7555-01-P

NATIONAL SCIENCE FOUNDATION

Request for Information; Datasets to Conduct Research on Computer and Network Systems

AGENCY: National Science Foundation.

ACTION: Request for Information.

SUMMARY: The Division of Computer and Network Systems of the National Science Foundation seeks public input from the research community on the specific needs for datasets to conduct research on computer and network systems.

DATES: Please send comments on or before 5:00 PM Eastern time on May 21, 2021. Submit comments via the SurveyMonkey link found in the "Instructions to Submitters" below.

ADDRESSES: Email comments to: Dr. Nicholas Goldsmith, AAAS Science & Technology Policy Fellow at nicgolds@nsf.gov . Send written submissions to: Division of Computer and Network Systems, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, VA 22314. Submit comments via https://www.surveymonkey.com/r/RFIDCLSurvey.

FOR FUTHER INFORMATION CONTACT: Contact Dr. Alex Sprintson, NeTS Program Director at asprints@nsf.gov or Dr. Nicholas Goldsmith, AAAS Science & Technology Policy Fellow at nicgolds@nsf.gov or call (703)-292-8950.

SUPPLEMENTARY INFORMATION: The ubiquity, structure, and use of communication networks and computing systems have changed dramatically

over the last decade. The technology trade-offs that have enabled these networks and systems are becoming increasingly more complex with convergence across computer systems (spanning mobile, edge, fog, and cloud computing, etc.), application accelerators, distributed systems, network stacks, wireless systems, and wired network domains, thereby decreasing the efficacy of traditional model-based approaches. As a result, researchers are increasingly relying on machine learning and other data-intensive techniques to lead the development of next-generation, high-performance networks and computer systems. This necessitates the availability of representative datasets that can inform such research. Furthermore, representative datasets will enable the Networking Technology and Systems (NeTS) and Computer Systems Research (CSR) communities to contribute to innovations in Advanced Wireless and Artificial Intelligence, both of which have been identified as strategic priority areas for the Nation.

Addressing current and future research areas may require access to specific types of datasets that capture a broad range of practical settings and navigate through a complex set of design trade-offs. Researchers utilizing machine learning and other artificial intelligence techniques may need large, labeled data to use as training and testing sets, to test algorithms and protocols that they have developed, or to assess the viability of their design methodologies. More generally, datasets can motivate research questions or identify areas to target in future work. Equitable access to data is also essential for replicable and reproducible research.

Additionally, identification of the specific dataset needs of the research community may motivate the collection of specific new types of data or the creation of new tools for accessing and analyzing data. Existing or future NSF

infrastructure investments, such as the Platforms for Advanced Wireless Research (PAWR), may be important venues for collecting the identified data.

This Request for Information (RFI) seeks input from the community on the specific needs related to collecting, sharing, and utilizing public or private datasets for networking and computer systems research, and any challenges associated with each. The input could identify requirements for datasets that may include, but are not limited to, spectrum data, physical layer data, network and Internet measurement data, workload data, power/performance data, and other systems data. NSF recognizes that some datasets currently exist but is interested in needs that are not currently met by these existing datasets, conventions or formats that may broaden the usability of the data, and ways in which additional high-quality datasets may be made available to the research community. NSF is interested in assessing where research progress is slowed due to the lack of datasets that may either already exist or can be generated using existing infrastructure (including NSF-funded infrastructure). NSF may use the responses to this RFI to inform and refine future investments.

Instructions to Submitters

NSF invites individuals and groups of individuals to provide their inputs via the online submission form (link below). The submission form requires the following information:

- Contact person name and affiliation.
- Valid contact email address.
- Additional author name(s) and affiliation(s)
- Research domain(s), discipline(s)/sub-discipline(s) of the author(s), including either NeTS, NeTS-Wireless, or CSR.

- Title of the response.

- Abstract (maximum 200 words) summarizing the response.

- Question 1 (maximum 1000 words) – Data Needed for Research. *State whether or not your research requires datasets. If your research requires datasets, describe whether or not you have access to the needed datasets with sufficient quality; and describe what type of data would address your current need for datasets if it is not being met. NSF is interested in where the lack of datasets and/or the quality of datasets may be holding back research, what datasets would help take research to the next level, and the proportion of researchers that have a need for datasets.*

- Question 2 (maximum 600 words) – Ability to Contribute. *Describe the type of datasets you may be able to contribute to the research community and any barriers to making these datasets available to the research community over at least a seven-year period.*

- Question 3 (maximum 600 words) – Privacy. *Describe the concerns, either as a user and/or a data provider, that you may have in maintaining and ensuring data privacy, in anonymizing data, and in the effects of data anonymization on data quality. Specific ideas to address data privacy and anonymization concerns are also welcome.*

- Question 4 (maximum 600 words) – Format and Metadata. *Describe any suggested formats or standards with which datasets should conform. Describe the types of metadata which should be included with data, as well as particular parameters of concern in the data collection or generation.*

- Question 5 (maximum 600 words) – Other Considerations. *Any other relevant aspects that need to be addressed; or any other issues that NSF*

should consider, such as where such datasets may exist (e.g. Federal agency, industry, service providers, international partners) and intellectual property concerns.

- Checkbox to consent to NSF's use and display of the submitted information, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode). *NSF anticipates making submissions publicly accessible through a website.*

To respond to this RFI, please use the official form available at https://www.surveymonkey.com/r/RFIDCLSurvey . We recommend writing out your responses in a separate document, and then pasting them into the response fields on the form.

NSF will use the information submitted in response to this RFI at its discretion and will not provide comments to any responder's submission. The information provided will be analyzed, may appear in reports, and may be shared publicly on agency websites. Respondents are advised that the government is under no obligation to acknowledge receipt of the information or provide feedback to respondents with respect to any information submitted. No proprietary, classified, confidential, or sensitive information should be included in your response. The government reserves the right to use any non-proprietary technical information in any resultant solicitation(s), policies, or procedures.

Authority: 42 U.S.C. 1861 et al.

Dated: April 8, 2021.

Suzanne H. Plimpton,

Reports Clearance Officer,

National Science Foundation.