Cancer Registries: Secondary Data Sharing Fact Sheet

CHALLENGE

State cancer registries are the only source of rich, well-curated data on every cancer patient in the US. As such, registries serve as a valuable source of outcome data for downstream research on cancer etiology, prevention, and disparities. Much of this research is funded by the National Institutes of Health (NIH) which requires researchers to share data in compliance with the NIH Policy for Data Management and Sharing (NIH DMS Policy) or other policies, or in response to specific requirements in funding opportunity announcements. However, restrictive and varying state registry data sharing policies limit the researcher's ability to further share data that includes cancer information from the state registries (e.g., secondary data sharing).

PURPOSE

The North American Association of Central Cancer Registries (NAACCR) seeks to eliminate unnecessary barriers to data sharing. The goal is to maximize use of registry data for research and public and individual health interests while protecting patient privacy and supporting researcher compliance with the NIH DMS Policy. Within a partnership of registries, researchers, and standards setters, NAACCR will assist in identifying, developing, or promoting:

- Best practices for sharing of researcher data that incorporates registry data
- Standard registry approaches and policies
- Standard privacy protection policies and procedures
- National repositories to hold, protect and share deidentified data.

WHAT IS THE NIH DMS POLICY?

The NIH Policy for Data Management and Sharing (https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies) reinforces NIH's longstanding commitment to making the results and outputs of NIH-funded research available for further research through effective and efficient data management and data sharing practices. NIH DMS Policy defines data sharing as the act of making scientific data available for use by others (e.g., the larger research community, institutions, the broader public), for example, via an established controlled access repository. The NIH DMS Policy applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data. This includes research funded or conducted by extramural grants, contracts, Intramural Research Projects, or other funding agreements regardless of NIH funding level or funding mechanism.

WHAT IS REGISTRY DATA SHARING?

Registry data sharing is the release of data by cancer registries to external researchers. Nearly all registries can share cancer data for approved research purposes such as cancer epidemiology cohorts and other NIH-funded research. While many state cancer registries receive NIH funding, the NIH DMS Policy does not apply to registries; however, researchers receiving registry data are bound by the NIH DMS Policy.

Prepared by North American Association of Central Cancer Registries (NAACCR), 2/9/23.

WHAT IS RESEARCHER DATA SHARING?

Researcher data sharing is the release of data from an external researcher to a third party. Per the NIH DMS Policy, NIH-funded researchers are required to develop and implement data sharing plans. Typical data sharing methods include (1) researcher to researcher, (2) researcher to pooling project, and (3) researcher to controlled access repository. For studies that incorporate cancer registry data into their datasets and plan to share the augmented data (secondary data sharing), this may require registry approval and documentation in the data use agreement with the researcher. Some registries do not allow researchers to further share data received from the registry and others require complex and inconsistent policies which add cost, delay, and effort by both registry and researcher and, at times, exclusion of otherwise eligible cancer patients.

WHY ALLOW SECONDARY DATA SHARING?

Sharing scientific data augmented with cancer registry outcomes accelerates biomedical research discovery, enhances research rigor and reproducibility, provides accessibility to high-value datasets, and promotes data reuse for future research studies. Ultimately, the sharing of scientific data expedites the translation of research results into knowledge, products, and procedures to improve human health. Data sharing expands the public's investment in cancer research by reducing:

- barriers to investigating rare cancers and exposures.
- effort and costs to collect data
- delay to discovery
- cancer disparities

Limitations on secondary sharing of registry data severely impacts the ability of NIH-funded researchers to meet the NIH DMS Policy and, in turn, the amount of potential knowledge gained about cancer risk, prevention, treatment, and outcomes. An editorial on the ethical obligation to share data for the common good is provided in the Summer 2022 edition of the *Journal of Registry Management* (here).

HOW CAN REGISTRIES HELP?

Cancer registries are integral in advancing researcher ability to share registry data. As such, NAACCR is requesting that registries initiate local discussions about how best to facilitate secondary data sharing in the interests of cancer research and public and individual health at the national level. Below are questions that registries should ask their organization's administration, IRB, and local decision makers:

- What are the concerns about secondary data sharing as required by the NIH DMS Policy?
- On what authority are data sharing prohibitions or limitations, if any, imposed?
- What information would be needed to allow for support of secondary data sharing policies?
- What is the definition of identifiable or deidentified data that is used for data releases?
- Are there specific provisions (e.g. de-identification of data) that could be written into a data sharing agreement to address concerns about secondary sharing of cancer registry?

