

Early-stage and Established Biomedical Data Repositories and Knowledgebases

FY23 Concept Clearance



*January 27, 2022
NIH Council of Councils*

Concept Clearance Overview

Concept Clearance: Reissue (PAR-20-089, PAR-20-097)

Title: Early-stage and Established Biomedical Data Repositories (DR) and Knowledgebases (KB) (**Mechanism U24**)

Objective/Purpose: To seek data repository and knowledgebase resource projects that deliver scientific impact, employ and promote good and efficient data management, engage the user community, and govern to address biomedical data life-cycle and long-term preservation.

Funds Available and Anticipated Number of Awards: Contingent upon NIH appropriations and submission of meritorious applications

Award Project Period: up to 5 years

Council Action: Vote for continued support of Biomedical Repositories and Knowledgebases

Goals of the Program

An NIH program to support investigator-initiated, sustainable data resource development driven by critical research needs



* Including NIH desirable characteristic of repositories [NOT-OD-21-016](#)

** Require a sustainability plan which includes transitions and sun-setting

Background and History

Solicitation U24: [PAR-20-089](#) U24: [PAR-20-097](#)

Project Requirements: Support for repositories and knowledgebases that deliver scientific impact, employ and promote good and efficient data management, engage the user community, and govern to address data preservation.

No prior NIH-wide data resource program; databases and knowledgebases were primarily funded as research grants

- **Mechanism:** PARs for Repository and Knowledgebase (U24)
- **Review:** CSR Special Emphasis Panels (SEPs)
- **Awarded:** To date 30 applications have been reviewed. NIH has awarded 7 applications (23% success rate) as of October 2021.
- **Participation:** 10 ICs - NCI, NHLBI, NHGRI*, NIAAA*, NIAID*, NIDA, NIEHS*, NIGMS*, NIMHD, NINDS*

* Institutes with awarded applications

Application Requirements

Program Requirements:

Scientific significance of proposed resource

Demonstrate community needs and engagement

Address quality of data, services, & efficiency of operations

Document & implement governance processes

Additional Requirements:

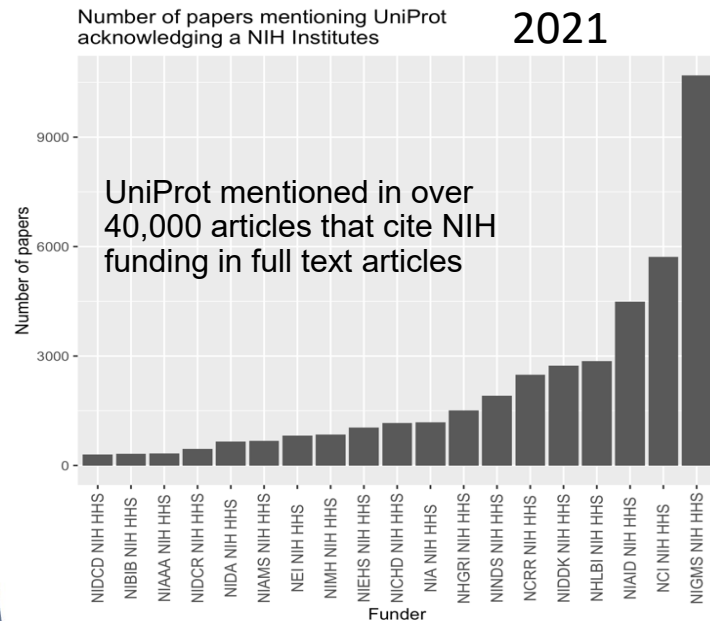
Project Management Plan

Resource Sharing Plan

Resource Sustainability Plan

UniProt: A measure of success

UniProt aims to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.



