July 3, 2019

STM Response to the Request for Comments on the Cross-Agency Priority Goal: Leveraging Data as a Strategic Asset: Phase 3 – <u>Federal Data Strategy Action Plan Year 1</u>

The International Association of Scientific, Technical and Medical Publishers (STM) is the leading global trade association for academic and professional publishers. It has over 150 members in 21 countries, who each year collectively publish nearly 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. Our members, the majority of which are small businesses, 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 supports publishers in their mission to advance research worldwide, including through the dissemination and use of data. STM publishers serve society by supporting innovation, openness and the sharing of knowledge. One way in which we do this is by developing standards and technology to ensure the information we share is of high quality, trustworthy and easy to access. It is with this perspective in mind that we welcome this opportunity to respond to the "Request for Comments on the Cross-Agency Priority Goal: Leveraging Data as a Strategic Asset: Phase 3" as published on June 4, 2019 in 84 FR 25730, Docket Number: USBC-2019-0001.

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 develop and implement the Federal Data Strategy, particularly through the Year 1 Action Plan. We believe that the Action Plan highlights the appropriate steps to advance the government's goals and public interest in achieving greater access to data.

We would like to offer our community's expertise and knowledge in the handling and sharing of research data as the government moves ahead with the action plan, especially in view of the ambitious deadlines that have been set for implementation. Throughout the academic and research community, which includes but is not limited to scholarly publishers, there have been huge strides in enabling and expanding the sharing and use of data. In particular, communities have developed infrastructure, standards, and policies to enable data sharing, including the use of Digital Object

Identifiers (DOIs) through CrossRef and DataCite, interlinking protocols developed in frameworks like SCHOLIX, and efforts to improve metadata. These could be harnessed to support the Federal Data Strategy. The government should utilize existing standards and practices to ensure that its efforts to better leverage federal government data are aligned and consistent with efforts within the broader data community for non-governmental data. We appreciate that this is articulated in Practice 20.

While we support the goals and approach taken thus far with the Federal Data Strategy, we remain concerned about the risk inherent in the continued lack of definition around the terms to which the federal data strategy applies. The Action Plan references "federal and federally sponsored data" and the terms "government data" and "data assets" are also used in various places. Definitions of these terms would provide clarity for agencies that are preparing implementation plans, as would clarifications as to which operational data is in scope. We recommend that definitions focus the scope of the Federal Data Strategy on the "programmatic, statistical, and mission-support data" that originates with, and therefore belongs to, the federal government. Information and works that may be acquired or maintained by the government, but which are subject to private sector intellectual property rights, such as journal articles, books, or proprietary datasets, should not be subject to the same dissemination or other requirements that may be applied to data created by the federal government. If this is not the case, the government 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 private sector investment and commercialization. Carefully defining the scope and meaning of "government data" will ensure the greatest possible impact of the data assets on the public and the economy.

A clear articulation of "government data" would be consistent with the OPEN Government Data Act title of the Foundations of Evidence Based Policymaking Act of 2018, which calls for consideration of intellectual property rights in setting federal data policy, and the previously finalized Practices of the Federal Data Strategy, which call for recognition and safeguards of these rights (e.g. in Practice 31). Such consideration is also consistent with government practices and principles for the use of intellectual property across the government, including practices of the various agency libraries, and the treatment of IP under the Bayh-Dole Act of 1980.

With these two points in view – interest in leveraging existing data sharing efforts, and a need to protect intellectual property and other proprietary interests – we offer additional input on the specific actions listed in the Year 1 Action Plan, focusing on the five areas for which specific and responsive comments were requested.

1. Identify additional actions needed to implement the Federal Data Strategy that are not included in the draft Action Plan and explain why.

We do not believe that there are additional actions needed for the Action Plan, but only edits and clarifications to a few of the 16 actions listed, which we detail below.

2. Identify additional actions that would align with or complement ongoing Federal data initiatives or the implementation of new legislation, such as the Foundations for Evidence-based Policy Making Act of 2018 and explain why.

