Division of Cancer Epidemiology and Genetics
Data Sharing Policy

11/21/2012

I. Background

As part of the Division of Cancer Epidemiology and Genetics' mission to discover the causes of cancer and the means of its prevention, DCEG is committed to sharing research data to further advance science, improve public health, and leverage the investment made by the U.S. taxpayers in the Division's research program.

Data sharing must be conducted in a manner consistent with federal, state, and local laws and policies concerning privacy, confidentiality, protection from discrimination, the honoring of informed consent, as well as other restrictions related to institutional review board (IRB) approvals. For some studies, there may be data that cannot be shared because of conditions placed upon DCEG by the individuals or organizations who supplied the data to DCEG for the original research.

II. General Considerations

A. Protecting Privacy and Confidentiality

DCEG has the highest commitment to protecting the identity of study participants against undesired intrusions (privacy) and to limiting access to study information that might individually identify them (confidentiality). All direct identifiers are omitted from datasets; however, because study data often include extremely detailed personal information (*e.g.,* age, gender, race, residence, employment histories, medical histories), it may be possible in rare cases through complex analysis and with outside information, to identify specific subjects. Therefore, DCEG will redact data as necessary to protect study subject privacy, particularly for data requested under the Freedom of Information Act (FOIA) mechanism (see Section III.D). Less data redaction may be necessary under Data Transfer Agreements (DTA) that contain provisions assuring that the researchers will not try to learn the identity of subjects or link the shared data with individually identifiable records from other sources (see Section III.C).

B. Timing of Data Sharing

While DCEG is committed to data sharing, we believe the optimal time to provide data is after the publication of findings from a final data set. The reasons for this are related to NIH policy and data quality. The NIH data sharing policy allows the initial investigator to have first and continuing use of data that he or she has generated, although it does not permit prolonged exclusive use. Having the opportunity to publish study findings first acknowledges the time and effort, typically involving years, which have been invested by researchers in the design and conduct of their studies. It allows them to complete their work and to be appropriately recognized for their efforts. Furthermore, prior to publication, study data sets cannot be considered "final." Throughout data collection and analysis, various actions are undertaken to

detect out-of-range, illogical, and inconsistent data, to check original sources, and to correct data as needed. Because this process can continue up to publication, obtaining data prior to that time is not advisable.

C. Assistance from DCEG Investigators

In response to outside requests, DCEG will provide the necessary documentation (*e.g.*, questionnaires, coding manual, data dictionaries) to understand and analyze study data. Further assistance from DCEG investigators, however, may not be possible due to time constraints and the competing demands of their positions. Researchers requesting data through mechanisms other than a mutually–agreed upon scientific collaboration with a DCEG investigator should not expect assistance beyond the provision of standard study documentation.

D. Cost

Due to fiscal constraints, DCEG reserves the right to require payments to cover the cost of producing data sets and accompanying documentation.

III. Methods to Access Data

There are various methods to access data, depending on whether there are established study-specific application and review procedures, whether the requestor prefers to collaborate with DCEG investigators or to work independently, and whether special precautions are needed to protect the privacy of study subjects or the confidentiality of study data.

A. Study-specific Application and Review Procedures

Cohort Studies: Several of DCEG's large prospective cohort studies have established procedures whereby researchers can request data. Some have regularly scheduled calls for proposals, while others accept proposals on an ongoing basis. Typically, researchers must submit proposals detailing their research plans. Information on the following studies can be found at these websites:

Agricultural Health Study: http://aghealth.nci.nih.gov
Alpha-Tocopherol, Beta-Carotene Study: http://atbcstudy.cancer.gov
NIH-AARP Diet and Health Study: http://dietandhealth.cancer.gov
PLCO Study: http://prevention.cancer.gov/plco/eems
Radiologic Technologists Study: http://radtechstudy.nci.nih.gov

Genomic Data: Genomic data generated by DCEG are archived in the National Center for Biotechnology Information's Database of Genotypes and Phenotypes (dbGaP). Researchers may obtain dbGaP data only with the permission of a Data Access Committee (DAC), which operates according to the policies established by the NIH Scientific Oversight Committee. The requestor must stipulate in a Data Use Certification that they will comply with federal, state, and local policy; will only use the data for the specified research use; will not identify study participants; will not transfer the data; and other provisions. An institute official must vouch that the

requestor is a bona fide researcher, and must also stipulate to the provisions related to privacy and use of the data. Additionally, an IT Director at the requesting institution must be identified to vouch that the data are being stored and maintained in a secure fashion. The federal Genetic Information Nondiscrimination Act, which makes it illegal for health insurance companies, group health plans, and most employers to discriminate based on genetic information, is also relevant to these requests.

Information on submitting a Data Access Request for data stored in dbGap is available at: https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login . *(Will add the CGEMS-specific site when new DCEG website is launched. Currently, it links to dbGaP site.)*

B. Collaboration with DCEG Investigators

Many researchers interested in accessing data from DCEG studies prefer to do so by establishing a scientific collaboration with DCEG investigators. Through these formal collaborations, the DCEG team members can share their expertise and knowledge of the intricacies of the study data, which greatly enhances the research endeavor. A collaborative agreement is developed which details the data to be shared, the roles and responsibilities of the parties involved, and consistency of the proposed research with the study informed consent and IRB approvals.

Researchers should contact the appropriate DCEG investigator directly (http://dceg.cancer.gov/about/directory) to determine if a mutually agreeable scientific collaboration can be established.

C. Data Transfer Agreement

DCEG data for collaborative and non-collaborative research will be released under the terms of a Data Transfer Agreement, which must be signed by the DCEG Director, the recipient scientist, and an official representing the recipient scientist's institution. The agreement will specify the proposed research plan, identify who will have access to the data, and contain provisions related to protecting privacy. These provisions will require that: 1) the requesting researcher will neither attempt to link nor permit others to link the dataset with individually-identifiable records from any other dataset, and 2) the researcher will not attempt to use the dataset or permit others to use it to learn the identify of any person. In addition, the relevant NIH IRB must provide an assurance that the proposed research and data transfer are consistent with informed consent. If the agreement is for non-collaborative research, DCEG investigators will not be able to assist the requestor beyond provision of standard study documentation.

Researchers should contact the appropriate DCEG investigator directly (http://dceg.cancer.gov/about/directory) to request data and develop a Data Transfer Agreement.

D. Freedom of Information Act Request (FOIA)

FOIA requires federal agencies to provide access to agency records except to the extent that such records are protected from public disclosure by one of nine exemptions or three law enforcement

exclusions. The exemptions that most often apply to DCEG records are those that prohibit disclosure of records that are: 1) specifically exempted by another federal law (*e.g.*, the Privacy Act of 1974) (Exemption 3); 2) privileged inter-agency or intra-agency communications that are pre-decisional and part of an agency's deliberative process (Exemption 5); and 3) personnel, medical, or similar files, release of which would constitute a clearly unwarranted invasion of personal privacy (Exemption 6). Requestors should note that because there are no restrictions placed on use or further dissemination of records obtained through a FOIA request, DCEG takes special care to protect subjects from invasion of privacy that might result from linking the dataset to individually identifiable records from any other dataset. FOIA allows the agency to recover part of the cost of responding to a request. DCEG investigators will not be able to assist the requestor to use or interpret the data beyond provision of standard study documentation. Details on FOIA are available at http://www.nih.gov/icd/od/foia/index.htm.

FOIA requests for DCEG records should be submitted to the National Cancer Institute FOIA Coordinator, Building 31, Room 10A48, 9000 Rockville Pike, Bethesda, MD 20892.