

EDRN External Data Sharing and Reuse Policy Funding Cycle V

Consortium Statement

[The NCI's Early Detection Research Network \(EDRN\)](#) is an integrated infrastructure to develop and validate biomarkers and imaging methods for early detection and risk assessment for cancer. EDRN is committed to the timely sharing of resources both within and outside of the EDRN network to assist in primary and secondary data analyses, promote reproducibility of research results, and enable their clinical translation. EDRN Cycle-V funded researchers and NCI program staff work together to ensure that data is broadly shared in a manner consistent with the [EDRN blinding policy](#), EDRN Request For Applications (RFAs) requirements, informed consent of participating human subjects, and approved data sharing plans. The EDRN policy references relevant NIH/NCI policies and guidelines because the EDRN policy builds upon or refines those policies.

Once approved by the Executive Committee, the EDRN Data Sharing Policy will be shared and adopted by all EDRN consortium members and their institutions as a legally binding document. After that, the policy will be reviewed annually by the EDRN Data Sharing and Informatics Subcommittee. If appropriate, proposed changes will be submitted to the EDRN Steering Committee for approval. The policy and subsequent updates will be made broadly available on the public [EDRN website](#).

Purpose

This document was developed by current members of the EDRN consortium and NCI officials to facilitate the external (broad) access and use of experimental data, clinical data, metadata, and algorithms (hereafter referred to as “*data*”) generated by EDRN Cycle-V funded efforts. *Data* does not include lab notebooks, preliminary analyses, peer reviews, or physical objects. Material, specimen, and preclinical model use agreements are described elsewhere. The document compliments “EDRN Internal Data Sharing Agreements” and other documents that guide data sharing among the consortium members and collaborators. These guidelines apply to both pre-publication and published *data* once it have been made available to the public.

This document will describe data types to be preserved and broadly shared, standards to be applied to *data if applicable*, related tools and software, data repositories, timelines, data access and reuse options, and oversight plans. Broad sharing ensures fair, equitable access and secondary use of data by the wider research community. It is not equal to open data and should never compromise participant privacy or program goals (e.g., validation of cancer biomarkers).

This document aims to balance the needs of data producers with the goal of advancing the cancer research enterprise, while maintaining compliance with NIH guidelines and policies.

Applicable NIH Data Sharing Policies

Since all EDRN Cycle V applications were submitted before January 25, 2023, they fall under several NIH Data Sharing policies that are specific to certain types of research (e.g., clinical trials, large-scale genomic data - **see Table 1**). These policies complement expectations outlined in Notices of Funding Opportunity (FOAs) and individual data sharing plans. Under the 2003 Data Sharing Policy ([NOT-OD-03-032](#)), EDRN-funded investigators are expected, to the extent possible, to make the primary data (initial raw data) publicly accessible *no later than the time of an associated publication acceptance of the main findings from the final dataset or the end of the performance period of the award*. For definitions see [Appendix A](#).

The 2003 NIH policy requires an approved data-sharing/resource-sharing plan for [large applications](#) (awards of \$500k direct cost per year and over). Applications submitted after January 25, 2023, must follow the [2023 NIH Data Management and Sharing Policy \(DMS\)](#), which includes a submission of a Data Management and Sharing (DMS) Plan within the funding application or proposal. The two policies are very similar, but the new DMS Plan spells out all the expected elements (data types, tools, standards, data preservation, access, and reuse considerations).

Applications subject to the NIH 2014 Genomic Data Sharing (GDS) Policy ([NOT-OD-14-124](#)) must submit an [Institutional Certification](#) (IC) and a Data Sharing Plan ([DSP](#)) at Just-In-Time (JIT). The GDS policy requires that primary genomic data be shared [once it has been cleaned](#) (e.g., the analytic dataset is finalized). Any exceptions to the guidelines due to co-funding agreements or legal agreements (e.g., CRADAs) must be prospectively reviewed by the NCI and will only be granted in rare circumstances.

NCI program staff determines if the project falls under the GDS policy before the award is made and can request revisions if needed. They may request an [NCI DSP template](#) or a full DSP plan, which will help determine if data submission needs to be mandated based on the GDS policy. NCI program staff can also [mandate data sharing](#) for smaller-scale genomic projects focused on examining rare cancers or rare-cancer subtypes, under-studied populations, mitochondrial DNA sequencing, etc. For examples of genomics based data types and data levels, see [Appendix A](#) and the [NIH guidance on Data Submission Expectations](#).

Genomic data, which is not subject to GDS, does not have to be submitted to NIH-supported repositories and does not need Institutional Certification. However, investigators still have the responsibility to ensure the protection of privacy and confidentiality using de-identification methods, choosing a repository that has security and control plans in place, etc.

