NIH Plan to Increase Findability and Transparency of Research Results Through the Use of Metadata and Persistent Identifiers

Background

Each year, the National Institutes of Health (NIH) issues over 58,000 awards that support up to 37,000 researchers conducting research in all 50 states. Through this investment, NIH-supported research results in hundreds of thousands of unique publications and scientific datasets. When publications and scientific data can be quickly identified and contextualized within the landscape of biomedical and behavioral research, the value of these research outputs increases and can lead to improvements in health. To deliver on the promise of the Nation's investment in research, information uniquely identifying research outputs and information related to who conducted the research, where the research was conducted, when the research was conducted, and how the research was funded must be created and made publicly available.

Recent technological advances in the ways in which scientific information is described and found have made it easier for both researchers and other members of the public to utilize publications and data to conduct biomedical and behavioral research, compare research results, and assess the contributions from individual researchers, institutions, awards, publications, and scientific data. NIH recognizes that more can be done to increase access to research outputs by consistently using metadata describing the "who, where, when, and how" of research outputs. Creating and sharing standardized, publicly available metadata on research outputs will enable researchers, clinicians, students, and other members of the public to rapidly locate, contextualize, and analyze the outcomes of research. Proper metadata use will also allow NIH to track the outputs of research investments, offering added transparency and public accountability of federal funds spent on biomedical and behavioral research. Consistent use of identifiers for research outputs and the "who, where, when, and how" will enable cross-linking of the research outputs, as well as the researchers and awards.

In this Plan, NIH presents a combination of efforts that are already underway and planned efforts that will increase the findability of research outputs and promote transparency in the research process, ultimately bolstering the impact of NIH's research while fostering trust in the research that NIH supports. This Plan also addresses expectations from the fourth section of the 2022 White House Office of Science and Technology (OSTP) [Memorandum on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research](#) (the 2022 OSTP Memo) and is an update to section IV, "Mechanisms to Increase Findability and Transparency of Research" of the [NIH Plan to Enhance Public Access to the Results of NIH-Supported Research](#). Specifically, NIH is outlining how researchers and institutions should use metadata to consistently identify themselves and their research outputs, as well as requirements for NIH-supported (i.e., both NIH-funded and NIH-managed) repositories to collect and make relevant metadata publicly available and searchable. These efforts are informed by NIH's experience in managing and hosting repositories and are intended to increase the findability and usability of all NIH research.

Below, the OSTP expectations are listed in italics, and the responsive sections of this Plan are provided in bold, immediately after each point.

Federal agencies should, consistent with applicable law:
  a) *Collect and make publicly available appropriate metadata associated with scholarly publications and data resulting from federally funded research, to the extent possible at the time of deposit in a public access repository. Such metadata should include at minimum:*
      i.   *all author and co-author names, affiliations, and sources of funding, referencing digital persistent identifiers, as appropriate;*
      ii.   *the date of publication; and,*
      iii. *a unique digital persistent identifier for the research output;*

Section II outlines NIH's commitment to and additional plans for collecting and making publicly available appropriate metadata and digital persistent identifiers for publications and data. Section II.B. specifically outlines planned requirements for providing digital persistent identifiers for data in NIH-supported repositories.

  b) *Instruct federally funded researchers to obtain a digital persistent identifier that meets the common/core standards of a digital persistent identifier service defined in the NSPM-33 Implementation Guidance, include it in published research outputs when available, and provide federal agencies with the metadata associated with all published research outputs they produce, consistent with the law, privacy, and security considerations.*

Sections I.A. and I.B. outline NIH's existing requirements and guidance for researchers to obtain a digital persistent identifier and additional expectations to include it in published research outputs. Section I.C. also offers guidance for researchers to provide appropriate metadata with data submissions.

  c) *Assign unique digital persistent identifiers to all scientific research and development awards and intramural research protocols that have appropriate metadata linking the funding agency and their awardees through their digital persistent identifiers.*

Section III outlines NIH's plans for providing digital persistent identifiers for awards.

NIH Actions to Increase Findability and Transparency of Research Results Through the Use of Metadata and Persistent Identifiers

NIH recognizes the value of the diverse repository systems that already exist. This Plan is intended to maintain flexibility for researchers and minimize administrative burden. This Plan is not intended to be prescriptive for other uses of metadata and related identifiers, such as metadata describing data types, research methods, reagent identifiers, or related demographics or unique identifiers assigned to individual research participants. This Plan only addresses metadata for research outputs (specifically, publications and scientific data), awards, and researchers.

