

Population Association of America Association of Population Centers

Office of Government and Public Affairs

1436 Duke Street • Alexandria, VA 22314

www.populationassociation.org • www.popcenters.org • 301-565-6710 x 1006



Population Association of America President

Dr. Eileen Crimmins
University of Southern California

Vice President

Dr. Sara Curran
University of Washington

President-Elect

Dr. Robert Hummer
U. of North Carolina-Chapel Hill

Vice President-Elect

Dr. Marcia Carlson
University of Wisconsin-Madison

Secretary-Treasurer

Dr. Bridget Gorman
Rice University

Past President

Dr. John Casterline
Ohio State University

Board of Directors

Dr. David Bloom
Harvard University

Dr. Deborah Carr
Boston University

Dr. Jennifer Dowd
King's College, London, UK

Dr. Pamela Herd
Georgetown University

Dr. Emily Hannum
University of Pennsylvania

Dr. Jennifer Johnson-Hanks
University of California, Berkeley

Dr. Hedwig Lee
Washington University in St. Louis

Dr. M. Giovanna Merli
Duke University

Dr. Mary Beth Ofstedal
University of Michigan

Dr. James Raymo
University of Wisconsin, Madison

Dr. Jenny Trinitapoli
University of Chicago

Dr. Kathryn M. Yount
Emory University

Association of Population Centers President

Dr. Kathleen Cagney
University of Chicago

Vice President

Dr. M. Giovanna Merli
Duke University

Treasurer

Dr. Marcia Carlson
University of Wisconsin-Madison

Secretary

Dr. Jeffrey Morenoff
University of Michigan

March 9, 2020

Sent via email to: OpenScience@ostp.eop.gov

To whom it may concern:

The comments below are submitted on behalf of the over 3,000 members of the Population Association of America (PAA) (www.populationassociation.org) and the over 40 federally supported population research centers at U.S. based research institutions comprising the Association of Population Centers (APC) in response the notice in the Federal Registrar, [“Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research.”](#)

We are gratified to see that many of the stated desired characteristics of data repositories for storing data resulting from federally funded research align with the priorities of our organization and its members. We agree particularly with the following characteristics:

(A) *Persistent Unique Identifiers*. Unique identifiers will support data discovery, reporting, and research assessment. Repositories that support such identifiers would provide a standardized way of citing data products, aligning the incentives of academic rewards for principal investigators with the scientific community’s data sharing needs.

(B) *Long-term sustainability*. Experience in the population sciences has highlighted that technologies change quickly, and the value of historical data is sometimes limited because researchers cannot quickly and easily use such data with modern computing technologies.

(C) *Metadata*. We also agree that the distribution of metadata is crucial. Many fields have developed standards for the distribution of metadata, and such standards are crucial to enabling data discovery and reuse.

(E) *Access*. The call for data to be broad, equitable, and maximally open.

(F) *Free and Easy*. Making data free and easy to access are in keeping with traditions in the population sciences, as exemplified by our leading archive the Inter-university Consortium for Social and Political Research (ICPSR).

(H-K) *Secure. Private. Common Formats. Provenance*.

We also agree that storing data in repositories that are *secure* (H), *private* (I), have *common formats* (J) and clarifies the *provenance* (K) are critical objectives that can maximize the utility of data to researchers while also ensuring research participants that their data will be handled in ways that can produce societal good without personal harm.

We would also like to highlight certain tensions that exist within the proposed framework. Achieving the objectives of building repositories that are "Free & Easy to Access and Reuse" (F) and provide "Long-term sustainability" (B) have real costs, especially for large complex data sets that are often not adequately supported on Federal grants that fund data collection. Another tension is that designing repositories that make data more accessible and easier to reuse often means relying on existing technologies (e.g., for formatting and storing data), while ensuring the long-term sustainability of data in repositories often favors less efficient but more durable technologies. Achieving both of these objectives often involves considerable resources (e.g., because data must be stored in multiple formats). We believe the Federal government should make a modest investment to institutions to assure the long-term preservation and viability of research data.

Although we see the value of developing repositories to store data resulting from federally funded research, we are not in favor of placing these repositories within Federal agencies themselves. In the population sciences, so much of our data depend on the voluntary participation of citizens. Trust is an essential feature in gaining this voluntary participation, and the advent of massive computation and databases has left the public fearful of data sharing. We worry that the creation of massive Federal data repositories could be viewed suspiciously by the public, especially in the face of historical examples of Federal agencies sharing data to the detriment of the participating subjects. The Census Bureau is a model, having established data sharing firewalls across Federal agencies to address the public's concerns about how their data could be accessed and used and to ensure the agency's ability to collect complete and accurate data. The value of central data storage and distribution is enticing, but we recommend a cautious approach that balances a need to maintain the public confidence with the data sharing needs of the research community. We note that strengthening and maintaining well-known and stable repositories that have been created by non-governmental organizations, such as ICPSR might be a viable alternative to Federal data repositories as these repositories may not raise public fear of inappropriate data sharing within the Federal government.

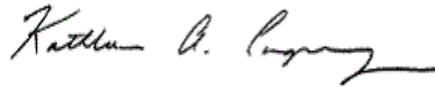
Finally, we would like to advocate for another characteristic of data repositories that was not mentioned in the Federal Registry notice, which is that there should be more mechanisms to protect sensitive, confidential data stored in repositories from being obtained through Freedom of Information Act requests and/or court orders. The Federal government currently provides two such mechanisms that protect confidential and sensitive human subjects data: the National Institutes of Health can issue Certificates of Confidentiality for funded and unfunded health-related survey efforts, while the Department of Justice can provide Privacy Certificates for DOJ-funded research projects. There are many examples of Federally-funded research projects that fall outside these narrow definitions and thus cannot be protected against the risk of disclosure, which compromises the ability of researchers to guarantee confidentiality to human subjects and also makes it difficult for researchers to obtain data from some government agencies. We advocate for the creation of a unified that would provide legal protections for all federally funded human subjects research.

Thank you for providing the opportunity for PAA and APC to comment on this important topic. We are eager to be a resource to the Office of Science and Technology Policy as it proceeds with any plans to encourage enhanced Federal data sharing and management policies.

Sincerely,



Eileen Crimmins, Ph.D.
President
Population Association of America



Kathleen A. Cagney, Ph.D.
President
Association of Population Centers