Registries are also encouraged to adopt the templated Virtual Pooled Registry Data Use Agreement (VPR DUA) which contains terms and conditions that address researcher data sharing, including sharing with pooling projects and repositories. The VPR DUA can be found here: https://www.naaccr.org/vpr-fact-sheets/

Prepared by North American Association of Central Cancer Registries (NAACCR), 2/9/23.

Appendix 1: Terms and Definitions

The NAACCR Secondary Data Sharing Task Force produced these definitions to accompany the Secondary Data Sharing Fact Sheet and establish standard terminology for secondary data release of population-based cancer registry data by downstream researchers.

Term	Definition	Reference
Registry Data Sharing	Registry data sharing is the release of data by cancer registries for inclusion in downstream research.	
Downstream Research	Research that includes registry data, has been approved by a registry (and/or their IRB, if applicable), and proceeds with appropriate agreements in place between the registry and researcher.	
Augmented Data	Registry data combined with other data collected by a third party. For example, cancer registry data linked with national cohort data would be considered augmented data.	
Scientific Data	Under the NIH DMS Policy, this is data commonly accepted in the scientific community as being of sufficient quality to validate and replicate the research findings. This may include datasets generated from existing data. However, not all data generated during NIH-supported research will constitute scientific data and not all data are appropriate to share.	https://sharing.nih.gov/faqs#/data-management-and-sharing-policy.htm
Secondary Data Sharing (see also Researcher Data Sharing)	Further sharing of data collected by the downstream research with a third party, including external investigators, pooling projects, and/or repositories that allow further access. Controlled access repositories are recommended for datasets that include registry data (augmented data).	
Researcher Data Sharing (see also Secondary Data Sharing)	Researcher data sharing is the release of data from downstream research to third party. Per the NIH DMS Policy, NIH-funded researchers are required to develop and implement data management and sharing plans to avoid duplicate data collection. Typical data sharing methods include (1) researcher to researcher, (2) researcher to pooling project, and (3) researcher to controlled access repository. For studies that incorporate cancer registry data into their datasets and plan to share the augmented data (secondary data sharing), this may require registry approval and documentation in the data use agreement with the researcher.	

NIH Data Management and Sharing (DMS) Policy	The NIH Data Management and Sharing (NIH DMS) Policy, effective January 25, 2023, reinforces NIH's longstanding commitment to making the results of NIH-funded research (e.g. scientific data) available for further research through effective and efficient data management and data sharing practices, including controlled access repositories. The NIH DMS Policy applies to all research, funded or conducted in whole or in part by NIH, including extramural grants, contracts, Intramural Research Projects, or other funding agreements regardless of NIH funding level or mechanism.	https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies
De-identified Data	Although cancer registry data are not generally considered to be governed by HIPAA provisions, the "Safe Harbor" method (45 CF 164.514(b)(2)(i)) of removing the 18 identifiers is commonly used as the standard for de-identification. The NIH Genomic Data Sharing Policy is an example of an NIH data sharing policy that explicitly uses the HIPAA safe harbor de-identification standard.	https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html https://sharing.nih.gov/genomic-data-sharing-policy
Controlled Access Repository	While there is no single standard, NIH DMS Policy guidance uses "controlled access" to refer to measures such as requiring data requesters to verify their identity and the appropriateness of their proposed research to access protected data. Historically, "controlled" access contrasted with "public", "unrestricted", or "open" access. At a minimum, the Secondary Data Sharing Task Force recommends that a controlled-access repository meets the following requirements: • Relies on appropriate standards to de-identify data (pursuant to HIPAA 45 CFR 164.514(b)(2)(i)), including using random, unique Subject IDs. • Reviews data access requests. • Controls/audits dataset access and download • Requires a data use agreement with repository recipients that prohibits further data sharing and includes requirements for maintaining confidentiality and destroying data when project is completed. A model of a federal controlled access data repository is the database of Genotypes and Phenotypes (dbGaP) as described by Charlisse Gaga-Anan, JD, MA, Program Director, Epidemiology and Genomics Research Program (EGRP), Division of Cancer Control and Population Sciences (DCCPS), National Cancer Institute (NCI) October 14, 2022.	https://sharing.nih.gov/data-management-and-sharing-policy/protecting-participant-privacy-when-sharing-scientific-data/designating-scientific-data-for-controlled-access https://grants.nih.gov/grants/guide/notice-files/not-od-22-213.html https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf