



AMERICAN SOCIETY OF HEMATOLOGY

2021 L Street, NW, Suite 900, Washington, DC 20036 **ph** 202.776.0544 **fax** 202.776.0545 **e-mail** ASH@hematology.org

March 10, 2020

2020

President

Stephanie Lee, MD, MPH
Fred Hutchinson Cancer Research Center
1100 Fairview Avenue N, D5-290
PO Box 19024
Seattle, WA 98109
Phone 206-667-5160

President-Elect

Jane N. Winter, MD
Northwestern University
Robert H. Lurie Comprehensive Cancer Center
676 N. Saint Clair Street, Suite 850
Chicago, IL 60611

Vice President

Martin Tallman, MD
Memorial Sloan-Kettering Cancer Center
1275 York Avenue
Howard Building 718
New York, NY 10065
Phone 212-639-3842

Secretary

Robert Brodsky, MD
Johns Hopkins University
Ross Building, Room 1025
720 Rutland Avenue
Baltimore, MD 21205
Phone 410-502-2546

Treasurer

Mark Crowther, MD
McMaster University
50 Charlton Avenue East
Room L-301
Hamilton, ON L8N-4A6
Canada
Phone 1-905-521-6024

Councillors

Alison Loren, MD, MS
Bob Lowenberg, MD
Belinda Avalos, MD
John Byrd, MD
Cynthia Dunbar, MD
Arnold Ganser, MD
Agnes Lee, MD, MSc, FRCPC
Joseph Mikhael, MD, FRCPC, Med

Executive Director

Martha Liggett, Esq.

Lisa Nichols, PhD

Assistant Director for Academic Engagement
National Science and Technology Council Subcommittee on Open Science
Office of Science and Technology Policy
1650 Pennsylvania Avenue, NW
Washington, DC 20504

Comments submitted online to OpenScience@ostp.eop.gov

RE: Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting from Federally Funded Research [FR Doc. 2020-00689]

Dr. Nichols:

The American Society of Hematology (ASH) appreciates the opportunity to provide comments on the Draft Desirable Characteristics of Repositories to Consider for Managing and Sharing Data Resulting from Federally Funded or Supported Research, as proposed by the Subcommittee on Open Science (SOS) of the National Science and Technology Council's Committee on Science. ASH's journals, *Blood* and *Blood Advances*, currently mandate that datasets be accessible by reviewers and editors at the time of paper submission and must be publicly available as of the date of publication. Given ASH's journal policy with respect to data sharing, we support the effort by the SOS to improve the consistency of information that Federal Agencies provide to scientists on the long-term preservation of data resulting from Federally funded research, along with the effort to improve and support the discoverability, management, and sharing of data.

ASH represents more than 18,000 clinicians and scientists worldwide, who are committed to the study and treatment of blood and blood-related diseases. These disorders encompass malignant hematologic disorders such as leukemia, lymphoma, and multiple myeloma, as well as non-malignant conditions such as sickle cell disease, thalassemia, bone marrow failure, venous thromboembolism, and hemophilia. In addition, hematologists are pioneers in demonstrating the potential of treating various hematologic diseases and continue to be innovators in the field of stem cell biology, regenerative medicine, transfusion medicine, and gene therapy. ASH membership is comprised of basic, translational, and clinical scientists, as well as physicians providing care to patients.

In general, the draft desirable characteristics as proposed in both sections I and II of the proposal are reasonable, and ASH appreciates that "Federal agencies would not plan to use these characteristics to assess, evaluate, or certify the acceptability of a specific data repository" since different public sharing solutions may be needed given the various types of research data. If access to data generated by Federally funded research is done appropriately, it will enhance research transparency and accuracy, as well as foster the

reproducibility and reliability of the data. More importantly, it will provide an opportunity to analyze data in new ways that might further enhance scientific discovery and promote collaborative interactions.

The Society has a few suggestions and questions for OSTP to consider as it drafts its final set of characteristics that Federal funding agencies can use when issuing guidance around data sharing and management. First, the draft characteristics in both Sections I and II currently do not address who is responsible for maintaining, updating and disbursing the data once it is deposited and ensuring that use is compliant with requirements (the government repository, the non-government repository, or the investigator/institution). For example, in the case of restricted use agreements and protected health information, how will Federal agencies monitor whether the investigator/institution complies? Second, the Society is concerned that the RFC did not address where the resources will come from to collate, share, and store all these data in a manner compliant with a new policy. Federal agencies should allow for grant dollars to be used to comply with data sharing and storage policies. Please also consider applying any policy changes to newly funded studies and not to ongoing projects that were not designed to deposit data with a compliant repository and do not have budgets to support this work. Third, it would be helpful to clarify that this repository guidance is intended for deposition of primary data collected by federally funded investigators and does not apply to data collected by others, i.e., an investigator, federally funded or not, should never deposit someone else's data. The Society also offers our comments on the following draft data characteristics:

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.

