

4 March 2020

Lisa Nichols
Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, DC 20504
openscience@ostp.eop.gov

Re: RFC Response on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research

Dear Dr. Nichols:

We appreciate the many ongoing opportunities for continued dialogue with OSTP and the Administration on how to best to promote openness and sharing – consistent with our commitment to promote sustainable Open Science. We especially appreciate OSTP’s recognition that publishers are a valued partner for addressing these questions.

The International Association of Scientific, Technical and Medical Publishers (STM) is the leading global trade association for academic and professional publishers. It has more than 150 members in 21 countries who each year collectively publish more than 66% of all journal articles and tens of thousands of monographs and reference works. STM supports our members in their mission to advance research worldwide. As academic and professional publishers, learned societies, university presses, start-ups and established players, we work together to serve society by developing standards and technology to ensure research is of high quality, trustworthy and easy to access. We promote the contribution that publishers make to innovation, openness and the sharing of knowledge and embrace change to support the growth and sustainability of the research ecosystem. As a common good, we provide data and analysis for all involved in the global activity of research.

The majority of our members are small businesses and not-for-profit organizations, who represent tens of thousands of publishing employees, editors, reviewers, authors and readers, and other professionals across the United States and world who regularly contribute to the advancement of science, learning, culture and innovation throughout the nation. They comprise the bulk of a \$25 billion publishing industry that contributes significantly to the U.S. economy and enhances the U.S. balance of trade.

STM represents publishers across the entire spectrum of science, technology, medicine and the humanities, and is therefore uniquely positioned to discuss the Desirable Characteristics for All Data Repositories (section I). We look forward to continuing our efforts to partner with OSTP, SOS, and individual Federal agencies on these topics.

STM commends OSTP and the SOS for developing these characteristics, which are broadly consistent with those that we are utilizing in our [2020 Research Data Year](#) and also those supported by international initiatives such as the Research Data Alliance (RDA), in which STM is an active participant. We agree with the proposed use and application of the desirable characteristics, in particular that it would be inappropriate to provide “an exhaustive set of design features” or “use these characteristics to assess, evaluate, or certify the acceptability of a specific data repository.” Data sharing is a rapidly-developing field and being too prescriptive at this point could stifle innovation and reduce competition. In addition, specific fields and groups of practitioners may have different needs from those that could be described for all data repositories. Therefore, this flexibility is key.

STM agrees with the SOS that any proposed characteristics of desirable repositories should be consistent with those broadly accepted in research communities. Such criteria would ideally be the result of collaborative efforts by multiple stakeholders in the scholarly ecosystem and are therefore community endorsed. STM’s own efforts to identify and recommend repositories includes the latter requirement as a central characteristic. The identification of ISO 16363 Standard for Trusted Digital Repositories and CoreTrustSeal Data Repositories Requirements as an exemplar. We also greatly appreciate the explicit mention of the FAIR principles in the background section as a motivator for the specific characteristics. STM has been recognized as a member of the FAIRsFAIR project (<https://www.fairsfair.eu/>) in the European Union, and would welcome the opportunity to bring some of these principles and expertise to support OSTP’s efforts in this area.

With respect to the “Desirable Characteristics for All Data Repositories,” we support each of the characteristics that are included. We would like to highlight in particular the importance of “A. Persistent Unique Identifiers” (PUIDs), and encourage the use of widely used and interoperable types of DOIs rather than the creation of government- or repository-specific ID types. We encourage the SOS to work with the RDA to ensure alignment of these IDs.

One criterion that the SOS may want to consider softening is “D. Curation & Quality Assurance.” Of course, repositories that offer curation services are to be preferred over repositories that do not. However, these services are not yet developed enough or consistently deployed across the repository ecosystem, even among the higher-quality data repositories. Although expert curation and quality assurance (including peer review) are important themes and are desired in all data repositories, the other items within this list are more fundamental to identifying appropriate data repositories.

It might be useful to add to the characteristic list two organizing ideas that are implicit in the set of desirable characteristics but may not be completely evident to agencies and Federally funded investigators using the list. In particular, although many of the listed features are in line with the FAIR principles, it might be useful to explicitly highlight these principles in the list of criteria as they are accepted as an effective means to communicate the desired characteristics of repositories. In addition, as noted above, it would be constructive for many of the characteristics (e.g. PUIDs, metadata, reuse tracking, security and privacy) to utilize community endorsed standards and approaches. In this context, it might be useful to add a characteristic “Aligned with community endorsed standards” to highlight the importance of non-proprietary approaches to many of the issues shared by data repositories.

Finally, we would like to suggest a few additional characteristics for consideration. These potentially could be included in a supplemental list of “Additional Characteristics for Consideration of Data Repositories,” to which “curation & quality assurance” could also be moved. These selection criteria can be used or seen as “nice-to-haves”:

- “Fit to subject”: Subject specific repositories are usually superior to generic ones. Repositories that are built and designed for specific disciplines are better catered to the specific needs and requirements of academic disciplines, and therefore should be preferred over generic repositories.
- “Size and scalability”: Larger repositories are, in general, to be preferred over smaller ones. The larger a database, the more useful it becomes due to network effects (e.g. it allows its users to find comparable datasets, find connections with related research, and prevents data being distributed over different databases).
- “Mirroring”: To keep data stored safely, repositories should maintain mirror sites, preferably over different geographical locations.

With respect to the feasibility of the proposed list of desired characteristics, we believe this to be a reasonable list that most responsible and appropriate data repositories for agencies and researchers would be able to meet the set of characteristics for. However, the degree to which an individual repository addresses each of the desired characteristics will vary significantly. This remains a key reason to maintain the list as guidance, rather than as requirements. The list is also generally consistent with those used by several certification schemes, as well as supported by the wider scholarly ecosystem.



The global voice of scholarly publishing

A significant challenge going forward will be to support and guide researchers and federal agencies towards the most appropriate repositories to meet their data sharing needs. This RFC, and the ongoing efforts by OSTP and the SOS to support data sharing are an excellent step in the right direction. Such efforts will need to be coordinated across universities, non-federal funders, publishers, scholarly societies, and others who engage in and support the American research enterprise. Publishers stand ready to work with NSTC, OSTP, and Federal agencies on all of these issues going forward, and welcome additional opportunities to engage and collaborate.

Very truly yours,

A handwritten signature in black ink that reads "Ian Moss". The signature is written in a cursive style and is positioned above a horizontal line that extends to the right.

Ian Moss
CEO