Additionally, for this Plan, the term "repositories" refers to archives, databases, and systems designed to accept data submissions (including bibliometric and scientific data) to organize, validate, archive, preserve, and distribute data, whether unrestricted or controlled, to the broader scientific community and the public. Metadata are defined as information conveyed with publications and scientific data upon deposit in a repository to ensure proper attribution and versioning. In this Plan, metadata include information on author names and affiliations, funding, and publication dates. Persistent identifiers (PIDs) are a type of metadata that are defined in this Plan as globally unique, persistent, machine resolvable and processable, and have an associated metadata schema. Examples of PIDs include Digital Object Identifiers (DOI) and Open Researcher and Contributor ID (ORCID) iDs.

I. Using and Submitting Metadata and PIDs

Greater use of standardized metadata and PIDs by researchers when submitting manuscripts and data, applying for funding, reporting research results, and reporting progress on research projects will enable more robust cross-linking of authors, funding sources, and research outputs. The ongoing and planned efforts outlined in this section will help ensure that metadata and PIDs for researchers and research outputs are used to the maximum and appropriate extent, making research outputs easily findable and usable for researchers, clinicians, students, and community members alike.

I.A. Obtaining an ORCID iD
Existing NIH policies will soon require researchers to obtain and uniquely identify themselves using a PID. By using a standard identifier, researchers can easily connect their research outputs, allowing other researchers, institutions, and funders to transparently track outputs, even across different repositories and research platforms. This linkage can help ensure that credit is given where credit is due and simplifies the process for researchers when submitting funding applications, research publications, and more.

Plan: Starting in May 2025, NIH will require all senior and key personnel named in NIH grants to link their eRA profiles with a free, uniquely assigned ORCID iD (see Notice [NOT-OD-24-](#)

163). Senior and key personnel will be required to use their ORCID iD in the [SciENcv](#) system to submit Common Forms (for the Biographical Sketch, Current and Pending (Other) Support) and the NIH Biographical Sketch Supplement. This builds on the 2019 requirement for individual trainees, scholars, or others who are supported by specific training, fellowship, research education, and career development award mechanisms to link their eRA profiles with an ORCID iD (see Notice [NOT-OD-19-109](#)).

All NIH intramural principal investigators will also be required to obtain an ORCID iD, with the requirement for other scientists as determined by the Deputy Director for Intramural Research.

I.B. Including Metadata and PIDs when Submitting Manuscripts
Promoting the widespread use of ORCID iDs for researchers will reduce author ambiguity by uniquely distinguishing the individual from other researchers with the same or similar name and continuing to uniquely identify the same individual across any name changes. Using ORCID iDs can also reduce the administrative burden for researchers when submitting manuscripts and updating digital curriculum vitae. For many journal submission systems, ORCID iDs are a log-in option for submitting authors.

Plan: For new manuscripts that were generated with NIH support, NIH plans to:
- Expect NIH-supported institutions to ensure all authors who are named senior and key personnel use ORCID iDs when submitting new research manuscripts to the NIH Manuscript Submission system and the journal submission system, as allowed.
- Expect NIH intramural principal investigators to use ORCID iDs when submitting new research manuscripts to the NIH Manuscript Submission system and the journal submission system, as allowed.
- Expect NIH-supported authors to submit metadata consistently, such as affiliations and names, when ORCID iDs are not accepted.

ORCID iDs can also be used to link co-author information. To that end, as allowed by the NIH Manuscript Submission system or the journal submission system, NIH plans to:
- Encourage NIH-supported institutions to consider the use of ORCID iDs for all other authors of a manuscript that was generated with NIH support.

I.C. Submitting Metadata and PIDs when Depositing Scientific Data in Repositories
Scientific data resulting from NIH-supported research should be appropriately shared under the [NIH Data Management and Sharing Policy (DMS)](#). Sharing data without additional context substantially limits the ability of data to be reused for testing the validity of research findings, combining datasets for other analyses, or exploring new frontiers of discovery. Submitting appropriate metadata and PIDs to repositories is a fundamental part of ensuring that data can be found and properly contextualized. Maximum metadata use can also help improve data usage and increase citations rates.