Investigators conducting NIH-funded clinical trials are also expected to register their studies with the www.clinicaltrials.gov per policy (<https://grants.nih.gov/grants/guide/notice-files/not-ca-15-011.html>).

Table 1. NIH and NCI Data Sharing and Public Access Policies.

Policy	Final Policy Link	Effective Date
2003 NIH Data Sharing Policy	NOT-OD-03-032	10/01/2003
2023 NIH Data Management and Sharing Policy (does not apply to EDRN cycle V)	NOT-OD-21-013	01/25/2023
NIH Genomic Data Sharing Policy	NOT-OD-14-124	01/25/2015
NCI Clinical Trial Access Policy	NOT-CA-15-011	01/28/2015

Data types that fall under these policies may include (but are not limited to) the following:

- Genomics and other -omics
- Clinical Research including clinical data and clinical trial reporting
- Imaging (Medical, Cellular and Histopathology)
- Social, Behavioral, Epidemiology, Population, Surveillance and Implementation Science
- Preclinical Models and Biological Studies
- Biochemical, Biophysical and Immunological Studies
- Metadata - data that describes other data at the study, file, and individual level
- Algorithms and Models

Data must be of sufficient quality to validate and replicate research findings. Data that are not provided publicly need an accompanying explanation with the exception. For example, datasets that cannot be practically digitized with reasonable efforts or datasets linked to a blinded reference sample set.

Justifiable ethical, legal, and technical considerations for limiting sharing of data include:

- Explicit federal, tribal, state, local, or tribal law, regulations, statutes, guidance, and institutional policies prohibit the sharing and use of scientific data generated from research involving human participants
- Privacy, safety, rights, or confidentiality of human research participants would be compromised
- Participants would be at greater risk of re-identification and protective measures such as de-identification and Certificates of Confidentiality would be insufficient
- Informed consent will not permit or limit the scope of sharing or future research use
- Existing consent (e.g., for previously collected biospecimens) prohibits sharing or limits the scope or extent of sharing and future research use
- Restrictions imposed by existing or anticipated agreements with other funders, partners, repositories, Health Insurance Portability and Accountability Act (HIPAA) covered entities that provide Protected Health Information under a data use agreement
- Datasets cannot be practically digitized with reasonable efforts

Reasons not generally justifiable to limit sharing include:

- Data are considered too small
- Researchers anticipate data will not be widely used
- Data are not thought to have a suitable repository

EDRN-specific Data Sharing Expectations

The EDRN Data Sharing Policy aims to be consistent with and build upon the above NIH guidelines and policies to enable the following:

- an open and collaborative environment, including the timely release of both pre-publication and published *data* to the broader scientific community;
- opportunities for use of *data* by the scientific community and for publication of research using such data in high-impact journals;
- protection of human subjects' privacy and confidentiality of their data;
- protection of intellectual property; and
- appropriate recognition and citation of both contributors and users of *data*.

To comply with the RFA-specific requirements ([RFA-CA-21-033](#), [RFA-CA-21-034](#), [RFA-CA-21-035](#), [RFA-CA-22-039](#), [RFA-CA-22-054](#), and [RFA-CA-22-040](#)), all EDRN projects, regardless of the amount of direct costs requested for any one year, must have a Data Sharing Plan (DSP) or Data Management and Sharing Plan (DMSP). The plan describes how investigators will manage and share data during the entire funded period. This includes information on data storage, access policies/procedures, preservation, metadata standards, and distribution approaches.

Additionally, EDRN investigators will follow terms and conditions incorporated into their Data Sharing Plan/Notice of Award and consistent with the informed consent of human subjects participating in these projects. If there are restrictions limiting data sharing (e.g., a lack of informed consent, legal restrictions by state or tribal laws), investigators will describe such limitations in the DSP and provide reasonable justifications and alternate data sharing plans whenever possible.

EDRN-specific justifiable considerations for limiting sharing of data include the following:

- The EDRN blinding policy will not permit or limit the scope of sharing or use; e.g., data associated with [Specimen Reference Sets](#) that have not been yet exhausted

Since its inception, the EDRN has established rigorous biomarker development guidelines including biomarker reference sets and the “arm-length-away” blinding policy. Under this policy, a biomarker discovery lab should not be driving the validation trial. Discovery lab biomarkers are validated by an EDRN Biomarker Reference Lab. The validation study is coordinated by the EDRN Data Management and Coordinating Center (DMCC). For this reason, data and samples associated with validation studies and reference sets should remain blinded until the samples are exhausted.