- This category would be better titled “Data Quality.” We also recommend deleting “expert” which is vague and adding instead, “... provide, training for data capture methods, data architecture designed to maximize data quality, automated or manual data cleaning, and/or conduct internal data validity evaluation.”

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.

- ASH recommends rephrasing this description as follows: “Provides broad and equitable access to datasets, as appropriate and consistent with legal and ethical limits required to maintain privacy and confidentiality.” We recommend deleting “maximally open access” since access might be determined by considerations beyond legal and ethical limits required to maintain privacy and confidentiality, such as the requestor (e.g., a foreign government), planned use of the data (e.g. scientifically questionable projects), or prioritization (in the case of limited resources). The requirements under *Request Review* suggest that there will be some review of the request before data are released.

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.

- ASH recommends deleting “free of charge” as this prohibits any cost sharing in providing the data. As noted above, ASH's journals, *Blood* and *Blood Advances*, currently mandate that datasets be accessible by reviewers and editors at the time of paper submission and must be publicly available as of the date of publication. We would also like to highlight the need to ensure Federal funding includes costs involved with making datasets and their metadata accessible, if the intent is to make data free of charge to others. Costs may include creation of the dataset, documentation, review and approval procedures, data storage and access, data security procedures and data maintenance. Requiring that datasets be available free of charge shifts substantial costs and responsibilities to the investigators and data depositors rather than the users. Many datasets will not be of interest to other investigators, and ASH hopes that data access plans can be calibrated to the likelihood of use. A model where data repositories can recoup some of their costs from users provides a sustainable model

that unburdens investigators while providing the best practices in repository management in an efficient manner. This is especially important if long-term sustainability is required beyond the funding period.

- ASH recommends deletion of “and with broadest possible terms of reuse or documented as being in the public domain” as repositories may wish to reserve the right to review data requests to make sure they meet certain criteria.
- Who will make sure that analyses are correct, and attribution is given? Disclaimers should be required if the primary team is not involved in the reanalysis.

The Society also has questions and suggestions related to Section II, Additional Considerations for Repositories Storing Human Data, as follows:

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).

- Restricting access to appropriate use consistent with original consent is critical. Who bears responsibility to ensure that data requestors’ use of the data is consistent with the original consent? These terms are usually delineated in Data Use Agreements or through IRB review, but it does not appear that these oversight mechanisms will be used with data requestors.

C. Privacy: Implements and provides documentation of security techniques appropriate for human subjects’ data to protect from inappropriate access.

- Protection of clinical trial data as it relates to increasing the inclusion of minority patients in clinical trials is especially important. A conclusion from a recently published paper (PMC2990341) about biospecimen repositories was that “Minority blood donors are less likely to participate in biospecimen repositories than Caucasians, though other variables also influence participation. The reluctance of minority donors to participate in repositories may result in a reduced number of biospecimens available for study and a decreased ability to definitely answer specific research questions in these populations.” Mistrust of data use was discussed in the study as a reason for a lack of participation in biospecimen repositories; similarly, ASH is concerned that this same mistrust could be the reason for low participation rates of minority patients in clinical trials focused on therapies that would benefit patients with sickle cell disease or multiple myeloma. It is not just important that repositories of clinical trial data are protected from unauthorized users (i.e., law enforcement), but equally if not more important that these communities trust that their data are protected. The desirable characteristics might also include a description outlining who authorized and unauthorized users are.

Again, ASH appreciates the opportunity to provide comments on desirable characteristics for managing and sharing data from federally funded or supported research and remains available for consultation as the National Science and Technology Council Subcommittee further refines the characteristics. We also call your attention to [comments](#) submitted on January 15, 2020, in response to the National Institute of Health’s Data Sharing and Management Policy, for further reference. Please use Suzanne Leous, ASH Chief Policy Officer, as your point of contact at sleous@hematology.org or 202-292-0258, if you require additional information from the Society on this matter.

Sincerely,



Stephanie Lee, MD, MPH
President