16 November 2018

STM Response to the Request for Comments on the Cross-Agency Priority Goal: Leveraging Data as a Strategic Asset: Phase 2

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 members include non-profit scientific and scholarly societies, commercial publishers, and university presses who work collectively to ensure broad access to and use of the latest scientific and scholarly information. The majority of our members are small businesses and not-for-profit organizations, who represent tens of thousands of publishing employees, editors and authors, 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 $10 billion publishing industry that contributes significantly to the U.S. economy and enhances the U.S. balance of trade.

STM welcomes this opportunity to respond to the "Request for Comments on the Cross-Agency Priority Goal: Leveraging Data as a Strategic Asset: Phase 2," as published on October 17, 2018 in 83 FR 52379, Docket Number: USBC-2018-0017-0001. As requested, we have indicated which questions we are addressing in each part of our response.

Frameworks for organizing the practices (question 1)

The draft practices are organized around five broad objectives, each of which is a reasonable and laudable goal. At the same time, these broad objectives seem to interact and overlap significantly, and many of the draft practices could be categorized under several of the objectives. In addition, several of the objectives would be best seen as permeating all of the practices. For example, from the perspective of a stakeholder organization, honoring stakeholder input and leveraging partners should be included in all practices of the Federal Data Strategy, and not simply limited to some of the practices listed.

Therefore, we recommend that the practices instead be organized as suggested by question 1 according to the stage of that data lifecycle and the government's role in the data's use to which it applies.

Additional practices (question 2)

The International Association of Scientific, Technical and Medical Publishers (STM) shares the administration's goal to leverage data and information to enable innovation and economic growth.

We applaud the efforts of the Department of Commerce, the Small Business Administration, and the White House Office of Science and Technology Policy to harmonize and improve data practices across government.

However, we remain concerned that the Federal Data Strategy is not as explicit as it could be in defining the scope and meaning of "federal data" or "government data."

We therefore strongly recommend that a practice be added that encourages agencies to clearly distinguish between "federal data" and "nongovernment data" and ensure that the latter are handled in a manner than respects intellectual property and other private sector rights. Under the categorization proposed with the draft practices, this would be consistent with the objectives both to "promote efficient use of data assets," and to "honor stakeholder input and leverage partners." In the alternative categories proposed in Question 3 of the request for comments, this would fall under "data creation, collection, or acquisition."

In particular, "federal data" should be limited to "programmatic, statistical, and mission-support data" that originates with and therefore belongs to the federal government. This could be data collected directly by government agencies; data collected by a private sector agency under the direction of or contract with an agency; or data which is purchased under a license that provides the government with explicit rights to use and distribute the information as envisioned in the Federal Data Policy. Information and works that may be acquired by the government, but which are subject to private sector intellectual property rights, should not be considered "federal data" and certainly should not be subject to the same dissemination or other requirements that may be applied to data created by the government. In addition, information that is not data at all, but is instead content, should be outside the scope of the Federal Data Strategy. These include, but are not limited to, journal articles and books that are published by STM member organizations.

The lack of clarity in the current principles and draft practices risks undermining the market mechanisms that enable broad access to and use of information products that relate to or are derived from government data, as well as the economic benefits that derive from commercialization. Ensuring that agencies clearly distinguish between "federal data" and "nongovernment data" will protect intellectual property and ensure the greatest possible impact of the data assets on the public and the economy.

It would also be helpful to include a practice related to ensuring proper attribution and linking of data assets, including reference to the original owners of data, as appropriate. Although this is alluded to in draft practice 22, separating out the practice of attribution and linking is consistent with best practices in the field. Such a practice would include the use of established data citation standards that ensure that the provenance of data can be tracked, that data ownership if applicable is registered, that due reference can be given to the generators, collectors, curators and stewards of the data (ORCID), and that links are in place to related data publications (SCHOLIX) so the context of the data is available.

In our submission to Phase 1, STM recommended several additional practices which could be useful additions to the draft list: partner with existing data providers for the identification of standards and best practices for data management and the interoperability of data repositories; create clear rules for citation, modification and privacy; improve links between data and applications of that data; facilitate cyber infrastructure; and advance policies and funding to ensure the long-term

sustainability of data archives. Some of these overlap with or may be implicit in items proposed in the draft practices list, but could also stand as their own practices in the Federal Data Strategy.