Consistent with the OPEN Government Data Act title of the Foundations for Evidence-based Policy Making Act of 2018, one of the fundamental actions that must be undertaken to implement the Federal Data Strategy is to clearly define and delineate what exactly is included in government data assets, and what is not. With respect to the Federal Data Strategy, we recommend clear differentiation between "federal and federally sponsored data," specifying what these two categories encompass, and the expectations and obligations for each. This would align with efforts to coordinate implementation activities and educate data officers about their responsibilities within the Federal Data Strategy.

3. Identify any actions in the draft Action Plan that should be considered for omission and explain why.

The 16 articulated Actions are all important and useful for the implementation of the Federal Data Strategy, and all should be pursued in the Action Plan.

4. For each action, provide any edits and additional detail to ensure that they accurately and effectively describe needed activities, responsible entities, metrics for assessing progress, and timelines for completion.

Action 1: In the guidance provided by the OMB Data Council on the implementation of the Federal Data Strategy, the Action Plan should clarify that this guidance would include defining the scope of the strategy. In particular, guidance should provide that requirements under the Federal Data Strategy be limited to "programmatic, statistical, and mission-support data" that originates with and therefore belongs to the federal government.

Action 2: In addition to providing career paths and career training broadly for personal development, the training catalogue should emphasize the need for training and credentialing on the legal, ethical and licensing requirements to ensure appropriate handling of data. This Action should work in tandem not only with Action 14, but with the implementation of Action 3.

Action 3: Coordinating with non-government stakeholders, especially academic and professional associations, as articulated in this Action is key to achieving the goal in the short timeframe provided. Much of the existing work of research communities and frameworks used by publishers in data sharing could be brought to bear here. Also, in addition to Artificial Intelligence, GSA should consider in its fitness for use assessments the potential use of Federal Data for Text and Data Mining applications and consider in the framework how Federal agencies will respond to these requests consistent with ethical standards, licensing requirements, and protection of privacy and intellectual property(also to be considered in Action 6).

Action 5: The repository tools and resources should include guidance to inform data officers on respect for and protection of intellectual property rights. This should be scoped both to provide support for agencies in avoiding undermining intellectual property rights and to support intellectual property rights to incentivize the development of tools that use government data. Stakeholders, including STM members, have been actively engaged in gathering resources, tools and information sources that could be helpful to the government's efforts, Worthing through the Research Data Alliance and other community groups. GSA could be more efficient by working with STM members and others to develop a list of resources to populate the Repository of Federal Data Strategy Resources and Tools

Action 7: Making sure that the use and dissemination of data respects confidentiality and privacy requirements is critical, and so is the protection of intellectual property rights. In addition to helping users "understand whether the data is collected under a pledge of confidentiality or privacy," the metadata for data shared by agencies should also include any licensing requirements or restrictions.

Action 9: "Intellectual property rights" should be added to the list of considerations for protection of "Federal data and models" which currently includes "safety, security, privacy, and confidentiality."

Action 14: Well curated datasets and metadata are prerequisites for making data discoverable, usable and interoperable/machine-readable. "Data curation" should be added to the list of skills required, to make explicit the importance of data curation to the success of the Federal Data Strategy. Mention of this critical skill could be added to the workforce requirements. One way to highlight this critical skill would be highlighting it in step 1 as follows: "Identifying critical data skills for the agency (and relevant sub-units) in the areas of analysis and evaluation, data management, *data curation*, and privacy protection;"

5. For each action, provide information about the implementation resources necessary to ensure success of the action steps.

The Action Plan is an ambitious strategy, especially in the 12-month timeframe provided. By leveraging non-government stakeholder knowledge through use of existing data standards (Practice 20) and partnerships (Practice 36), the need for additional resources could be kept to a minimum.

STM and its members welcome the opportunity for continued engagement and potential collaboration in advance of these goals.

Sincerely,

Michael Mabe
CEO STM