FASEB
Federation of American Societies
for Experimental Biology

Representing Over 130,000 Researchers

301.634.7000   9650 Rockville Pike
www.faseb.org   Bethesda, MD 20814

FASEB response to NIH Request for Information (RFI): Seeking Stakeholder Input on the Need for an NIH Administrative Data Enclave (NOT-OD-19-085)

Topic 1: Examples of NIH mission relevant biomedical and behavioral research using a data enclave that cannot be pursued currently.

The Federation of American Societies for Experimental Biology (FASEB) is particularly interested in the possibility of increased access to administrative/grant-associated data through the proposed NIH data enclave. Below we highlight four examples of analyses that would benefit from expanded data access and could be used to inform NIH policies pertaining to biomedical workforce development.

1) Assessment of outcomes of policy changes intended to support new/early stage investigators. Such an analysis would require application data, including the grant program(s) applied to, age of applicants, and other demographic data. Findings could be used to inform future NIH policies and initiatives to support specific career stages. Specifically, FASEB would be interested in utilizing this enriched administrative data set to do a more complete assessment of investigator outcomes for renewal rates of first R01/R01-equivalent awards and second R01/R01-equivalent applications to determine differences by gender, race and ethnicity, and research project topic.

2) Evaluation of interventions to broaden participation in STEMM fields. NIH and other research agencies support many programs to broaden STEMM participation. Current data sources make it difficult to compare outcomes of multiple independent programs. An NIH data enclave would facilitate cross-referencing of NIH IMPAC II data with program participation records to conduct multi-program assessments. Such analyses could help identify approaches or programmatic components associated with desired outcomes. IMPAC II data would also be useful for establishing control groups for such assessments.

3) Identification of research topic trends across demographic groups (age, type of institution, degree, etc.). This type of assessment would require access to application data, including proposal keywords and applicant demographic information. Analyses could identify research areas in which investigators from underrepresented groups are particularly active; fields of increasing or decreasing PI interest; and gaps in research activity. Furthermore, network-based analyses could examine emergence and features of cross-disciplinary research activities. Such evaluations could provide NIH leadership with insights regarding strategic deployment of funds across research and training programs.

4) Assessment of trainee outcomes by funding mechanism. Research project reports provide information regarding the number of laboratory staff and trainees supported by a particular grant. Comparing the outcomes of individual trainees supported by different grant mechanisms, e.g., T-, F- or R-series grant awards, would allow NIH to assess the utility of different funding strategies for developing the biomedical workforce and update agency policies.

The American Physiological Society • American Society for Biochemistry and Molecular Biology • American Society for Pharmacology and Experimental Therapeutics
American Society for Investigative Pathology • American Society for Nutrition • The American Association of Immunologists • American Association of Anatomists
The Protein Society • Society for Developmental Biology • American Peptide Society • Association of Biomolecular Resource Facilities
The American Society for Bone and Mineral Research • American Society for Clinical Investigation • Society for the Study of Reproduction • The Teratology Society
The Endocrine Society • The American Society of Human Genetics • American College of Sports Medicine
Biomedical Engineering Society • Genetics Society of America • The Histochemical Society • Society for Pediatric Research • Society for Glycobiology
Association for Molecular Pathology • Society for Redox Biology and Medicine • Society For Experimental Biology and Medicine
American Aging Association • U. S. Human Proteome Organization • Society of Toxicology

Topic 2: Whether the benefits of the proposed data enclave are worth repurposing NIH research funds to establish, maintain, and operate the data enclave.

Data-driven approaches to science policy issues, particularly those related to research funding portfolios and development and retention of the scientific workforce, are critical for maximizing federal investments in research. By providing controlled access to expanded administrative and scientific data the proposed data enclave could engage researchers and analysts beyond NIH in the application of unique perspective and analytical tools to assess NIH programs and investments. Other federal agencies offer comparable data access. For example, the National Science Foundation's National Center for Science and Engineering Statistics offers restricted expanded use of core datasets, such as the Survey of Earned Doctorates. Such data access increases the utility of collected information and resulting evaluations can provide powerful insights into key policy areas, such as workforce development.

While FASEB agrees that the proposed data enclave could lead to more robust analyses of the research and workforce supported by NIH funding, we do not support repurposing research funds for this initiative for two key reasons. First, the RFI does not provide projections for start-up and maintenance costs. Therefore we recommend that NIH conduct a preliminary assessment of start-up and maintenance costs and share the outcomes of this assessment with the stakeholder community prior to making the final decision regarding whether or not to proceed with establishing a data enclave.

Second, although the data enclave would expand the audience for de-identified NIH administrative and research information, the agency itself would likely remain the main user of this information. Therefore, in the event that NIH opts to proceed with establishing a data enclave, we strongly recommend that associated expenses be built into the agency's administrative budget and not tap into funds designated for research.

Topic 3: Preferences and considerations about accessing a data enclave only at a designated physical location or within a virtual environment.

FASEB's preference is that access to the data enclave be offered within a virtual environment rather than a designated physical location. While the latter would allow NIH to have tighter controls on access and data protections, we are concerned that it could also limit opportunities for researchers to utilize this data. We encourage NIH to seek guidance from other federal agencies, such as the National Science Foundation or Centers for Medicaid and Medicare Services, regarding their policies and procedures for managing third-party data access.

Topic 5: Examples of procedures an organization would implement to ensure the highest level of data protections, as well as to monitor, document, and notify NIH of any unauthorized and/or inadvertent data breaches.

As previously noted, we encourage NIH to seek guidance from other federal agencies, such as the National Science Foundation or Centers for Medicaid and Medicare Services, as well as leaders from the private sector regarding their policies and procedures for managing third-party data access.

Topic 6: Examples of outputs from approved research and how these may be shared with NIH.

While FASEB has not entered into a Special Data Access Agreement (SDAA) or comparable agreement for expanded use of NIH data, we have used publicly available data and related information provided by the agency to inform our reports, publications, and policy statements. Access to more detailed or previously unavailable datasets could have strengthened and added more context to several of our analyses, including the 2015 report, "Sustaining Discovery in Biological and Medical Sciences," and the 2017 report and recommendations on "Maximizing Shared Research Resources."