Usage & Impact:

- Userbase has grown from 500,000 to over 800,000, a larger increase than in prior years
- **"UniProt: the universal protein knowledgebase in 2021."** *Nucleic Acids Research* 49, no. D1 (2021): D480-D489.
 - Cited in 750 articles in 2021; almost 3400 citations from original papers

- Mentions of UniProt in patents has significant increased to almost 3000
- Mentions of UniProt, in all PubMed papers increased from 0.25% to 0.75% over the last decade
- New metrics developed, e.g. co-authorship network

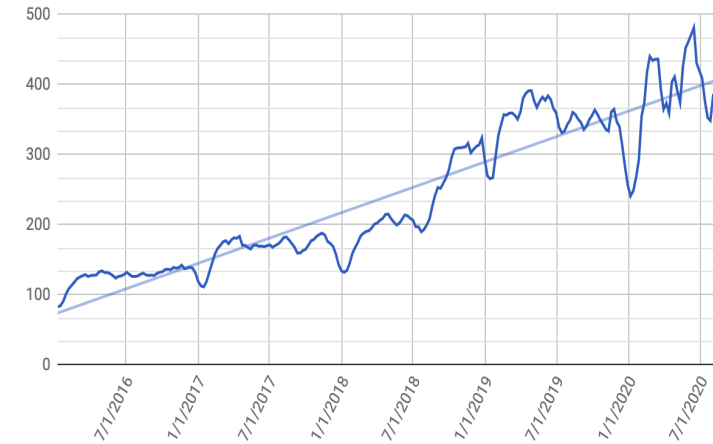
BioPortal: A Definitive Source in Ontology Knowledge

Usage & Impact:

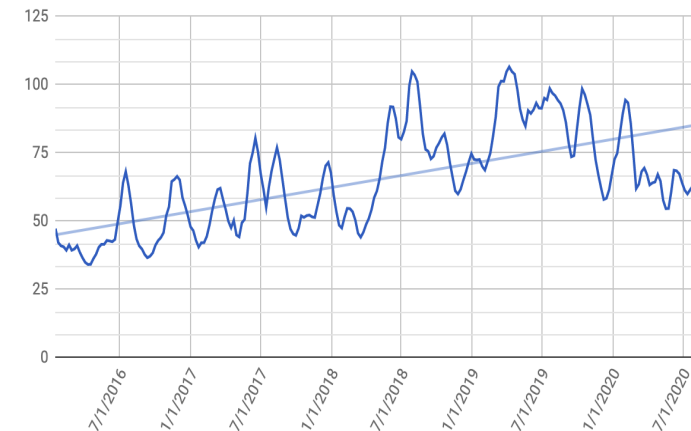
- Interaction with the BioPortal has grown from 300,00 to over 400,000 per month since funding; with 12,000 registered users
- **BioPortal:** enhanced functionality via new web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic acids research*. 2011 Jun 14
 - Cited in 732 publications



UI Page Views (thousands per month)



REST API App Server Requests (millions per month)



over 64 million API calls/month

International Committee on Taxonomy of Viruses (ICTV)

The only authoritative source of taxonomic information on viruses;
enriches the virus taxonomy with available virus properties

Usage & Impact:

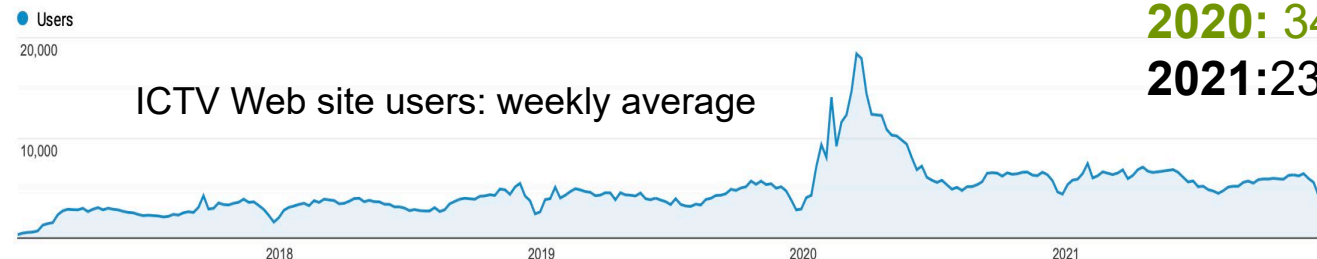
- Total unique users: 951,744; returning users: 17%
- Countries of origin: 230
- **Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)** *Nucleic Acids Research*, Volume 46, Issue D1, 4 January 2018
 - Cited in 508 publications, mentioned in over 5,000
 - Virus publications containing an ICTV taxon name: 774,031