EDRN-specific Expectations for Repository Selection for Broad Sharing of Open Access Data

EDRN investigators are encouraged to submit their Open Access Data (open data) to LabCAS to enable broad data sharing with the research community. See [Appendix A](#) for definitions and details. Secondary data users will have access to open EDRN data through the [EDRN Public Portal](#) without user authentication or authorization.

In addition to LabCAS, open EDRN data can be broadly shared through NIH- and NCI-supported repositories (e.g., see [BMIC repositories](#)). The choice of repositories for broad data sharing for EDRN investigators is not prescriptive, as long as the repositories provide persistent unique identifiers (PIDs), defined minimal metadata sets, quality assurance protocols, and tools to track data reuse ([NOT-OD-21-016](#)) and promote FAIR Principles ([www.re3data.org](#) and [www.fairsharing.org](#)). Data in these repositories is considered pre-competitive and will not be protected by intellectual property patents.

EDRN RFAs do not identify particular data repositories to be used for broad data sharing. EDRN investigators will select a data repository taking into consideration the following considerations:

- Use a domain-specific repository if available (see https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html);
- If no domain-specific repository is available, use a generalist repository (see https://www.nlm.nih.gov/NIHbmic/generalist_repositories.html) or institutional repositories, that make data available to the larger research community;
- Use cloud-based data repositories for large datasets;
- Include small datasets as supplementary material to accompany articles submitted to PubMed Central (see <https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/#suppm>);
- Use appropriate repositories for open and controlled data as detailed below

NIH Expectations for Repository Selection for Broad Sharing of Controlled Access Data

EDRN investigators will follow the best practices for protecting privacy when sharing human research participant data to meet the Common Rule and the Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule standards. Controlled data (even de-identified) is subject to access restrictions and will be made available to qualified researchers through controlled-access NIH repositories or NIH-trusted partners. Such data will be de-identified to the greatest extent possible while preserving data utility and shared only with explicit consent from the study participants.

To safeguard the privacy of human subjects participating in EDRN-funded research projects, sensitive data will be made broadly available via Controlled Access systems. Based on current NIH policies, access to data in the Controlled Access Data tier (controlled data) requires authorization and eRA Commons authentication. Controlled data generally includes individual-level data which use is restricted due to contractual terms or potentially individually identifiable data such as sequencing data (BAM, FASTQ files), germline variants, some images, and certain clinical data elements. In addition to genetic sequence or genotype data from living individuals,

controlled data may include data from “vulnerable” populations as defined using OHRP guidelines and data generated with restrictions or requirements for use as outlined in informed consents or legal agreements.

Genomic and transcriptomic data backed by EDRN will be shared in a manner consistent with data-sharing under the *NIH Genomic Data Sharing Policy* ([NOT-OD-14-124](#)). Suitable controlled-access repositories for genomic and phenotypic data include the Database of Genotypes and Phenotypes ([dbGaP](#)), the Sequence Read Archive (SRA), and NIH-designated [Trusted Partnerships](#) such as the NCI [Cancer Research Data Commons \(CRDC\)](#). Imaging data will be shared through The Cancer Imaging Archive ([TCIA](#)). Additionally, small image data linked to controlled genomic or clinical data will be submitted to the [NIH Database of Genotypes and Phenotypes \(dbGaP\)](#).

Prior to data submission, all controlled-access EDRN data will be registered through [dbGaP](#) to establish data use agreements and data use limitations for each dataset in compliance with the participants’ informed consent, protection of intellectual property and NIH policies, as summarized in the [NCI Data Sharing Guidelines](#). For all controlled access NIH repositories, access and submission of controlled-access datasets will be managed through [dbGaP](#) to enable a unified identity and access management for approved users.

NIH Expectations for Guidelines for Data Release

EDRN investigators are committed to timely release of data to meet the [NIH expectations](#). The data will be released via the EDRN Data Portal or other established data repositories. NIH data release expectations depend on the data level, which is defined by the extent of data processing information abstraction that has occurred. Examples of genomic data levels are shown [here](#). For other data types, such as proteomics and imaging, specific data levels may need to be worked out. Additional data types and timelines may be developed and modified by the Data Sharing Committee if needed.

For how long to share data, EDRN investigators will consider relevant expectations, including repository policies, record retention requirements from funders (NIH), and journal policies for minimum time frames and how long data will be useful.

NIH Guidelines for Data Use

The NIH guidelines aim to enable broad use and publication of data by the wider scientific community while, at the same time, protecting intellectual property and ensuring appropriate recognition and citation of both the generators and users of data. In keeping with this, data generators agree not to arbitrarily restrict the use of the data by others, and data users are strongly encouraged to act in a manner that is consistent with this policy in its entirety. EDRN data in broadly shared through public repositories is considered pre-competitive and will not be protected by intellectual property patents.

