COGR

Council On Governmental Relations

An Association of Research Institutions

December 8, 2021

Submitted via Federal eRulemaking Portal: http://www.regulations.gov

Dorothy Aronson, Chief Data Officer
Office of Information and Resource Management
National Science Foundation
2415 Eisenhower Ave
Alexandria, VA 22331

RE: Comments in Response to National Science Foundation Notice of a New System of Records [86 F.R. 62217 (Nov. 9, 2021)]; Docket No. NF_FRDOC_001_2796

Dear Ms. Aronson:

The Council on Governmental Relations (COGR) is an association of 200 public and private U.S. research universities and affiliated academic medical centers and research institutes. COGR concerns itself with the impact of federal regulations, policies, and practices on the performance of research conducted at its member institutions. COGR works to address important government interests associated with funding of research, while minimizing unnecessary administrative burden to better enable researchers to engage in the open academic inquiry that is the hallmark of fundamental research. COGR and its member institutions recognize the importance of ensuring research integrity and responding to inappropriate foreign influence on federally funded research, and we have worked to develop and promote effective practices in this area,[1] while fostering international collaborations that are vital to the success of the U.S. and global scientific enterprise.

COGR welcomes the opportunity afforded by the National Science Foundation (NSF) to submit comments in response to the above-captioned "Notice of a new system of records" under the Privacy Act of 1974 (hereafter referred to as the "system of record notice" or "SORN).[2] The NSF SORN describes a new system of records named NSF-77 Data Analytics Application Suite ("NSF-77").[3] Per § 552a(e)(4) of the Privacy Act, publication and

[1] *See, e.g.,* COGR, Foreign Influence – Practical Considerations in Developing an Institutional Response (Aug. 2021); Principles for Evaluating Conflict of Commitment Concerns in Academic Research (Sept. 2021).
[2] We also appreciate NSF meeting with representatives from COGR and other associations and institutions on Nov. 29, 2021, at a presentation hosted by the American Association of Universities (AAU) to provide information on how NSF-77 will be used (hereafter the "NSF Presentation").
[3] 86 F.R. 62217 (Nov. 9, 2021).

solicitation of comments on the NSF SORN is required before NSF-77 can be deployed, to inform decisions concerning individuals or to share individual data with other government agencies. COGR understands the need for federal agencies to utilize tools such as NSF-77 to analyze the vast amount of data they receive from individuals and institutions, and we recognize that when employed appropriately, such tools can reduce administrative burden on agencies, institutions, and individuals.

In reviewing the NSF SORN, COGR and its member institutions noted the description of the NSF-77 tool's intended use is quite generalized and vague. The purpose of a SORN is to promote transparency regarding the new system of records by providing the public with adequate notice about the systems' contents and use and processes for its operation.[4] In sum:

> SORNs must be clear, unambiguous, and understandable to the general public while fulfilling the necessary legal requirements of the Privacy Act. The publication requirements of the SORN are intended to:
>
> 1. Prevent the creation of a system of records without first giving individuals an opportunity to review and comment ***on the purpose and routine uses*** for which their information is collected.
>
> 2. Help individuals locate systems of records that are likely to contain information pertaining to them.[5]

Overall, the NSF SORN does not provide researchers with adequate notice of how their information will be used and disclosed, and therefore is not adequate under the Privacy Act. In addition to this overarching concern about transparency and adequate notice, COGR also notes the following more specific issues:

- Privacy Act Provisions: The NSF SORN does not provide sufficient information to permit assessment of: (a) how NSF-77 will fulfill certain of the NSF SORN's enumerated purposes[6]; or (b) how NSF will address several important requirements of the Privacy Act in its use of NSF-77.
- Rules of the Road: The NSF SORN does not describe the guidelines and processes, or "rules of the road," that the agency will apply in utilizing NSF-77.
- Individual Information Access/Amendment Mechanism: The mechanism outlined in the NSF SORN by which researchers may access/amend their individual information collected by NSF-77 is too slow and cumbersome to permit rapid validation or correction of information. Further, the tool might

[4] 5 U.S.C. § 552(a)(e)(4).
[5] U.S. Office of Personnel Management System of Records Notice (SORN) Guide (Apr. 2010) at p. 6 (emphasis supplied).
[6] 86 F.R. at 62217-18.

better accomplish its objectives if used in a proactive, as opposed to a retroactive, fashion.

