

Register for the Annual Data Sharing Symposium

National Cancer Institute sent this bulletin at 04/24/2025 09:30 AM EDT

News from NCI's Office of Data Sharing

Subscribe to the Newsletter

April 2025

Register for the Third Annual Data Sharing Symposium

The NCI Office of Data Sharing presents
the Third Annual Data Sharing Symposium:

How Data Advances the Impact of Cancer Research

September 30 – October 1, 2025



[Register now.](#)

On September 30 and October 1, 2025, the NCI Office of Data Sharing (ODS) is hosting the ***Third Annual Data Sharing Symposium: How Data Advances the Impact of Cancer Research*** inside the [NCI Shady Grove Campus in Rockville, MD](#).

Researchers, clinicians, advocates, and, most importantly, patients all contribute to a vast and complex ecosystem of data that can be used to drive the NCI's mission to advance scientific knowledge about cancer to help people live longer and healthier lives.

A foundational shift to a culture of broad data sharing and collaborative science holds immense promise for more rapid advances in cancer research. We invite scientists, clinicians, policymakers, patients, advocates and trainees across government, academia and industry to learn from one another and explore ways to maximize the benefits of these efforts. Please join us for engaging discussions on advancing cancer research through data sharing and reuse.

Platforms Enhancing the Findability of Cancer Data

Data sharing is a powerful catalyst for advancing cancer research. At the NCI Office of Data Sharing, one of our goals is to ensure that NCI-funded cancer data adheres to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) as these tenets accelerate the re-use of cancer data, paving the way for innovations that ultimately improve patient care. Here we will focus on NCI platforms that enhance the **Findability of cancer data** so that it is readily discoverable by the cancer research community. The NCI has a suite of resources that champion data sharing in cancer research: the [Cancer Research Data Commons \(CRDC\)](#), the [Childhood Cancer Data Initiative \(CCDI\)](#), the [Database of Genotypes and Phenotypes \(dbGaP\)](#), the [Surveillance, Epidemiology, and End Results Program \(SEER\)](#), the [Index of NCI Studies \(INS\)](#), and the [NLM Dataset Catalog](#).

A flagship infrastructure program at the NCI, the **Cancer Research Data Commons (CRDC)**, is a secure, cloud-based data science infrastructure designed to empower researchers to accelerate discovery. CRDC connects diverse data types—from pre-clinical to clinical, genomics and proteomics to imaging—through innovative analytical and visualization tools. By harmonizing these data under a standardized model, CRDC enables analyses that drive precision medicine and large-scale collaborative efforts.

The **Childhood Cancer Data Initiative (CCDI)**, created as a model for all cancer data sharing, ushers in a new era of pediatric cancer research by centering data collection on a vulnerable, underrepresented population. Dedicated to gathering comprehensive clinical, molecular, and research data from every child, adolescent, and young adult diagnosed with cancer—no matter where they receive care—CCDI creates an interconnected data ecosystem that accelerates diagnosis, informs personalized treatment protocols, and enhances survivorship.

The **Database of Genotypes and Phenotypes (dbGaP)**, a longstanding platform launched in 2006, serves as a critical repository for deciphering the complex relationships between genetic variations and clinical manifestations. Its controlled-access framework protects patient privacy and enables researchers to leverage these genetic insights to decode the molecular underpinnings of cancer.

The **Surveillance, Epidemiology, and End Results Program (SEER)** provides a population-based perspective essential for understanding cancer trends across the nation. SEER offers robust datasets that inform epidemiological research, guide public health policies, and address disparities in cancer care with data collection on cancer incidence, treatment outcomes, and survival rates.

To enhance the findability of data within the NCI resources mentioned above, the NCI Office of Data Sharing is actively developing the **Index of NCI Studies (INS)**. The goal of the INS is to enable access to information about the large amount of data generated by NCI-supported studies so that NCI leadership, cancer researchers, and the public can rapidly assess the state of science at a single location. Similar efforts exist outside NCI and cancer research data to curate and catalog this information, such as the newly-released **NLM Dataset Catalog beta**, which is curated by the National Library of Medicine and lists biomedical datasets available from various repositories and allows users to search for and navigate specific biomedical datasets by linking descriptive data.

Collectively, these data sharing champions are reshaping the cancer research landscape by breaking down traditional barriers to access, fostering transparency, and accelerating innovative discoveries to ultimately improve patient care. Their integrated platforms and

collaborative nature empower the cancer research community to translate raw data into lifesaving insights, paving the way for enhanced diagnostics, personalized therapies, and improved clinical outcomes.

Announcements

NCI SRP Releases Latest SEER Data and Statistics

NCI's Surveillance Research Program released SEER's latest data and statistics on April 16, 2025. The [SEER website](#) now shows data and statistics from the November 2024 SEER submission and U.S. Mortality through 2023.

This includes:

- [SEER Incidence Data](#), 1975-2022
- [SEER*Explorer](#) - NEW rates and trends by rural/urban county groupings!
- [Cancer Stat Facts Sheets](#)

The updated [Impact of Revisions on 2020 U.S. Population Estimates](#) is included with this release.

The [Surveillance Research Program website](#) has also been updated to reflect the new statistics, including:

- New versions of Joinpoint and DevCan
- Updated Delay Method documentation

Version 2.0.0 of the Bridge2AI-Voice Dataset is Now Available

The Bridge2AI Voice Program recently released **Version 2.0.0** of the **Bridge2AI-Voice Dataset**. It is now available through registered access on [PhysioNet](#). This release includes 16,738 voice recordings from 442 participants, collected across four clinically relevant domains voice disorders, neurological and neurodegenerative conditions, mood and psychiatric disorders, and respiratory diseases.

The dataset represents a significant step toward advancing the use of voice as a digital biomarker in precision medicine. The data were acquired using a harmonized, multi-disorder protocol specifically designed to support research at the intersection of health, acoustics, and artificial intelligence. It is intended to enable a wide range of machine learning and clinical investigations aimed at transforming healthcare diagnostics through voice.

[Learn more about the Bridge2AI Voice Program.](#)

We'd love to hear from you! Contact the ODS team at: nciofficeofdatasharing@mail.nih.gov

National Cancer Institute | National Institutes of Health | U.S. Department of Health & Human Services | [USA.gov](#)

[Manage your subscriptions](#)

Subscribe to updates from National Cancer Institute

Email Address e.g. name@example.com

Share Bulletin



Powered by



[Privacy Policy](#) | [Cookie Statement](#) | [Help](#)