Growth in Users

2017: 97,496
2018: 138,237
2019: 167,300
2020: 342,891
2021: 233,480



<https://ictv.global/>



Reissue of Program supports Resource Life Cycle

PAR 1: Early-Stage biomedical repositories or knowledgebases

- **Objective:** Fund early-stage data repositories or knowledgebases
- **Mechanism:** U24
- **Duration of Award:** Up to 4 years
- **Budget:** \$300K direct costs per year
- **Due Dates:** 2 receipt dates per year, Sept and Jan

PAR 2: Established biomedical repositories or knowledgebases

- **Objective:** Fund established data repositories or knowledgebases
- **Mechanism:** U24
- **Duration of Award:** Up to 5 years
- **Budget:** No limit; must reflect actual needs of the project.
- **Due Dates:** 2 receipt dates per year, Sept and Jan

Reviewed by CSR Special Emphasis Panel

Require Letter of Intent and encourage ICO consultation prior to proposal submission

Program aligns with the stage of the repository or knowledgebase

Early-stage
DR-KB



Established
DR-KB

Early-stage development of biomedical repositories or knowledgebases

Early-stage is defined as initial development of a data repository or knowledgebase with potential for an increase in usage and adoption in the community.

Established repositories or knowledgebases

Established repositories and knowledgebases are defined as high-value data resources that are near or at optimal research community penetration and are critical research resource as demonstrated by their usage, utility, and impact.

Focus of the Reissue

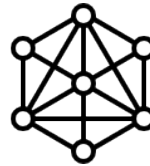
What is new in the reissue

Include domain-specific mission as a science driver

Adopt best practices - NIH desirable characteristics of repositories

Require open metrics – KPIs to measure scientific impact

Align support with stage of the repository or knowledgebase



Benefits

Lower barriers for data sharing; reduce or eliminate silos

Allow discovery and use of data (increase transparency)

Optimize efficiency of operations and costs

Disentangle data resources from research projects

This program is fundamental to creating a modernized biomedical data ecosystem

Other Data Resource Models

ELIXIR Core Data Resources

<https://elixir-europe.org/platforms/data/coredata-resources>

Resource indicator Categories*:

- Scientific focus; quality of science
- Community served
- Quality of service
- Legal and funding infrastructure, and governance
- Impact and translational stories

* Durinx C, McEntyre J, Appel R et al. Identifying ELIXIR Core Data Resources. F1000Research 2017, 5(ELIXIR):2422



DOE Public Resources (PuRe)

<https://science.osti.gov/Initiatives/PuRe-Data>

Resources require:

- Stewardship
- Oversight model
- Data management
- Resource operations
- Scientific impact