Each of these points is discussed in greater detail in the remainder of this letter, and we respectfully request that NSF issue a revised NSF SORN that addresses our suggestions.

Privacy Act Provisions:

Fulfillment of the NSF SORN's Stated Purposes: Section 552a(e)(1) of the Privacy Act requires agencies that maintain a system of records to "maintain in [those] records only such information about an individual as is relevant and necessary to accomplish a purpose of the agency required to be accomplished" by law. The NSF SORN lacks any discussion about how the categories of records in the system will accomplish NSF-77's following listed Purposes of the System[7] (the "Unaddressed Purposes"):

- Purpose 2 - Evaluating impact and return on investment of awards
- Purpose 3 (in part) - Providing necessary analyses for the specified strategic priorities of equity and partnerships
- Purpose 5 - Tracking career development, mentorship and outcomes of education grants and their training activities
- Purpose 6 - Merging internal data to facilitate agency organizational efficiency and portfolio analysis
- Purpose 8 - Informing pre-onboarding and onboarding evaluations of NSF staff

The NSF SORN's categories of covered individuals, system records, and record sources are primarily focused on serving the ends of research security, monitoring international collaborations, and detecting commitment overlap and scientific overlap, i.e., Purposes 1, 3 (in part), 4, and 7 (the "Addressed Purposes"). The Unaddressed Purposes are equally, if not more, important than the Addressed Purposes from a privacy and utility perspective, but the NSF SORN's vagueness makes it extremely difficult to determine if NSF-77 is maintaining only those records relevant and necessary to achieve these aims, as required by the Privacy Act. COGR suggests that the NSF SORN be modified to address why the covered records are necessary to achieve the Unaddressed Purposes and how they accomplish this task.

Additional Privacy Act Requirements: The Privacy Act of 1974 enumerates eleven requirements that federal agencies must follow in maintaining systems of records.[8] In comparing the information in the NSF SORN against these items, we noted the NSF SORN does not provide adequate detail to determine whether or how NSF is addressing the following Privacy Act requirements:

---

[7] 86 F.R. at 62217-18

[8] 5 U.S.C. §§ 552a(e)(1)-(11); *see, generally*, U.S. Dept. of Justice, Overview of the Privacy Act (2020 ed.) ("DOJ Overview").

- **5 USC §552a(e)(2) "Each agency that maintains a system of records shall . . . collect information to the greatest extent practicable directly from the subject individual when the information may result in adverse determinations about an individual's rights, benefits, and privilege under Federal programs."**

In its description of "Record Source Categories," the NSF SORN states that "publication and patent information published by PIs will be obtained from third parties that compile related public information," including but not limited to "Clarivate (Web of Science), Elsevier (Scopus), Dimensions, USPTO, PubMed, arxiv databases, ORCID and Google Scholar."[9] However, NSF has not provided any justification for why it is not practicable to collect such information directly from researchers, as called for by the Privacy Act. The NSF SORN should address this basic Privacy Act requirement.

This omission is particularly noteworthy because NSF has not explained how the third-party data sources assure that information they obtain is in fact published "by PIs." Even assuming accurate attribution to PIs, the system of records described in the NSF SORN covers graduate, postdoctoral, and undergraduate students[10] -- individuals who do not generally fall within the category of "senior personnel" required to make proposal/report disclosures to NSF.[11] The NSF SORN fails to explain how NSF will attempt to collect information directly from these individuals "to the greatest extent practicable," as required by the Privacy Act.[12] Further, the NSF 2021 Proposal and Award Policies and Procedures Guide ("2021 PAPPG") places a three-page limit on Biosketches (the primary vehicle by which publications/patents would be disclosed) and states that listed publications should be limited to ten items.[13] In short, covered individuals may be subject to NSF inquiries, and potentially sanctions, for inconsistencies that arise, in part, from third-party data over which they have no control or from self-reported data for which NSF applies disclosure limits. We suggest that NSF specifically address in the NSF SORN the issue of collecting publication and patent information directly from all categories of covered individuals and consider removing any limits placed on listing publication/patent information in Biosketches or other NSF forms, if that data is to be maintained in this system for the stated purposes.

