

## EDRN Data Sharing Policy

*April 28, 2015 – EDRN Data Sharing and Informatics Subcommittee*

### Purpose

The release and sharing of reproducible, well-defined data is an essential goal for the Early Detection Research Network and the National Institutes of Health (NIH) to accelerate the discovery of biomarkers and technologies for the clinical application of early cancer detection. EDRN funded members are required to capture their data in a stable, public repository for long term preservation and sharing per the NCI/EDRN requirement<sup>1</sup>. EDRN provides a public repository, the EDRN Catalog and Archive Service, for capturing and sharing data.

### Covered Data

This policy is intended to inform the long-term distribution of journal-published data from experiments funded by the Early Detection Research Network (generally, but not exclusively, under BDL, BRL, or CVC grants). If the findings of an EDRN-funded publication report the data or conclusions from the data, they are covered by this policy. Data may be presented orally or in posters at EDRN or other conferences without being covered by this policy.

### Classes of Data

- **Raw data:** if possible the preferred format would be the raw data converted to an open format, as defined by the appropriate discipline, e.g., HUPO PSI, GCS. If that is not easily done the files and databases generated directly by instrumentation will be accepted. EDRN *requires* public accessibility of raw data for journal-published research that it has funded.
- **Scripts and Software Configurations:** the processes by which data analysis has taken place may be described in EDRN Protocols or may be treated as a class of data to be shared. The goal of this sharing is to provide information sufficient for independent reproduction of data analysis. This information may be made available as supplementary information to journal publications. EDRN *recommends, but does not require* public accessibility of scripts and software configurations.
- **Data Processing Outputs:** many data analysis pipelines contain multiple stages, each associated with its own outputs. EDRN *encourages* the inclusion of as many layers of data analytical

outputs as possible to service different communities in data re-use. However, EDRN *requires* the processed data used in the publication.

## Timelines for Data Sharing

Required data from journal-published research articles supported by EDRN funding must be deposited to a public repository within 30 days of article acceptance. It is desirable to have data deposited in external systems referenced and cataloged within the EDRN to show the comprehensive results of EDRN funded research. EDRN will provide support to the investigators to capture data in EDRN repositories that can also be cited.

In some instances there may be scientific reasons why the data may not be shared within this timeline. An example is data from analyses of not-yet-depleted EDRN reference sets, where the EDRN policy is not to release data that could compromise the blind of still-extant reference sets. Investigators may request a waiver of the timeline for data sharing by providing their scientific rationale to the Executive Committee whose decision on whether to grant a waiver is final.

## Compliant Repositories

Researchers may combine EDRN and other open repositories for data sharing. The EDRN Informatics Center maintains the eCAS (Catalog and Archive Service) system for organizing and distributing experimental data sets to the public. NCBI resources such as the Gene Expression Omnibus and the Sequence Read Archive can also house these data and can be referenced by EDRN systems. While data may be housed at non-EDRN sites, the EDRN Informatics Center must be informed of the accession numbers associated with these deposits for each linkage in eCAS. For a repository to qualify as compliant with this policy, it must meet the following criteria:

- Persistent: the repository may be expected to remain accessible for the next five years.
- Private: protected health information must comply with HIPAA Privacy and Security Rules.
- Indexed: the repository allows reference to particular data sets by means of an identifying code.
- Free: the repository must not charge fees for data access.

## Accessibility

The EDRN will provide multiple levels of access to the data during data preparation and peer review to ensure that only authorized users will be able to access the data. Following publication and/or release of data, in accordance with the NIH data sharing policy it will be made public.

## Other Resources

- NIH Data Sharing Policy: [http://grants.nih.gov/grants/policy/data\\_sharing/](http://grants.nih.gov/grants/policy/data_sharing/)
- NIH Health Information Privacy: <http://www.hhs.gov/ocr/privacy/>
- EDRN Public Portal: <http://edrn.nci.nih.gov>
- HUPO proteomics standards initiative: <http://www.psidev.info>
- GSC: <http://gensc.org/projects/mixs-gsc-project/>
- Sharing Data from Large-scale Biological Research Projects Fort Lauderdale, January 2003: <https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>
- The Amsterdam Principles: <http://www.ncbi.nlm.nih.gov/pubmed/19344107>

---

<sup>1</sup> “The Early Detection Research Network: Biomarker Development Laboratories (U24) Request for Application.” *National Institutes of Health*. 19 August 2009. 5 March 2014. <http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-09-017.html>

“The Early Detection Research Network: Biomarker Reference Laboratories (U24) Request for Application” *National Institutes of Health*. 19 August 2009. 5 March 2014. <http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-09-019.html>

“Early Detection Research Network: Clinical Validation Centers (U01) Request for Application”. *National Institutes of Health*. 19 August 2009. 5 March 2014. <http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-09-018.html>