BILLING CODE XXXX-XX-X

OFFICE OF SCIENCE AND TECHNOLOGY POLICY

Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research

AGENCY: Office of Science and Technology Policy (OSTP).

ACTION: Request for Comments; extension of comment period.

SUMMARY: The White House Office of Science and Technology Policy is seeking public comments on a draft set of desirable characteristics of data repositories used to locate, manage, share, and use data resulting from Federally funded research. The purpose of this effort is to identify and help Federal agencies provide more consistent information on desirable characteristics of data repositories for data subject to agency Public Access Plans and data management and sharing policies, whether those repositories are operated by government or non-governmental entities. Optimization and improved consistency in agency-provided information for data repositories is expected to reduce the burden for researchers. Feedback obtained through this Request for Comments (RFC) will help to inform coordinated agency action.

DATES: The comment period for the request for information published 01/17/2020 at 85 FR 3085, is extended. Comments will be accepted until 11:59 p.m. ET on March 17, 2020.

ADDRESSES: Comments should be submitted online to: OpenScience@ostp.eop.gov. Email submissions should be machine-readable [pdf, word] and not copy-protected. Submissions

should include "RFC Response: Desirable Repository Characteristics" in the subject line of the message.

Instructions: Response to this RFC is voluntary. Each individual or institution is requested to submit only one response. Submission should not exceed 5 pages in 12 point or larger font, and should be paginated. Responses should include the name and organizational affiliation(s) of the person(s) filing the comment. Additionally, to assist in analyzing responses, respondents are requested to indicate the primary scientific discipline(s) in which they work (e.g., life sciences, physical sciences, social sciences) and their role (e.g., researcher, librarian, data manager, administrator). Comments containing references, studies, research, and other empirical data that are not widely published should include copies or electronic links of the referenced materials. Comments containing profanity, vulgarity, threats, or other inappropriate language or content will not be considered.

Comments submitted in response to this notice are subject to FOIA. Responses to this RFC may also be posted, without change, on a Federal website. Therefore, we request that no business proprietary information, copyrighted information, or personally identifiable information (beyond filing name and institution) be submitted in response to this RFC.

In accordance with FAR 15.202(3), responses to this notice are not offers and cannot be accepted by the Government to form a binding contract. Additionally, those submitting responses are solely responsible for all expenses associated with response preparation.

2

FOR FURTHER INFORMATION CONTACT: Lisa Nichols at OpenScience@ostp.eop.gov, (202) 456-4444.

SUPPLEMENTARY INFORMATION:

Background

The Subcommittee on Open Science (SOS) of the National Science and Technology Council's Committee on Science (https://www.whitehouse.gov/ostp/nstc/) convenes more than twenty Federal departments and agencies (hereafter "agencies") that support research and development (R&D). It aims to advance open science and foster implementation of agency Public Access Plans that were developed in response to the 2013 White House Office of Science and Technology Policy (OSTP) memorandum entitled "Increasing Access to the Results of Federally Funded Scientific Research" that called for improved access to data and publications resulting from Federally funded R&D. [For more information on agency Public Access Plans, see https://www.cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html. For more explanation regarding Federally funded research data, see 2 C.F.R. § 200.315(e)(3).] One goal of the Subcommittee's efforts is to improve the consistency of guidelines and best practices that agencies provide about the long-term preservation of data from Federally funded research, including suitable repositories for preserving and providing access to such data, considering agency missions, best practices, and relevant standards. According to OMB Circular A-81, section 200.315, "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." [See:

3

https://www.federalregister.gov/documents/2013/12/26/2013-30465/uniform-administrative-requirements-cost-principles-and-audit-requirements-for-federal-awards#sec-200-315.] These efforts are consistent with and supportive of other Administration priorities, such as the Federal Data Strategy and its associated set of Practices to leverage data as a strategic asset [For more information on Federal Data Strategy Practices, see https://strategy.data.gov/practices/].

In support of its work, the SOS has developed a proposed set of desirable characteristics of data repositories for data resulting from Federally funded research. The proposed characteristics could apply to repositories operated by government or non-governmental entities. They draw from agency experience in developing and supporting data repositories and build on existing information for selecting repositories that agencies developed as part of their public access policies. Through public comment, the SOS aims to refine and develop a common set of characteristics that Federal R&D-funding agencies can use to support their Public Access and data sharing efforts.

These characteristics are not intended to be an exhaustive set of design features for data repositories. Federal agencies would not plan to use these characteristics to assess, evaluate, or certify the acceptability of a specific data repository, unless otherwise specified for a particular agency program, initiative, or funding opportunity. Rather, the set of characteristics is intended to be used as a tool for agencies and Federally funded investigators when, for example, they are:

- Assisting Federally funded investigators in identifying data repositories to use for storing and providing access to research data (e.g., when funding agencies do not host the data and/or have not designated specific repositories for use);

- Identifying specific repositories that a Federal agency might designate for use for particular types of research data resulting from Federally funded research;

4

- Developing Federal agency repositories to store data resulting from Federally funded research;

- Informing external data repository developers and managers of the characteristics desired by Federal agencies for storing and preserving data resulting from Federally funded research;

- Evaluating data management plans that propose to deposit research data in a repository that is not operated by a Federal agency.

Consistent with their Public Access Plans, SOS member agencies have proposed characteristics to help support discoverability, management, and sharing of research data, in a user-friendly manner, consistent with principles becoming widely adopted in the research community to make data findable, accessible, interoperable, and reusable (FAIR). [For information on the FAIR principles, see https://www.go-fair.org/fair-principles.] The proposed characteristics are intended to be consistent with criteria that are increasingly used by non-Federal entities to certify data repositories, such as ISO16363 Standard for Trusted Digital Repositories and CoreTrustSeal Data Repositories Requirements, so that repositories with such certifications would generally exhibit these characteristics. SOS member agencies also anticipate that many repositories without such certifications would exhibit them as well. While the desirable characteristics are intended to be enduring, Federal agencies might update them periodically to reflect changing expectations, rapid evolution of research and technology, and practices related to data management and sharing.