- **5 U.S.C. § 552a(e)(5) "Each agency that maintains a system of records shall . . . maintain all records that are used by the agency in making any determination about any individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination."**

---

[9] 86 F.R. at 62218. Additionally, we note that the information sources cited in the NSF SORN are far more extensive than those mentioned by NSF in the NSF Presentation.
[10] 86 F.R. at 62218 ("Categories of Individuals Covered by the System").
[11] *See, generally*, NSF 2021 Proposal & Award Policies and Procedures Guide (PAPPG), §§ II.C.2.f & h concerning requirements for senior personnel to disclose biographical sketch and current and pending support information.
[12] The NSF SORN also does not describe how NSF will collect information on these individuals from third parties or from NSF proposals and annual/interim reports.
[13] *See,* 2021 PAPPG § II.C.2.f.i.(c).

As noted, the NSF SORN states that data from third parties on publications and patents will be compared with individual disclosures to identify inconsistencies. The NSF SORN does not, however, provide any details about how NSF, or these third parties, will ensure the quality of the records they provide. We suggest that NSF include within the NSF SORN details about how the accuracy, relevance, and completeness of these third-party documents will be assured. Without this information, it is difficult, if not impossible, to assess the validity of any noted inconsistencies. Additionally, transparency regarding data quality assurance measures will foster stakeholder trust in the system.

Rules of the Road

The NSF SORN employs exceptionally broad language in several areas. For example, it lists as a record source "third parties that compile related public information" without any limitation, and it encompasses a wide swath of routine uses, including sharing of information regarding identified inconsistencies with multiple law enforcement/security/intelligence agencies, as well as "federal agencies contributing to cross-governmental forums on research security."[14] In the context of a SORN, such wide-ranging language may be useful to ensure general notice of a system of records' uses. From the perspective of a researcher struggling to make sense of continually unfolding disclosure guidance, however, this language can appear as an unnecessarily broad dragnet, particularly when considered in the larger context of the outcomes of recent China Initiative cases in academia.[15]

The NSF SORN makes an oblique reference to verifying any noted inconsistencies[16] "according to internal guidelines and review processes,"[17] but it provides no details on these guidelines/processes, or "rules of the road." Without this detail, the NSF SORN fails as a clear notice to researchers of the routine uses that will be made of their records. Further, the lack of transparency on this point coupled with the NSF SORN's overall vagueness may undercut researchers' trust in the funding agency.[18]

COGR urges NSF to amend the NSF SORN to include the express details of these "rules of the road" for comment prior to their implementation. Promoting full transparency and engaging in formal dialog regarding such rules will help researchers and institutions understand NSF's boundaries governing the use of NSF-77 and foster community trust that the system will be deployed in a defined and fair manner. At a minimum, we hope that NSF will publish these "rules of the road" as draft guidance on which the research community can comment.

---

[14] 86 F.R. at 62218-19.
[15] *See, e.g.,* Strohm, C., "DOJ Curbs Trump-Era Zeal for China Spy Probes as Cases Fail," Bloomberg News, (Oct. 27, 2021).
[16] At the NSF Presentation, agency representatives helped to mitigate concerns about how NSF-77 will be operated by providing detailed information about guardrails that NSF would employ in its use, particularly regarding the verification of any inconsistencies that are identified. As these "rules of the road" are not included in the NSF SORN, however, our comments here focus the document's current text.
[17] 86 F.R. at 62219.
[18] See, e.g., Mervis, J., "Two Surveys Document Harmful Impact of China Initiative on Researchers," Science (Oct. 28, 2021).

Individual Information Access/Amendment Mechanism

The NSF SORN references the NSF's Privacy Act regulations at 45 CFR Part 613 with respect to mechanisms for accessing and amending individual information in NSF-77. Unfortunately, the described processes do not provide for quick turn-around. For example, the regulations state that the agency will "make reasonable effort to act on a request for access to records within 20 days of its receipt,"[19] and it can take up to 30 days for the agency to process requests to amend records.[20] Additionally, the individual is responsible for paying access costs.[21]

This standard approach to record access and amendment will be extremely burdensome on individual researchers given the large number and categories of records in the system (e.g., DOI-citable S&T materials, co-author connections, citations of other papers in present and future publications; publication content, other media sources, etc.[22]), and the fact that many of the records will come from third parties over whom the requester has no control. In light of the extremely broad uses to which the federal government may put the data, the risk to individuals of an adverse action resulting from incorrect information is high.