Under the NIH DMS Policy, NIH already expects researchers to select the most appropriate data repository, if particular repositories are not specified by NIH and/or Institute, Center, Office (ICO) policy(ies) and Notice of Funding Opportunity(ies) (NOFOs). When selecting a scientific data repository, researchers are encouraged to consider the Desirable Characteristics of Data Repositories to enable the discovery, reuse, and citation of data, using schema that are appropriate to, and ideally widely used across, the community(ies) that the repository serves (see Notice NOT-OD-21-016 and Guidance by the Subcommittee on Open Science).

Plan: When depositing scientific data, submitters are encouraged to meet the repository's best practices, which may further require or encourage submitters to provide additional metadata. Submitters should provide such metadata, consistent with the law and considering privacy and security. When submitting data and as allowed by the data repository (including controlled-access repositories), NIH also plans to:
- Expect institutions to ensure that submissions of scientific data that were generated with NIH support include the following metadata:
  - ORCID iDs/PIDs and names for contributing senior and key personnel,
  - affiliations (or other PIDs for affiliations[1]) for contributing senior and key personnel, and
  - funding sources.
- For all other contributors, encourage institutions to consider consistently providing these metadata and PIDs to the repository, as able.

NIH recognizes that data repositories are varied and continue to evolve, and that at this time, some repositories may not meet all of the Desirable Characteristics of Data Repositories. NIH does not intend for this expectation to prohibit researchers from selecting repositories that may otherwise be the most appropriate choice. NIH plans to continue to work to improve the options available to the scientific community.

I.D. Reporting PIDs to NIH
It is essential that NIH demonstrates accountability to the public, in terms of where its money is invested and what the research investment produces. Including PIDs for publications and shared scientific data in funding applications and proposals ensures proper attribution and unambiguous reference to the research output. Similarly, including PIDs in research performance progress reports and on annual reporting guarantees proper attribution and reporting to NIH, allowing linkage between investment and research outputs.

---

[1] Because affiliation information can become ambiguous over time if inconsistently reported, award recipient organizations may optionally work with PID providers such as the Research Organization Registry (ROR) or International Standard Name Identifier (ISNI). Alternatively, award recipient organizations may encourage any NIH-supported research contributors within their organization to use a defined, consistent affiliation text and a specific PID, as available, when reporting research outputs.

Plan: NIH will expect:
- NIH-supported institutions and NIH intramural principal investigators to provide PIDs for publications and shared scientific data when citing these products in NIH funding applications or proposals.
- NIH-supported institutions and NIH intramural principal investigators to include PIDs for publications and shared scientific data in research performance progress reports (see Notice [NOT-OD-24-175](#)) or annual reporting, as able.

I.E. Citing and Cross-Linking Metadata and PIDs
Identifying research outputs with the "who, where, when, and how" of the research will improve the utility of research outputs, enable researcher contributions to quickly be assessed, and allow for a better understanding of the impact of NIH support. Connecting scientific data with their associated publications to the maximum extent possible will bolster their utility in conducting additional analyses.

Plan: NIH acknowledges that community practices have already set standards on how and when to cite data in research outputs. NIH seeks to encourage researchers to continue to follow the community best practices. Additionally, NIH plans to:
- Encourage the use of available PID providers for publications, scientific data, and the cross-linking of metadata between publications and/or scientific data.
- Encourage researchers to add their research outputs to their ORICD iD records.

II. Collecting and Making Metadata and PIDs Publicly Available

NIH manages and funds a variety of repositories for publications and data. Some, such as PubMed Central® (PMC), NIH's full-text archive of biomedical literature, have a rich history of collecting and making metadata for research outputs publicly available. The ongoing and planned efforts outlined in this section will allow research outputs to be more easily found and unambiguously linked to the contributing researchers, their affiliations, and funding sources.

II.A. Collecting Publicly Available Metadata and PIDs for Publications at NIH
Publicly available and searchable metadata are critical to discovering and contextualizing published research. NIH is already committed to collecting metadata and PIDs related to publications through PMC.[2] Publications resulting in whole or part from NIH support are required to be deposited into PMC. As part of the submission process, PMC collects publication metadata and makes them publicly available and searchable.