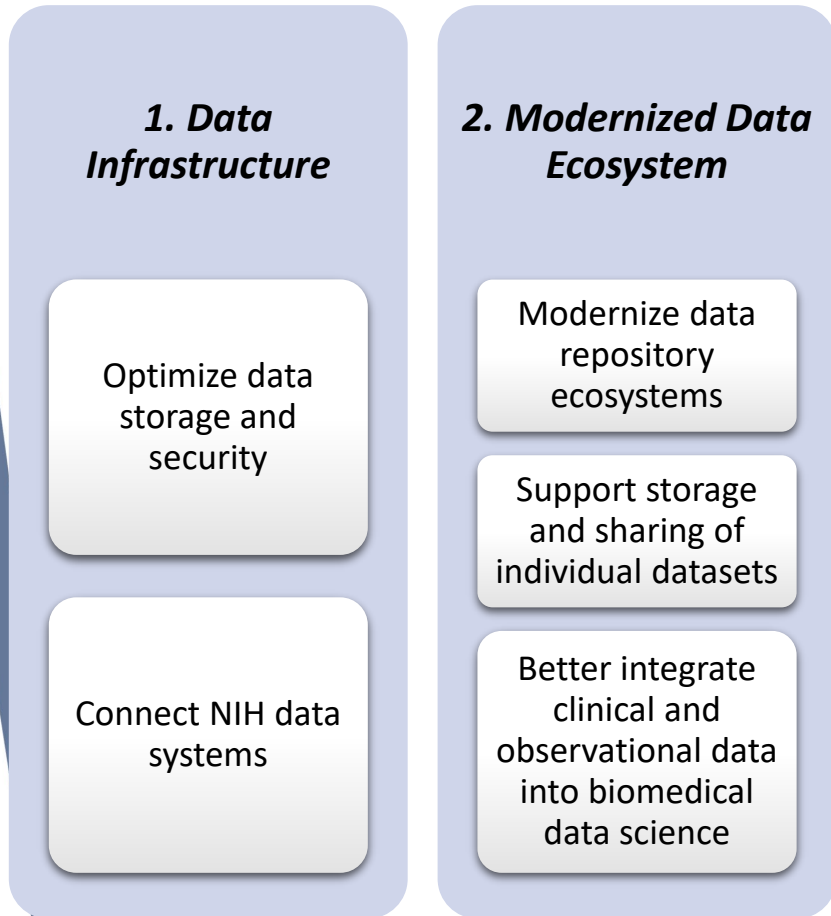


The Global Biodata Coalition announced (1/17/2022) the initial round of Global Core Biodata Resource selection based on the above ELIXIR indicators.

<https://globalbiodata.org/scientific-activities/>

Modernizing the NIH Data Ecosystem

Data resources are key enablers of modern biomedical research



Preparing the NIH data ecosystem for more effective sharing and the new [NIH Data Management and Sharing Policy](#) (NOT-OD-21-013)

- 1. Strengthen existing NIH data & repositories** (through FY21 & 22 supplements)
 - Improve their alignment to the [Desirable Characteristics for Data Repositories](#)
 - Improve the AI/ML-readiness of NIH-supported data
 - Support enhancement of software tools for open science
- 2. Supporting data types that may not have a home**
 - Incorporating generalist repositories into the NIH data ecosystem (Generalist Repository Initiative)
- 3. Establishing a Data Repository and Knowledgebase Ecosystem** (reissue funding opportunities for NIH data resources)
 - Biomedical Data Repository FOA ([PAR20-089](#))
 - Knowledgebase FOA ([PAR20-097](#))

Concept Clearance Vote

Concept Clearance: Reissue (PAR-20-089, PAR-20-097)

Title: Early-stage and Established Biomedical Data Repositories (DR) or Knowledgebases (KB) (**Mechanism U24**)

Objective/Purpose: To seek data repository and knowledgebase resource projects that deliver scientific impact, employ and promote good and efficient data management, engage the user community, and govern to address biomedical data life-cycle and long-term preservation.

Funds Available and Anticipated Number of Awards: Contingent upon NIH appropriations and submission of meritorious applications

Award Project Period: up to 5 years

Council Action: Vote for continued support of Biomedical Repositories and Knowledgebases

Discussion

FY 20-21 Awards

Project	Principal Investigator	Institution	Title
GM143402-01	MUSEN, MARK A	STANFORD UNIVERSITY	BioPortal: An Expansive Knowledgebase of Biomedical Entities and Relations
AI162625-01	LEFKOWITZ, ELLIOT	UNIVERSITY OF ALABAMA AT BIRMINGHAM	Virus Taxonomy: A Community Knowledgebase Supporting Virus Research
ES033155-01	MATTINGLY, CAROLYN	NORTH CAROLINA STATE UNIVERSITY RALEIGH	Comparative Toxicogenomics Database (CTD)*
GM142435-01	HALFON, MARC S	STATE UNIVERSITY OF NEW YORK AT BUFFALO	REDfly: The regulatory sequence resource for Drosophila and other insects
HG007822-08	BATEMAN, ALEX	EUROPEAN MOLECULAR BIOLOGY LABORATORY	UniProt: A Protein Sequence and Function Resource for Biomedical Science *
NS122732-01	FERGUSON, ADAM	UNIVERSITY OF CALIFORNIA, SAN FRANCISCO	Pan-Neurotrauma Data Commons
AA029959-01	WU, SAMUEL	UNIVERSITY OF FLORIDA	Southern HIV and Alcohol Research Consortium Biomedical Data Repository