COGR requests that NSF consider alternate additional mechanisms for providing researchers with access to their information in the system. Given that NSF will need to refresh the data periodically, researchers should be able to easily access their personal information on a regular basis without cost, and they also should be afforded a streamlined process for addressing errors they detect. In this regard, COGR respectfully suggests that NSF consider providing researchers with no-cost direct access to NSF-77 tool to obtain reports on themselves at any time, including prior to proposal submission.

The approach set forth in the NSF SORN contemplates a retrospective look-back in which awarded proposals are compared against third-party information. Disclosure rules have been in a constant state of flux over the past two years, and depending on the look-back period used, NSF-77 reports may identify past innocent inconsistencies that would not occur under institutions' new policies, processes, and training in this area. Permitting researchers to use the NSF-77 tool prior to proposal submission would enable them to proactively identify and address inconsistencies before submission to NSF. Such a proactive approach recognizes that the vast majority of researchers are hard-working scientists doing their best to follow the rules and limits any "gotcha" aspect that may arise from using the tool solely in a retrospective fashion. Moreover, a proactive approach to the use of NSF-77 aligns with one of the laws the NSF SORN lists as providing the legal basis for NSF-77: the Foundations for Evidence-Based Policymaking Act of 2018.[23] Title II of this Act – the OPEN Government Data Act[24] -- amended 44 U.S.C. § 3506 to mandate that federal agencies develop open data

[19] 45 CFR § 613.3(a).
[20] 45 CFR § 613.4.
[21] 45 CFR§ 613.3(e).
[22] 86 F.R. at 62218 (Categories of Records in the System).
[23] Pub. L. 115-435 (Jan. 14, 2019).
[24] *Id.* at § 202.

plans (for data that does not concern monetary policy) that require, among other things, data collection mechanisms to be available in an open format.

Overall, if the purpose of the NSF-77 tool is to ensure that the disclosures sent to NSF are accurate and complete, then permitting researchers to access the tool for their submissions on the front end is the most direct approach to achieving this objective. Further, this approach helps to prevent chilling effects on international scientific collaborations that are essential to the U.S. scientific enterprise.[25]

COGR hopes that NSF will consider this "front-end" approach and take the steps necessary to make the NSF-77 tool (or at a minimum, reports generated from that tool) available to researchers for use in reviewing proposals before submittal, rather than solely employing a retrospective approach. Alternatively, if NSF continues with the retrospective approach, COGR respectfully suggests that the "look-back" period go back no farther than 2021 because of the evolution in disclosure requirements over the past two years.

Conclusion

Although not directly applicable to the system of records described in the NSF SORN, the Privacy Act sets forth an additional provision that bears consideration in this area. Specifically, § 552a(e)(7) of the Act states that agencies shall "maintain no record describing how any individual exercises rights guaranteed by the First Amendment," including the right to free association. As described in the NSF SORN, a core purpose of NSF-77 is to identify and analyze researchers' associations, whether they be research collaborators, co-authors, co-patent holders, or mentors/mentees. Given this fundamental right to choose with whom to associate, coupled with the fact that scientific collaborations are the lifeblood of discovery, we sincerely hope that NSF will give serious consideration to the suggestions herein. By publishing a revised NSF SORN that incorporates COGR's recommendations, NSF can assure compliance with the Privacy Act, as well as improve both the operations and aims of the NSF-77 tool.

Once again, we thank NSF for the opportunity to provide these comments on behalf of our member institutions. If you have any questions regarding this letter, please contact Kris West, Director, Research Ethics and Compliance at kwest@cogr.edu.

Sincerely,

Wendy D. Streitz
President

cc: Rebecca Keiser, Chief of Research Security Strategy and Policy

[25] *See, e.g.,* Qin, A., "As U.S. Hunts for Chinese Spies, University Scientists Warn of Backlash," New York Times (Nov. 28, 2021).