

## POLICY ON DATA MANAGEMENT, PRESERVATION AND ACCESS

### OVERVIEW

#### Scientific Data Management Plan (SDMP)

HEI-Energy is committed to transparency and data access. It is HEI-Energy's policy to provide access expeditiously to data for studies that it has funded and to provide those data in a manner that facilitates review and verification of the work but also protects the confidentiality of volunteers who may have participated in the study and contributors of third-party data, while respecting the intellectual interests of the original investigator of the work.

The majority of HEI-Energy's work comprises research, sponsored under contracts with HEI-Energy, and based in a variety of disciplines – exposure science, epidemiology, biostatistics, toxicology and chemistry; therefore, different types of data are expected to be generated depending on the nature of the study. HEI-Energy will take several approaches to ensure data integrity, preservation and access:

- Research solicitations issued by HEI-Energy include requirements for data access following completion of the study (*see Attachment*). Research solicitations include details of the investigator's and their university's obligations in this regard. Obligations for data preservation, access and storage are stipulated under terms of the contract under which HEI-Energy engages the investigator;
- Various quality assurance/quality control aspects of data generation, management, and storage during and after completion of research are also detailed in HEI-Energy's Quality Management Plan. Each study's QA plan is reviewed by the HEI-Energy Research Committee and staff. HEI-Energy also obtains the services of outside expert QA auditors. Any deficiencies in data management plans are communicated to HEI-Energy and to the investigators, and HEI-Energy ensures that corrective actions are taken.

The major responsibility for data management is delegated by HEI-Energy to the investigator. The location where data are stored and made accessible also varies, depending on the kind of data. If the investigator does not have a preference, HEI-Energy will use the Harvard Dataverse repository, which is an open source web application, designed to share, preserve, cite, explore, and analyze research data, and has more than 90,000 datasets. Dataverse stipulates standards and best practices for data use and metadata content that are consistent with EPA's requirements (see <http://best-practices.dataverse.org/>). Various levels of security and access can also be implemented, but HEI-Energy's goal is to make data access as simple as possible, while respecting the rights of the investigators who generated the data and the confidentiality of those who provide data or participate in the study.

For studies involving human subjects, specific human subject identifying information is stripped from the stored data. However, when using third party private data where data sharing is prohibited as part of the condition for the initial access or using public data that can be easily obtained by anyone, HEI-Energy requires the investigators to provide detailed instructions on how the data may be obtained by a requestor.

HEI-Energy requires that data be saved in a usable form for at least 10 years and that the investigator inform HEI-Energy when disposing of the data. HEI-Energy has provided supplemental funding to investigators to prepare and post their data. HEI-Energy is willing to and has paid for long term preservation of data and samples. Reasonable costs associated with transfer of data may be billed to the requestor.

In general, access to data, metadata and code are necessary for replication of initial findings; however, this is an evolving area and practices in the research community are rapidly changing. Additionally, other

considerations – such as access to suitable computing facilities, and familiarity with advanced analytical methods – may also be needed for replication.

One of the advantages of using a public repository, such as the Dataverse, is that once posted, the data and code are preserved and can be obtained at any time. Change in personnel can be challenging, but with good practices in preparation of meta data and inclusion of the code – in addition to well-prepared data files– mitigate the impact of change in personnel.

HEI-Energy allows reasonable costs of items such as rentals, leases, and computer costs or other equipment to be included in applicants' budget.

A small part of HEI-Energy activities comprises review of literature or other existing studies; as such, these reviews do not generate scientific data. In the interest of full transparency, HEI-Energy publishes the full protocol for such review activities and the reports produced by such expert panels are also peer-reviewed.

## **ATTACHMENT: HEI-ENERGY POLICY STATEMENT ON THE PROVISION OF ACCESS TO DATA UNDERLYING HEI-ENERGY FUNDED STUDIES**

The provision of access to data underlying studies funded by HEI-Energy is an important element of ensuring credibility, especially when the studies are used in controversial public policy debates. The open and free exchange of data is also an essential part of the scientific process. Therefore, it is the policy of the HEI-Energy to provide access expeditiously to data for studies that it has funded and to provide that data in a manner that facilitates review and verification of the work but also protects the confidentiality of any volunteers who may have participated in the study and respects the intellectual interests of the original investigator of the work.

This policy applies to all research funded by HEI-Energy; it is consistent with federal policy, as outlined in 2 CFR Part 200 - Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards, which requires access under the federal Freedom of Information Act (FOIA) to data from federally- supported research that was used in developing a federal agency action that has the force and effect of law.

In responding to FOIA requests through the U.S. EPA or other federal agency for HEI-Energy data that are subject to 2 CFR 200 requirements, HEI-Energy will follow the principles established in the amendments.

In responding to non-FOIA, direct requests to HEI-Energy for data, HEI-Energy will in general follow the principles described below, which are designed to be consistent with the principles contained in the recent 2 CFR 200 requirements, although specific cases may require other arrangements for providing access.

