American Society for
Biochemistry and Molecular Biology
11200 Rockville Pike, Suite 302
Rockville, Maryland
20852-3110

**Responses submitted: May 30, 2019**
**See NOT-OD-19-085 for additional background.**

**Background:** The NIH is committed to transparency about its research investments and currently makes grant award information available to stakeholders (e.g. grantee institutions, researchers, professional organizations, the public) through web-based self-service tools. Currently RePORTER provides the public a searchable public repository of NIH-funded projects, and ExPORTER provides bulk files on funded projects for download. These tools contain non-sensitive information on NIH funded projects, including the institutions and principal investigators funded by NIH, with project abstracts and basic administrative data on those grant awards.

In recent years NIH has noted an increasing demand for access to sensitive information collected via the grants process. Such data includes information on peer review outcomes, progress reports, as well as, demographic information such as age range, sex/gender, race and ethnicity of individuals listed in NIH grant applications, etc. A recent report released by the Advisory Committee to the NIH Director on Next Generation Researchers calls for an increase in NIH administrative data for members of the biomedical research community. To address this demand, the NIH is considering making sensitive data available in accordance with the federal system of record[1]; collection, maintenance and dissemination of the data governed by the Privacy Act 1974, as amended, in a secure data enclave accessible only upon request through an approved Special Data Access Agreement (SDAA).

## I. Examples of NIH mission relevant biomedical and behavioral research using a data enclave that cannot be pursued currently.

Appropriate stewardship of resources or products is best conducted in a transparent environment in which entities disbursing resources or products are subject to review and inspection by a third party. This approach has led to remarkable gains in safety of a wide range of products including consumer goods, food, automobiles, and airplanes. Similarly, the stewardship of federal tax dollars disbursed as research grants is a product of the National Institutes of Health (NIH) and a valuable resource for grantee institutions. Independent analyses by third parties, often academic researchers or scientific organizations, of funding patterns and levels as a function of a range of variables is not only valuable, it is also a prudent avenue for providing independent review and inspection of the results of the NIH grant process. Without access to a data enclave, researchers and scientific organizations cannot easily nor efficiently access information that correlate awarded grants with age, gender, or career stage of principal investigators. An example of this research was published in 2018 (Hechtman et al., 2018, PNAS, dx.doi.org/10.1073/pnas.1800615115) in which researchers from the National Institute for General Medical Sciences studied and analyzed the longevity of NIH funding as a function of gender of the principal investigator. It should be noted that as NIH employees these researchers had privileged access

American Society for
Biochemistry and Molecular Biology
11200 Rockville Pike, Suite 302
Rockville, Maryland
20852-3110

to sensitive information not available to outside researchers and groups. A second study (Charette et al., 2016, PLOS ONE, dx.doi.org/10.1371/journal.pone.0168511) featured authors who were exclusively employees of the Department of Health and Human Services (NIH and CDC employees). The authors examined demographic changes, particularly changes in the age of grantees over time. Again, the privileged access to sensitive data used in this study is not normally available for external researchers. These two examples are not meant to be exhaustive, but rather as exemplars of the types of high quality and high value research that can be conducted using data that would be available through a data enclave. It is disheartening to imagine the vast array of research and analyses that could be carried out by qualified external researchers at academic institutions and scientific societies and the insights and lessons we are missing because this research cannot be done in the current regulatory climate.

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

The research that will be made possible by the proposed data enclave will provide the opportunity for independent analyses of the NIH funding process, making the effort worth funding. Analyses can be paired with the history publicly stated NIH funding policies and programs to provide a view of the consequences, intended or unintended, of how changes in NIH policies and funding levels have had effects upon the landscape of NIH funded research and the biomedical research enterprise. Armed with these new analyses we anticipate that NIH will then be in an improved position to make further course corrections and adjustments to NIH funding policies and programs to ensure the long-term viability of a robust biomedical research enterprise. These changes could include efforts that increase the efficiency of funding policies and programs. We believe that benefits from these efforts will offset the cost of repurposing NIH research funds to establishing, maintaining, and operating the proposed data enclave.

**III. Preferences and considerations about accessing a data enclave only at a designated physical location or within a virtual environment.**

Although a designated physical location for the proposed data enclave may typically be thought to be more secure, advances in digital security and the unrelenting shift to digitization of workflows suggest that if at all possible, the proposed data enclave should be housed in a virtual environment. Among these advances, two-factor authentication methods paired with appropriate data storage approaches and policies at NIH can maintain confidence in security of sensitive data. Because the threat of data espionage and theft are real threats to building a centralized data enclave, data security should be a top priority.

A virtual data enclave would also eliminate the need for researchers to travel to a centralized location. If carrying out data analyses and studies using this data and supported by NIH grant award funds, a central physical location of the proposed data enclave would further increase costs of operating and maintaining the data enclave as the funds used for travel to the physical location would take away from funds that

American Society for
Biochemistry and Molecular Biology
11200 Rockville Pike, Suite 302
Rockville, Maryland
20852-3110

could be dedicated toward directly supporting research. A virtual location for the data enclave would also eliminate access issues for researchers who could not reasonably make accommodations to travel to a central physical location.

**IV. Quantity of seats desired if NIH decides to make a substantial investment to sponsor access to sensitive data as allowable under the applicable federal laws in a secure virtual or physical environment.**

Initially NIH may want to set the number of available seats to 10, particularly when piloting the data enclave with early users. Once the data enclave is fully running, NIH should evaluate the necessary number of seats and adjust seat availability based on demand from the community. Quarterly or biannual analyses of demand may be sufficient.

**V. List examples of outputs from approved research and how these may be shared with NIH.**

We expect outputs of approved research to take the format of peer-reviewed manuscripts. Two prime examples are studies published in PNAS by Blau & Weinberg and Levitt & Levitt in 2017 (Blau & Weinberg, 2017, PNAS, dx.doi.org/10.1073/pnas.1611748114; Levitt & Levitt 2017, PNAS, dx.doi.org/10.1073/pnas.1609996114). These studies accessed sensitive data under special data access agreements and the published papers reported aggregated data that was not identifiable. These critical analyses are examples of the high quality and value of research that can be conducted by experts outside of NIH when provided with access to sensitive data. Manuscripts from approved research should be made available to NIH and deposited in PubMed Central. The NIH should consider whether it would be feasible or useful for the processed and analyzed data products that include sensitive information from approved research to be shared with the organizers of the data enclave at NIH, or whether this processed data and analyses should be made available to approved subsets of NIH employees.

**VI. 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.**

Procedures to be implemented by institutions or organizations to monitor, document, and notify NIH of unauthorized or inadvertent data breaches may copy restrictions and procedures commonly used by institutions or organizations in restricting access to and protecting identifiable patient information in clinical studies. These may include procedures that dictate local data storage and access locations and protocols, and a dedicated information technology professional responsible for setting data access rights, methods, and logs of users and timestamps of data access.

American Society for
Biochemistry and Molecular Biology
11200 Rockville Pike, Suite 302
Rockville, Maryland
20852-3110

**VII. Examples of outputs from approved research and how these may be shared with NIH**

As described in our response to question V, we expect outputs from approved research to primarily take the form of published manuscripts.