November 16, 2018

Submitted Electronically

Ms. Karen Dunn Kelley
Under Secretary for Economic Affairs
U.S. Department of Commerce
1401 Constitution Avenue, NW
Washington, DC 20230

Re: Leveraging Data as a Strategic Asset Phase 2 Comments, Docket USBC-2018-0017

Dear Under Secretary Kelley:

ACM, the Association for Computing Machinery, is the world's largest and longest-established association of computing professionals, representing approximately 50,000 individuals in the United States and 100,000 worldwide. ACM is a non-profit, non-lobbying and non-political organization whose U.S. Technology Policy Committee is charged with providing policy and law makers throughout government with timely, substantive and apolitical input on computing technology and the legal and social issues to which it gives rise.

On behalf of the Committee, and in response to the Bureau of the Census' ("Bureau") October 17, 2018 request for public comment in Docket USBC-2018-0017 (83 FR 52379), I am pleased to submit the following initial recommendations for your consideration in connection with refinement of the National Data Strategy ("Strategy"). We also look forward to sharing further observations and comments early next year in Phase 3 of this proceeding.

ACM's U.S. Technology Policy Committee broadly urges the Department of Commerce and the Bureau to assure that the Strategy aligns with the "FAIR" Guiding Principles for scientific data management and stewardship to assure that information is "findable, accessible, interoperable, and re-usable" (*see, e.g., Scientific Data* Volume 3, Article 160018 (2016) (www.nature.com/articles/sdata201618). Among those Principles, we particularly urge the Bureau to specially support the ubiquitous adoption of:

- Metadata models that incorporate semantic linkages and granularly establish key terms and practices to describe types of data and how they are used. This goal will be complex, difficult and time consuming to meet at scale. We urge the Bureau, however, to

ACM U.S. Technology Policy Committee
1701 Pennsylvania Ave NW, Suite 200
Washington, DC 20006

+1 202.580.6555
acmpo@acm.org
www.acm.org/public-policy/ustpc

look to the medical field for potential "templates" in the form of current technologies such as SNOMED and UMLS, which incorporate and benefit from shared, structured ontologies and vocabularies;

- Widely embraced open interoperability standards (*e.g.,* REST, JSON, OAUTH) to permit the automated processing of shared government data through well-understood technologies that will allow programmers to create robust applications quickly; and

- Both broad principles and specific practices to maximize the effective use of data as detailed in the attached *Appendix*.

Thank you for the opportunity to participate in this proceeding. Should you have any questions regarding this submission, or if our member experts can be of further assistance, please contact ACM's Director of Global Policy & Public Affairs, Adam Eisgrau, directly at the number and email address below.

Sincerely,

James A. Hendler, Chair

cc: Mr. William Hawk, U.S. Bureau of the Census

ACM U.S. Technology Policy Committee
1701 Pennsylvania Ave NW, Suite 200
Washington, DC 20006

+1 202.580.6555
acmpo@acm.org
www.acm.org/public-policy/ustpc

U.S. Technology Policy Committee

Appendix

RECOMMENDATIONS REGARDING EFFECTIVE DATA USE

To maximize the effective of use of data, the ACM U.S. Technology Policy Committee recommends revision of the National Data Strategy to:

- Define **incentives** to motivate grassroots, bottom-up efforts to provide metadata, access rights, and documentation required for sharing. Tools that exploit metadata can provide such incentives, but must be chosen strategically so that they interoperate rather than selected separately on a project-by-project basis.

- Promote design for **flexibility** to maximize opportunities for data to be reused by multiple groups. We specifically recommend designs that:

    - Flexibly permit each data set to be uploaded, tagged and stored along the lines naturally suggested by the data, rather than being forced into predefined schemata;

    - Facilitate transformations from one schema to another, particularly for data providing high value; and

    - Enable pre-calculation of useful statistics, like sums and averages, and encourage the protection of such data sets by means of multiple methodologies, including differential privacy.

- Recognize and control for the fact that all data sets contain incorrect **outliers and errors**, specifically by:

    - Applying **automated tools** to eliminate outliers, identify constraint violations (*e.g.*, bad formatting or other inconsistencies) and, as possible, fill in missing values or fix errors. In some settings, problems and fixes should be reported to their original sources;

    - Acknowledging that **bias** in training sets may reflect past decisions and non-random samples that can and do render data sets ineffective or discriminatory, *e.g.*, by favoring particular demographics, or otherwise reflecting discrimination intrinsic to society or subsets of it.

ACM U.S. Technology Policy Committee
1701 Pennsylvania Ave NW, Suite 200
Washington, DC 20006

+1 202.580.6555
acmpo@acm.org
www.acm.org/public-policy/ustpc