This RFC, released on behalf of Federal agencies that are members of the SOS, aims to solicit public input on proposed characteristics for selecting or developing a repository for managing and sharing data that embody effective management and stewardship over data

resulting from Federally funded research. Feedback obtained through this RFC will help to inform the development of coordinated Federal agency technical and policy guidance on repositories for research data.

Request for Comments:

Federal agencies are specifically requesting public comment on the Draft Desirable Characteristics of Repositories to Consider for Managing and Sharing Data Resulting from Federally Funded or Supported Research, found below. The proposed characteristics include "Desirable Characteristics for All Data Repositories" (Section I), as well as "Additional Considerations for Repositories Storing Human Data (even if de-identified)" (Section II), found below. Note that Federal agencies are subject to additional requirements that must be met for repositories they manage or support, such as considerations of security, privacy, and accessibility.

Response to this Notice is voluntary, and respondents are free to address any or all of the topics listed below and should not feel compelled to address all items:

- The proposed use and application of the desirable characteristics (as described in the "Background" section above)

- The appropriateness of the "Desirable Characteristics for All Data Repositories" (Section I) for data repositories that would store and provide access to data resulting from Federally-supported research, considering:

    o Characteristics that are included

- o Additional characteristics that should be included

- Appropriateness of the characteristics listed in the "Additional Considerations for Repositories Storing Human Data (even if de-identified)" (Section II) delineated for repositories maintaining data generated from human samples or specimens, considering:

  - o Characteristics that are included

  - o Additional characteristics that should be included

- Considerations for any other repository characteristics which should be included to address the management and sharing of unique data types (e.g., special or rare datasets)

- The ability of existing repositories to meet the desirable characteristics

- Consistency of the desirable characteristics with widely used criteria or certification schemes for certifying data repositories

- Any other topic which may be relevant for Federal agencies to consider in developing desirable characteristics for data repositories.

DRAFT Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded or Supported Research

I.   Desirable Characteristics for All Data Repositories

A. **Persistent Unique Identifiers:** Assigns datasets a citable, persistent unique identifier (PUID), such as a digital object identifier (DOI) or accession number, to support data discovery, reporting (e.g., of research progress), and research assessment (e.g., identifying the outputs of Federally funded research). The PUID points to a persistent landing page that remains accessible even if the dataset is de-accessioned or no longer available.

B. **Long-term sustainability**: Has a long-term plan for managing data, including guaranteeing long-term integrity, authenticity, and availability of datasets; building on a stable technical infrastructure and funding plans; has contingency plans to ensure data are available and maintained during and after unforeseen events.

C. **Metadata**: Ensures datasets are accompanied by metadata sufficient to enable discovery, reuse, and citation of datasets, using a schema that is standard to the community the repository serves.

D. **Curation & Quality Assurance:** Provides, or has a mechanism for others to provide, expert curation and quality assurance to improve the accuracy and integrity of datasets and metadata.

E. **Access:** Provides broad, equitable, and maximally open access to datasets, as appropriate, consistent with legal and ethical limits required to maintain privacy and confidentiality.

F. **Free & Easy to Access and Reuse:** Makes datasets and their metadata accessible free of charge in a timely manner after submission and with broadest possible terms of reuse or documented as being in the public domain.

G. **Reuse:** Enables tracking of data reuse (e.g., through assignment of adequate metadata and PUID).

H. **Secure:** Provides documentation of meeting accepted criteria for security to prevent

unauthorized access or release of data, such as the criteria described in the International

Standards Organization's ISO 27001 (https://www.iso.org/isoiec-27001-information-

security.html) or the National Institute of Standards and Technology's 800-53 controls

(https://nvd.nist.gov/800-53).

I. **Privacy:** Provides documentation that administrative, technical, and physical safeguards

are employed in compliance with applicable privacy, risk management, and continuous

monitoring requirements.

J. **Common Format:** Allows datasets and metadata to be downloaded, accessed, or

exported from the repository in a standards-compliant, and preferably non-proprietary,

format.

K. **Provenance:** Maintains a detailed logfile of changes to datasets and metadata, including

date and user, beginning with creation/upload of the dataset, to ensure data integrity.

II.   **Additional Considerations for Repositories Storing Human Data (even if de-**

identified)

A. **Fidelity to Consent:** Restricts dataset access to appropriate uses consistent with original

consent (such as for use only within the context of research on a specific disease or

condition).

B. **Restricted Use Compliant:** Enforces submitters' data use restrictions, such as preventing

reidentification or redistribution to unauthorized users.

C. **Privacy:** Implements and provides documentation of security techniques appropriate for

human subjects' data to protect from inappropriate access.

D. **Plan for Breach:** Has security measures that include a data breach response plan.

E. **Download Control:** Controls and audits access to and download of datasets.

F. **Clear Use Guidance:** Provides accompanying documentation describing restrictions on dataset access and use.

G. **Retention Guidelines:** Provides documentation on its guidelines for data retention.

H. **Violations:** Has plans for addressing violations of terms-of-use by users and data mismanagement by the repository.

I. **Request Review:** Has an established data access review or oversight group responsible for reviewing data use requests.

Sean C. Bonyun,

Chief of Staff,

Office of Science and Technology Policy.

[FR Doc. 2020-04530 Filed: 3/4/2020 8:45 am; Publication Date: 3/5/2020]