

UC Office of the President

U.S. Office of Science and Technology Policy (OSTP)

Title

UC Comments on Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research (RFC Response: Desirable Repository Characteristics) (Docket ID OSTP-2020-0001)

Permalink

<https://escholarship.org/uc/item/6s53z1g1>

Journal

N/A

Author

UC Office of President

Publication Date

2020-03-17



OFFICE OF THE VICE PRESIDENT – RESEARCH AND INNOVATION

OFFICE OF THE PRESIDENT
1111 Franklin Street, 11th Floor
Oakland, California 94607-5200

March 17, 2020

Dr. Lisa Nichols
Assistant Director for Academic Engagement
Office of Science and Technology Policy
Submitted via email: OpenScience@ostp.eop.gov

RE: Docket ID [OSTP-2020-0001](#) Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research (RFC Response: Desirable Repository Characteristics)

Dear Dr. Nichols:

I write on behalf of the University of California (UC) system with regard to the Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research issued on January 17, 2020.

The UC system is comprised of ten research-intensive campuses, six medical schools, and three affiliated U.S. Department of Energy national laboratories. As a system, UC receives approximately \$6 billion annually in extramural research awards and is the nation's largest academic recipient of federally funded research and other university-based projects. In 2018, UC received more than \$2.95 billion in federal agency research funding.

The UC strongly values open science and applauds this Notice issued by the Office of Science and Technology Policy (OSTP) to solicit feedback and recommendations on approaches for ensuring long-term stewardship of, and broad public access to, data resulting from federally funded research.

We appreciate that OSTP wants to decrease burdens on researchers by setting data repository standards for federal agencies to provide optimization and improved consistency across the federal government's repositories. While UC generally agrees with the OSTP's Draft Desirable Repository Characteristics, we believe that in order to make this policy a success, OSTP should further promote harmonization across agencies and departments and consider the cost to curating and preserving research data. These considerations, provided below, are in addition to the comments on specific aspects of the Draft Desirable Repository Characteristics.

Promoting Harmonization Across Federal Agencies and Departments

UC recommends that OSTP promote harmonized regulatory guidance for data curation, preservation and sharing across all federal funding agencies. Currently, regulatory guidance varies between federal research agencies; this is due in part to the differences in types of data collected and varying practices between scientific disciplines. Improving regulatory alignment between agencies will help to improve greatly the overall research enterprise and reduce burdens and costs for both researchers and institutions.

One concrete way in which OSTP can promote harmonization is by encouraging the use of the Uniform Guidance OMB Circular A-81, section 200.315, definition of “research data,” referenced in this very Notice.¹ One recent example of how this definition is not used consistently is the issuance of NIH’s draft policy. The definition of “scientific data” as proposed by NIH expands beyond the Uniform Guidance definition by stating that scientific data includes recorded information that is “necessary to validate *and replicate* research findings” [emphasis added]. Aligning the definition of scientific data across all federal funding agencies will ensure proper management of scientific data and reduce confusion among the research community.

Cost to Research Data Curation and Preservation

Universities, their research administrators, librarians, and technology specialists are in a good position to advise investigators as they curate and preserve their research data. However, the cost of data curation and preservation is huge, and cannot be fully borne by individual universities. We believe that federal funding agencies must increase their support for the expansions of local and disciplinary data storage capacity to meet the need to maintain data in a usable format.

These costs are, in fact, a critical component of disseminating research and ensuring research quality. Thus, data management, de-identification, curation and preservation costs should be allowed as a direct cost by research granting agencies. We note that the current 26% cap on indirect cost recovery constrains universities’ ability to pay for the infrastructure and additional resources necessary to ensure public access to research results, particularly for biological data collected from medical patients as patient data is understandably subject to strict confidentiality protocols.

Feedback on Desirable Characteristics for All Data Repositories

We appreciate the comprehensive list of “Desirable Characteristics for All Data Repositories” (Section I) that provides information on how optimally to store and promote access to data resulting from federally-supported research.

¹ Uniform Guidance OMB Circular A-81, section 200.315 provides the following definition: “Research data means the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This “recorded” material excludes physical objects (e.g., laboratory samples). Research data also do not include: (i) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and (ii) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”

We particularly note that the recognition of datasets as a citable, authored sets of information will help protect the interests of individual stakeholders. In this regard, we suggest that data repositories provide each dataset with a citation that can be referenced in articles, resumes, etc. for both acknowledgement and for promoting reuse. To help ensure that citations and relationships between outputs (i.e. articles and related data) are indexed, repositories should send data and article relationships to DataCite, a central and open indexer for metadata.

We also note that the following attributes to the list of desirable repository characteristics should be included:

- *Persistent Unique Identifiers*: Repositories should support versioning of the Persistent Unique Identifiers like digital object identifiers, accession numbers, and others.
- *Metadata*: Repositories should implement best practices for standardized vocabularies in the metadata (such as, Crossref Funder Registry).
- *Free & Easy to Access and Reuse*: Repositories should be expected to implement Creative Commons licenses for published datasets.
- *Secure*: Adequate protection against security breach is important to protect the data from bad actors, both internal and external, to the US. Security measures would be different by area of science and should be set and evaluated by experts. We recommend that repositories provide documentation of its practices that prevent unauthorized access/manipulation of data.
- *Provenance*: Provenance tracking of datasets should be machine-readable.

Lastly, we strongly encourage the training for federal agency staff to act as partners with grantees and researchers in developing quality data and data management plans that include the elements laid out in the list of “Desirable Characteristics for All Data Repositories.” Data management is a constantly evolving field and agency partners should have the capacity to collaborate with researchers as data elements change.

Feedback on Additional Considerations for Repositories Storing Human Data (Even if De-Identified)

The access to and sharing of human subjects-related data is governed by a complex, fragmented set of ethical and legal requirements. Frameworks for accommodating these data, at scale, have not been developed. UC appreciates that the draft acknowledges the importance of human subject protections without including a mandate for IRBs to verify or otherwise be a gatekeeper for data sharing and management. UC recommends that the OSTP work across federal funding agencies to provide guidance on appropriate ways to maintain sensitive data, use cases when providing access to others would be appropriate, and language about the need for repositories themselves to have in place mechanisms for preventing or discouraging re-identification of de-identified data. We also ask OSTP to consider issuing guidance on standards for uncontrolled access, de-identification, application of confidentiality policies, consequences of participant withdrawal and ability for a participant to decline data sharing, and how requirements such as the Health Insurance Portability and Accountability Act, the European Union General Data Protection Regulation and other data protection laws apply. This would not only decrease administrative burden on researchers and grantee institutions, but also promote the goal of long-term data maintenance and accessibility.

Thank you for the opportunity to comment on this important issue and we look forward to continued engagement on this issue as further policies and other guidance is developed.

Sincerely,

A handwritten signature in black ink, reading "Lourdes G. DeMattos". The signature is written in a cursive, flowing style with a large initial 'L'.

Lourdes G. DeMattos
Acting Executive Director
Research Policy Analysis & Coordination
Office of Research & Innovation