Necessary edits/clarifications to practices (question 4)

As mentioned above and in our earlier submission, preserving and respecting IP rights is a key component of exercising responsibility with respect to data stewardship and governance. We appreciate that several of the draft practices, including 22, 44, 46, and 47, implicitly acknowledge the importance of the private sector in creating, curating, and preserving federal data. We strongly support these practices. Below we recommend additional clarifying language in the descriptions of these practices to ensure they fulfil this purpose. In addition, we recommend that the practices throughout emphasize the importance of working with the stakeholder community and leveraging existing resources, where possible. A specific suggestion to this end is made for draft practice 31. Our recommendations for edits to specific practices follow.

Draft practice 22, "Explicitly Communicate Allowable Use," is a critical practice, particularly for data that may be available from the government but which is not "federal data" as described above. Preserving and respecting IP rights is a key component of exercising responsibility with respect to data stewardship and governance. We greatly appreciate that the practice would be to "explicitly recognize any applicable intellectual property rights," but we are troubled by the use of the term "federally-hosted data," which could be interpreted broadly to include a wide variety of information which the government may have licensed but which the government may not have rights to further distribute. We recommend that this practice be revised to clarify that the Federal Data Strategy applies to "federal data," which is limited to that data that originates with the federal government.

Draft practice 44, "Leverage Partnerships," is a critically important practice. The Federal government has a unique opportunity to work in partnership with stakeholders in the public and private sectors to efficiently deploy scarce resources in support of data communication. Bringing together diverse sources of data, including government data, has the potential to create new opportunities and synergies that best leverage the federal investment. Direct government competition could undermine the development of a healthy community in which to share data. It might be useful to clarify that this practice is intended to support the development of private sector solutions, together with draft practice 46, to "innovate with partners."

Draft practice 46, "Innovate with Partners," is a valuable practice that appropriately emphasizes the economic benefits that derive from commercialization, and the potential of partnerships ensure the greatest possible impact of the data assets on the public and the economy. The practice should be clarified to encourage agencies not only to review competitions and collaborations, but to establish collaborations in the first place.

Draft practice 47, "Honor Proprietary Interests," is a practice that should be infused throughout the Federal Data Strategy to make sure that agencies comply with both the spirit and the letter of titles 17 and 35 with respect to intellectual property, as well as the treatment of IP under the Bayh-Dole Act of 1980. It might be helpful here to emphasize that not only should this practice apply to "commercially-licensed information and data," but to any information that may be subject to intellectual property rights. This could be achieved by adding a clause to that

effect as follows (underlined text added): "Protect proprietary interests related to commercially-licensed information and data <u>and information subject to Titles 17 and 35 with respect to intellectual property</u>, to honor business investments and promote economic growth."

Draft practice 31, "Leverage Data Standards" could more explicitly mention the importance of using existing standards where they exist. These include the Research Data Alliance (RDA), the National Information Standards Organization (NISO), CrossRef, DataCite, Opportunities for Data Exchange (ODE), and APARSEN, among others. Such collaborative approaches provide the best way forward towards broad access to and preservation of digital data. While draft practice 44 would "leverage partnerships," it is important that the government prioritize the use of existing standards to the greatest extent possible.

Draft practice 23, "Leverage Buying Power," is an appropriate management goal in the context of ensuring that the federal government is using its buying power to ensure that the procurement process is meeting federal goals and objectives. With that in mind, it may be useful to expand the purpose of this practice to emphasize that the goal of cost cutting should not be put above addressing the evolving needs of agencies with respect to data procurement. As data becomes ever more important to government operations, it could be that additional resources are needed to meet the unique needs of various agency customers.

In conclusion, we greatly appreciate the opportunity to submit these comments, and look forward to continuing to contribute to the development of the Federal Data Strategy, as articulated draft practice 40, "Assess the Needs of Stakeholders," which directs agencies to routinely engage with stakeholders for feedback and improvement. We look forward to continuing engagement as the strategy is deployed across the agencies and stand ready to collaborate on data strategies in the future.

Very truly yours,

Michael Mabe
CEO STM