Plan: PMC and PubMed will continue to collect and make publicly searchable all available metadata submitted to PMC, including author names, affiliations, funding information, publication date(s), and publication DOI. PubMed also supports fielded searching on author

---

[2] For a complete list of minimum requirements of metadata for publications, see PMC's [Minimum Data Criteria](#).

identifiers, such as ORCID IDs, that are supplied by the data provider. Other publicly available NIH resources, such as RePORTER and iCite, also index and make searchable many of these same metadata fields.

II.B. Collecting Publicly Available Metadata and PIDs for Scientific Data in NIH-Supported Repositories

Collecting uniform metadata for data can help streamline the data management process by reducing redundancies and improving the efficiency of a repository. Standardized metadata can also help researchers and the public easily find and track outputs from federally funded research, establish the integrity of the data, and use the data. To improve data repository management efficiency and public use, NIH plans to establish a standardized, minimum set of metadata, described below, that repositories should collect and make available to the public.

Plan: Scientific data repositories should be consistent with the NIH Desirable Characteristics of Data Repositories (see Notice NOT-OD-21-016 and Guidance by the Subcommittee on Open Science).[3] NIH plans to:

- Require NIH-supported scientific data repositories to collect and make publicly available the following minimum set of metadata, which may include:
    - submitter name, PID/ORCID iD for submitter, and affiliation (or PID for affiliation) or submitting organization name or PID,
    - names and PIDs/ORCID iDs for all research contributors,
    - affiliations (or PIDs for affiliations) for all research contributors,
    - funding sources (e.g., federal grant number),
    - associated publication PIDs,
    - submission date of the scientific data, and
    - publication date of the scientific data.
- Require NIH-supported scientific data repositories to provide a PID that uniquely identifies a given set of deposited data, if the data do not already have one issued independently by another PID provider, and make the PID for the data publicly available.
- Require NIH-supported scientific data repositories to support search functions for these metadata and PIDs.

NIH does not intend for this minimum set of metadata to prevent repositories from accepting or requiring the submission of any additional metadata, including metadata that make their holdings more interoperable. NIH will coordinate with other federal agencies and members of the scientific community to continue evaluating options on the most appropriate and cost-effective approaches for scientific data repositories to provide PIDs for data and adopt best practices.

---

[3] NIH recognizes that many data repositories already provide identifiers which are unique for a data repository or collaboration or data type but may not meet all the criteria to be a PID. NIH's Generalist Repository Ecosystem Initiative (GREI) suggests a number of existing approaches for NIH-funded data repositories to provide a PID service through collaboration with an existing PID-provider. Other ways to provide PIDs are also acceptable approaches.

III. Assigning Identifiers for NIH Awards and NIH-Conducted Research Projects

To continue to enhance the findability of NIH awards and NIH-conducted research projects, NIH plans to explore the most suitable option for a PID that could be issued for NIH awards and conducted research projects. NIH plans to coordinate with other federal agencies and evaluate the most appropriate options.

NIH currently assigns publicly available identifiers to all awards including grants, cooperative agreements, Other Transaction awards, fellowships, as well as intramural research projects. Within the NIH system, these award identifiers are persistent, unique, and have meaningful structure; however, these identifiers are not registered or indexed to ensure uniqueness beyond NIH. NIH award identifiers are also not machine resolvable and processable. NIH plans to consider exploring avenues to identify NIH awards and NIH-conducted research projects with globally unique, machine resolvable PIDs. NIH plans to coordinate this exploration with efforts of other federal agencies and relevant communities to assess how to best develop a robust, connected ecosystem where institutions, researchers, research outputs, and funding sources are linked consistent with Findable, Accessible, Interoperable, and Reusable (FAIR) Principles.

Conclusion and Next Steps

This Plan will help maximize access to and use of research outputs from NIH-supported research by reducing the ambiguity of the "who, what, where, and when" of research, which will make research outputs findable and transparently described. NIH recognizes that the broader community discussion about metadata and PIDs is still evolving, and additional approaches that maximize the use and utility of PIDs may be available in the future. Community efforts may also inform future actions.

NIH will collect input on this Plan from members of the scientific community, other members of the public, and OSTP; this input will be used in developing draft policy. NIH will also continue to be informed by other federal agencies through the Subcommittee on Open Science of the National Science and Technology Council. NIH intends for a final policy incorporating these elements to be issued by December 31, 2026, with an effective date not more than one year later in accordance with the 2022 OSTP Memo. NIH will iterate on approaches, as needed, to ensure that supported research produces maximal benefits for the scientific community and the public.