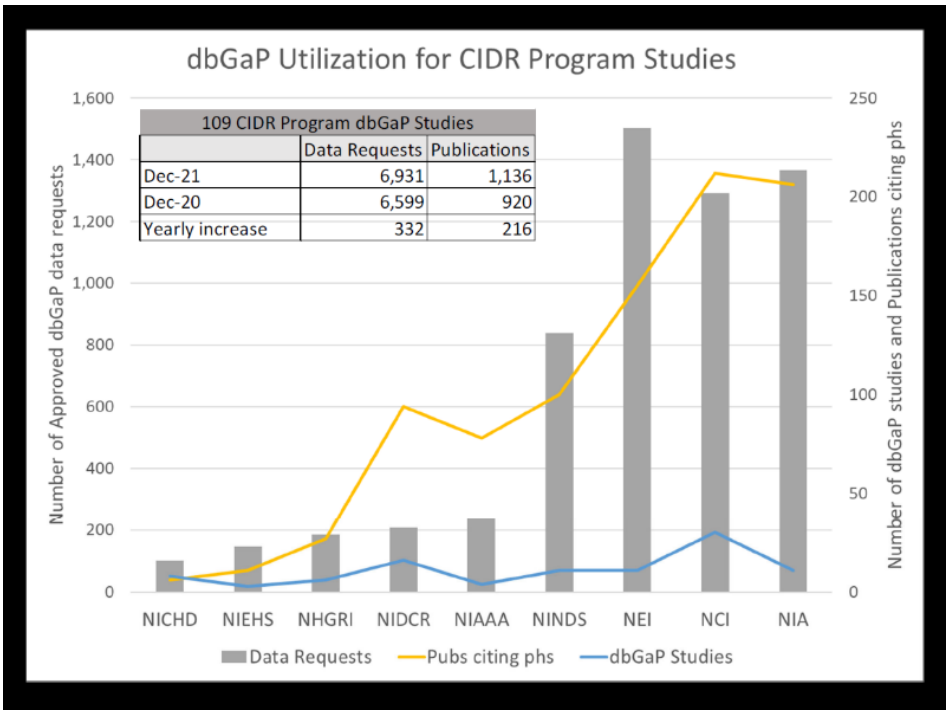
Epidemiology & Genomics Research Program: A Role Model for Data Sharing

Written by Sophia Coco

Responsible data management and sharing (DMS) are increasingly being recognized as vital for advancing scientific discovery. National Institutes of Health (NIH) has a long history of promoting data sharing dating back to 1999. Most recently, the NIH Final Data Management and Sharing Policy effective on January 25, 2023 aims to drive the culture towards broad data sharing in biomedical research by emphasizing how researchers should prospectively plan for maximized FAIR data management and sharing into their research process. It's anticipated that broad sharing will drive innovation in biomedical research while ensuring rigor and reproducibility.

At the National Cancer Institute (NCI), the Epidemiology and Genomics Research Program (EGRP) within the Division of Cancer Control and Population Sciences is one of the data sharing champion programs that has played a critical role in helping the implementation of the 2008 Policy for Sharing of Data Obtained in NIH Supported Genome-Wide Association Studies (GWAS), and later, the 2015 Genomic Data Sharing Policy. To achieve these goals, the program has consistently taken multi-pronged approaches by engaging relevant stakeholders, showcasing the impact of data sharing in cancer control research (for example, in the Up for a Challenge breast cancer genomics competition), and providing critical resources and education for the community. Two large programs from EGRP to highlight are the OncoArray from the GAME-ON Initiative and the Center for Inherited Diseases Research (CIDR) . Both OncoArray data and datasets from the CIDR resource (hosted at Johns Hopkins University) have been widely used by the research community, resulting in new insights and publication as shown in the figure below.

CIDR Data Usage: Data Access Requests and Publications



Furthermore, EGRP's dedication to data sharing is exemplified by the NCI Cohort Consortium, an extramural/intramural partnership to promote collaborations by pooling data and biospecimens collected from large, diverse populations. Funding opportunities for EGRP's cohorts also require robust data sharing beyond collaborations, to allow for a broader range of secondary research on cohort-generated data. EGRP utilizes the Cancer Epidemiology Descriptive Cohort Database (CEDCD) with information on the cohort studies to facilitate collaborations and data sharing.

To help implement the new DMS policy, the EGRP integrates resources and advice for researchers throughout their website. Future directions for the EGRP include developing a population science node of the NCI Cancer Research Data Commons, with the goal of facilitating FAIR data sharing in more areas of cancer research. As EGRP's Associate Director Dr. Kathy Helzlsouer explains, "It really takes a village to achieve the goals of sharing data in a FAIR way and make the fruits of our strong investment in cancer control research accessible to all to advance the health of the public. EGRP has been fortunate to have hard working data sharing champions, such as Drs. Elizabeth Gillanders, Charlisse Caga-Anan, and Pothur Srinivas, along with multiple members of our data sharing committee. These efforts will truly help provide new insights to improve cancer control."

EGRP Resources:

- Cohort Database: https://cedcd.nci.nih.gov/home
- Biospecimen sharing: https://epi.grants.cancer.gov/biospecimens/
- A list of GAME-ON and OncoArray datasets: https://epi.grants.cancer.gov/gameon/#datasharing
- NCI Cohort Consortium: https://epi.grants.cancer.gov/cohort-consortium/