

## Concept Clearance FY 2023

**INITIATIVE TITLE:** Early Stage and Established Biomedical Data Repositories (DR) and Knowledgebases (KB)

**INITIATIVE TYPE:** Renew (with minor modification)

**ACTIVITY CODE:** U24

**OBJECTIVE(S):** Two companion Funding Opportunity Announcements (FOA) to invite cooperative agreement (U24) applications – one announcement for early-stage biomedical repositories or knowledgebases and another for established biomedical data repositories or knowledgebases.

**DESCRIPTION:** The purpose is to continue the Funding Opportunity Announcements, supporting both early-stage and established biomedical data repositories or knowledgebases.

Repositories are defined as core data related to a particular system or systems that are available as a community resource and accept submission of relevant data from the community, validate, store, archive/preserve and distribute the data in compliance with the FAIR data principles. Knowledgebases are defined as data resources that collate, extract, organize, annotate, and link the growing body of information related to and relying on core datasets from one or more repositories, curate literature, and generate other forms knowledge and distribute them in compliance with FAIR data principles.

The program supports data resources that demonstrate a sufficiently significant or potential research usage and a userbase to warrant NIH support. The program focuses on impact and methods used to engage research community(ies), demonstrated impact through citations, data, or information, and demonstrate coordination across a field or fields of study. For the purposes of these FOAs, early-stage is defined as initial development of a data repository or knowledgebase with potential for a large increase in usage and adoption in the community. Established repositories and knowledgebases on the other hand are defined as high-value data resources that are near or at optimal research community penetration and are a critical research resource as demonstrated by their usage, utility, and impact.

Based on programmatic input across NIH, including the NIH repository and knowledgebase working group, the reissue of this solicitation includes modifications to better address NIH Institute-specific mission priorities, life-cycle stages, and NIH goals for data resources. The broadly applicable results of this program align with the goals as set forth in the NIH [Strategic Plan for Data Science](#).

Key aspects of the solicitation include:

- **Scientific Significance:** The resource should have a national and international significance, with relevance and an explicit mission statement that describes the importance to the NIH domains of science.
- **Community Needs and Engagement:** The resource should demonstrate that there is sufficiently significant or potential research usage and a userbase to warrant NIH support. Funded resources must demonstrate usage, utility, and scientific impact and openly share metrics.

- **Quality of Data and Services and Efficiency of Operations:** The resource should demonstrate adherence to the desirable characteristics of repositories and FAIR data principles. It should also use community standards for data and metadata, unique and persistent data identifiers, provenance of data, adequate infrastructure, and data storage. All applications require resource development, project development, and sustainability plans.
- **Governance:** The resource should document processes for developing terms of use, including appropriate processes for data licenses and policies for privacy and ethics. Resources funded through this mechanism require an external advisory committee.

Resources funded as early-stage resources as defined above may not exceed a budget of \$300K in direct costs per year for a period of four years. Applicants for early-stage resources are strongly encouraged to consult the NIH Institute prior to development of the application. Application budgets for established repositories or knowledgebases are not limited but need to reflect the actual needs of the proposed project and require Institute approval. All established resource applications and budgets above \$500K in direct costs/year require institute approval. Renewal applications are permitted.

**IMPORTANCE:** Accessible, well-maintained, and efficiently operated data resources are critical enablers of modern biomedical research. Data resources, through good data management practices, are the key to data and knowledge discovery, integration, and reuse, as outlined by the FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable digital objects). In order to sustain a healthy and productive data resource ecosystem it is critical to ensure that each component: (a) delivers scientific impact to the communities that they serve; (b) employs and promotes good data management practices and efficient operation for quality and services; (c) engages with the user community and continuously address their needs; and (d) supports a process for data life-cycle analysis, long-term preservation, and trustworthy governance.

These FOAs are dedicated to supporting data resources (repositories or knowledgebases) that are expected to (a) adopt good data management practices as outlined by the FAIR Data Principles allowing resources to form a trans-NIH data ecosystem; (b) lower barriers for data sharing; (c) allow greater usage and utility of the data for research and discovery; (d) discourage development of resources under research grant mechanisms and to discourage use of research mechanism dollars for resources; and (e) incentivize efficiency of operations to allow longer term preservation of data and enable greater sustainability of resources. Hence the need for this important program to realize NIH's vision for a modernized biomedical data ecosystem per the goals outlined in NIH's strategic plan for data science.

**HISTORY:** The funding opportunity announcements ([PAR-20-097](#) and [PAR-20-089](#)) were first launched in 2019 in partnership with ten NIH Institutes, Centers, and Offices (NIGMS, NHGRI, NINDS, NIMHD, NIEHS, NIDA, NIAID, NIAAA, NCI, and NHLBI). To date 30 applications have been reviewed in response to the solicitations and NIH has funded 7 applications (23% success rate). The applications are reviewed in a special emphasis panel in CSR.

Based on the success of the program and support of the NIH Scientific Data Council, ODSS requests concept clearance from the Councils of Councils to reissue this initiative with the proposed modifications.

**CONCEPT CLEARANCE DATE: January 27, 2022**

**COUNCIL REMARKS:**