NATIONAL SCIENCE FOUNDATION
2415 Eisenhower Avenue
Alexandria, Virginia 22314

NSF 18-060

Dear Colleague Letter: Advancing Long-term Reuse of Scientific Data

April 6, 2018

Dear Colleagues:

Through this Dear Colleague Letter (DCL), the National Science Foundation's (NSF) Office of Advanced Cyberinfrastructure (OAC) announces its intention to support initial exploratory activities toward the creation of social and technical infrastructure solutions that further NSF's commitment to public access. These solutions are a means to accelerate the dissemination and use of fundamental research results in the form of data that will advance the frontiers of knowledge and help sustain the Nation's prosperity well into the future.

NSF supports fundamental research grants that result in publications, primary data, samples, physical collections and other supporting materials created or gathered in the course of work performed under these grants [see NSF's Proposal and Award Policies and Procedures Guide (PAPPG) Chapter XI.D.4, https://www.nsf.gov/pubs/policydocs/pappg18_1/pappg_11.jsp#XID4 for details]. This particular DCL is focused on exploratory solutions that advance public access by reducing the barriers to data reuse within the scientific community, as guided by NSF's public access plan, Today's Data, Tomorrow's Discoveries (see https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf15052).

Specifically, this DCL encourages two types of funding requests: (1) proposals for Conferences (i.e., community workshops and other events) that are designed to bring together stakeholders to explore opportunities to converge on innovative solutions to advancing public access; and (2) proposals for Early-Concept Grants for Exploratory Research (EAGER) for high-risk/high-reward innovative concepts and pilot projects that yield new fundamental research discoveries from existing NSF-funded data or that ultimately result in deployment of ambitious, sustainable socio-technical infrastructure resources and capabilities that enhance and accelerate new discoveries from existing NSF-funded data. Research ideas that do not advance public access as narrowly defined in this DCL may be suitable for other solicitations such as Cyberinfrastructure for Sustained Scientific Innovation (CSSI) - Data and Software (see https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf18531).

SPECIFIC GUIDANCE TO PROPOSERS RESPONDING PURSUANT TO THIS DCL

This DCL encourages funding requests aligned with one of the following three tracks:

I. **Community track**: This track funds proposals for Conferences (i.e., community workshops and other events) that enable better data stewardship by the NSF research community, in particular of

data produced and used by the community in the conduct of research and education. Topics include community activities to organize stakeholders (e.g., discipline experts, data repository managers, and data appraisal experts) to explore:

- Community-specific agreements that identify the data of importance to the community; knowing what to keep helps determine what to throw away;
- Common data types (e.g., volumetric, image, etc.) across multiple disciplines to harness tools and best practices in data stewardship and use;
- Data repository findability, accessibility, interoperability, and reuse;
- The minimal descriptive information for findability and accessibility of data; and
- Best practices associated with data management plans.

II. **Data reuse track.** This track encourages reuse of data created as a product of NSF-funded research. Research ideas are sought in two areas as described below.

- EAGER proposals for high-risk/high-reward innovative studies that address development and testing of important science and engineering ideas and theories through use of existing data. Proposals that are responsive to this track may not involve collection of new data or field research; may not involve data created by an NSF Large Facility (see the list of NSF Large Facilities at https://www.nsf.gov/bfa/lfo/docs/large-facilities-list.pdf); and may not come from an investigator who is listed as a principal investigator (PI) or co-PI on an award that created the data set of use. Rather, proposals must:
    - Involve, for data proposed for use, publicly-available data generated through NSF funding; and
    - Agree to make public the details about their experiences reusing the data, including especially challenges associated with that reuse.
- Proposals for Conferences (community workshops) that creatively employ data challenges, meetups, hackathons, or related activities. These activities enable education and workforce development, along with novel use of existing data created through NSF funding. The majority of the data (but not all) must be publicly available and the result of NSF-funded activities.

III. **Socio-Technical Infrastructure.** This track encourages EAGER proposals for high-risk/high-reward innovative concepts and pilot projects that address one or more social and/or technical barriers that limit the findability, accessibility, and interoperability of research data in the US and internationally. Suggested topics include, but are not limited to, exploration of:

- Utility of persistent identifiers early in the data lifecycle that facilitate discovery, filtering, indexing, and routing of the data objects;
- Costs to repositories of legacy data objects made findable, accessible, interoperable, and reusable;
- Metrics for assessing findability and accessibility of data;
- Community-driven studies of data appraisal;
- Actions to reduce adverse use factors that fit the norms of a community; and
- Principles for generation of data that are consciously designed for reuse.

SUBMISSION PROCESS

The deadline for submission of Conference and EAGER proposals proposal submission date is May 23, 2018. Guidance on proposal preparation is given in Chapter II.E of the NSF PAPPG: for EAGER proposals see part 2 at https://www.nsf.gov/pubs/policydocs/pappg18_1/pappg_2.jsp#IIE2 and for Conference proposals see part 7 at https://www.nsf.gov/pubs/policydocs/pappg18_1/pappg_2.jsp#IIE7.

Proposals may be submitted via Fastlane or Grants.gov. NSF anticipates that all awards will be made by September 2018.

Conference requests in general should not exceed $50,000 for one- or two-year durations. EAGER proposals can be supported at up to $300,000 for up to two years.

PIs are urged to discuss the suitability of their ideas with Beth Plale at bplale@nsf.gov prior to submission.

Sincerely,
Jim Kurose
Assistant Director, Directorate for Computer and Information Science and Engineering