EDRN supports the use of data for a variety of purposes, such as development of novel analytical methods, identification of novel biomarkers for cancer prevention, detection, and diagnosis, and pathway characterization to facilitate the development of novel therapeutic interventions. Although all reasonable efforts have been taken by EDRN members to ensure the accuracy and reliability of data upon release to the public, such data is provided “as is”, and EDRN does not and cannot provide any warranty regarding the results that may be obtained from its use.

EDRN-specific Guidance for Data Use Agreements and Intellectual Property

All users of EDRN *data* are expected to adhere to these guidelines and comply with restrictions imposed by relevant regulatory bodies and the informed consent provided by the subjects from whom the data were derived. Currently, to gain access to any EDRN datasets, data users are required to sign a Data Use Agreements (DUA). For non-sensitive data, the requirement for a DUA may be removed in accordance with the 2023 NIH Data Sharing Policy ([NOT-OD-23-053](#)).

The data generator retains ownership of the data and materials they have generated and provided under a data use agreement (DUA) signed by data users. A DUA template will be developed by the Data Sharing Subcommittee. To maximize public benefit, EDRN encourages secondary users of *data* generated to avoid making intellectual property (IP) claims derived directly from primary *data* by EDRN grantees. This way all *data*, and conclusions derived therefrom, remain freely available, without requirement for licensing.

EDRN also recognizes the importance of developing IP on downstream discoveries to support full investment in products to benefit public health. If necessary, the *data* will only be shared under licenses that retain IP for commercialization. In such cases, EDRN expects users of *data* to implement licensing policies that do not obstruct further research, as outlined in the [NIH Best Practices for the Licensing of Genomic Inventions](#) and the [NIH’s Research Tools Policy](#).

EDRN-specific Guidance for Pre-publication Use

Data released to public databases is considered pre-publication until it appears in a peer-reviewed publication. All users of *data*, whether members of EDRN or of the wider community, should be aware of the publication status of *data* being used and treat it accordingly. Users of *data* are expected to act in accordance with the [Ft. Lauderdale guidelines](#) regarding use and publication of both pre-publication and published data, as follows:

- For up to one year following public release of data, users interested in publishing on pre-publication *data* should obtain the consent of the *data* generator(s) before using such *data* in their own publications, and the *data* generator(s) should not unreasonably withhold this consent. Initial requests should be made to the EDRN Steering Committee (SC). If a person who is publishing on pre-publication data based on data generated by others and the request is denied, this decision can be appealed to a SC member for further consideration.
- Users interested in publishing on pre-publication *data* should discuss their results with the *data* generator(s) and are encouraged to establish collaborations. However, EDRN members

are not required to collaborate with any outside investigators nor are the data users required to establish a collaboration with an EDRN investigator.

- **Users should cite the source and version of all *data* and should acknowledge the *data* generator(s) as specified in the Attribution section below.**
- All scientific community members, through their roles as journal and grant reviewers, should enforce a high standard of respect for the scientific contribution of *data* generators.

EDRN-specific Accessibility Guidance for Data Users

EDRN promotes dissemination of research findings using *data* as widely as possible through web-based portals, data repositories, scientific publications, scientific presentations, or other appropriate public dissemination mechanisms. As such, EDRN strongly encourages users to publish and disseminate derived data and results in accessible formats through open-access data repositories, preprint servers, and journals.

EDRN-specific Attribution Guidance for Data Users

EDRN expects that *data* users will acknowledge in all presentations and publications the relevant *data* generator(s) and corresponding grant(s) as follows: *“The results published here are in whole or in part based upon data generated by [Name, Center, Institution] DOIXXXX with support from [The NCI’s Early Detection Research Network \(EDRN\)](#), grant(s) [U24CAxxxxxx, U2CCAxxxxxx, and/or P01CAxxxxxx] and/or Task Order No. HHSN261100039 under Contract No. HHSN261201500003I. Citing of EDRN datasets should be done using persistent unique identifiers (PIDs), such as a Digital Object Identifier (DOI), which is used to locate specific digital objects.”*

EDRN-specific Appeal Process for Data Users

If parties are concerned about the compliance with these policies, they should bring the concerns to the attention of the NCI Program Official and chairs of the EDRN Steering Committee and if appropriate, it will be referred to the full Steering Committee for review.

EDRN Investigator:

Sign: _____ Print: _____ Date: _____

Signing official (name):

Sign: _____ Print: _____ Date: _____