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? Standards for Interoperability, Re-Use and Re-Purposing
Self-monitored by the investigator and reported to their advantage on followup grant applications by the same investigator

(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.
NEMA-radiology cooperation has produced DICOM image standards but pathology would benefit from encouragement. Standards can never be imposed but must be eased in incrementally by advantages accruing to the investigator community

(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?
DICOM because it was needed for clinical care interoperability between manufacturers. Hence the key incentives were driven by clinical healthcare operational needs before they were recognized for their greater research value.

(12) How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?
Moving as fast as possible and trumpeting success when it occurs

(13) What policies, practices, and standards are needed to support linking between publications and associated data?
“Shared data lists” and links to supplementary material is already occurring but hasn’t been fully absorbed or exploited by the science community as of yet. See the success stories by NIH TCGA, NCI-CIP, caBIG and other initiatives