1. *Research Data.* The research data that will be made available in response to requests will vary from study to study, but in general will consist of the recorded factual material commonly accepted in the scientific community as necessary to replicate and verify the original research findings. It will include digital records including analytical summary, metadata and computer codes, where appropriate, but will not include any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. The “recorded” material excludes physical objects (e.g., laboratory samples). Research data also excludes (a) trade secrets, commercial information, materials necessary to be held confidential by a researcher until published, or similar information which is protected under law; and (b) personal and medical information and similar information that is personally identifiable, and the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.
2. *Data Plan and Provision of Access to Research Data.* HEI-Energy will expect each Principal Investigator (PI) it funds to provide, at the outset of the study:
  - (a) a plan for organizing, protecting, archiving and making all data, data descriptions, analytical summary, metadata and computer codes described above available to HEI-Energy upon completion of the study,
  - (b) a plan to place the data, metadata, codes and other relevant information concerning the data and analyses, at a well maintained public repository; in cases where storage at such a repository is not possible, a clear plan for preservation and access to data from another platform should be described, and
  - (c) a plan for making the data available to other investigators following publication of the results as set forth under this policy.
3. *Data from Publicly Available Sources.* In cases where all of the data used are from publicly available data sets (e.g., air quality data from the EPA) and the analytic data set can readily

and expeditiously be recreated, HEI-Energy and/or the PI may, as an alternative, provide detailed descriptions of how to access and use these public data sets to recreate the analytic data set in lieu of providing the full analytic data set.

4. *Third Party Data.* In cases where the PI uses data for HEI-Energy sponsored research collected by a third party – whether public or private – and the PI is contractually bound with the third party to keep the data confidential, HEI-Energy and/or the PI will provide information on the process the third party has in place for access to the data and will direct the requestor to the third party to seek access to the data; wherever possible, HEI-Energy will facilitate this process. HEI-Energy will provide access to third party data only when such access is consistent with the confidentiality or other obligations HEI-Energy or its PIs have with respect to such data.
5. *Timing.* HEI-Energy will seek to provide access to data as expeditiously as possible after the completion, review and publication of the HEI-Energy Research Report (or Reports) resulting from the study. In doing so, HEI-Energy will, to the maximum practical extent, take into consideration the legitimate intellectual interests of the PI to have the opportunity to benefit from his or her intellectual endeavors and to publish subsequent analyses from the data set (including additional analyses funded by HEI-Energy). In some cases, e.g. for studies of particularly high regulatory importance being used to inform decisions over a short time frame, HEI-Energy may need to work to balance the PI’s interests against the need for interested parties to obtain access in a timely manner.
6. *Length of Data Retention.* PIs funded by HEI-Energy are required to retain all data generated in the course of HEI-Energy-funded research for at least ten (10) years from the date of publication of the research by HEI-Energy, or a longer period if required by a funding agency or third party data provider or as directed by HEI-Energy. HEI-Energy retains the right to access the data at any time during this period. If the PI has kept the data beyond this time, HEI-Energy will continue to have the rights to access to the data. At any point, and at least ninety (90) days prior to any alteration or destruction or other disposal of the data, the PI will notify HEI-Energy so as to enable the Institute to request such data under this provision.
7. *Responsibility and Reimbursement for Costs.* To the maximum extent possible, HEI-Energy will encourage the PI to be the primary sharer of the data. To the extent that providing the data would place an undue burden on the PI (e.g., in a situation where the sheer number of requests would not allow the PI to continue to conduct research or academic activities), HEI-Energy will be prepared to establish an alternative procedure for sharing the data. In either case, HEI-Energy will expect to receive from the entity requesting the data reasonable reimbursement for both the direct costs of providing the data, and PI and/or HEI-Energy staff time to gather, transmit, and explicate the data. To facilitate access to data from future and current studies in which HEI-Energy and the PI expect that the results have a high likelihood of being used in supporting a regulatory decision, HEI-Energy will consider requests from the PI for a reasonable budget of data archiving funds, to be provided as part of the project budget.
8. *Confidentiality.* Any requester of research data will be expected to obtain approvals and enter into the required data use agreements necessary to permit the requester access to such data. The requestor will be fully responsible for adhering to all such approvals from the appropriate agencies (e.g. the National Center for Health Statistics) or other third-party data providers. HEI-Energy will not knowingly itself provide, or require a PI to provide, information that can be used to identify a specific individual without the requester having already obtained all such necessary approvals.
9. *Responsibility of the Data Requester.* In addition to the payment of reasonable costs and the obtaining of any necessary confidentiality approvals, HEI-Energy will ask the data

requester, as would be normal courtesy in the scientific community, to inform both the PI and HEI-Energy promptly of any findings emerging from their analysis, to provide the PI an opportunity to respond to those findings prior to publication, to provide copies to both the PI and HEI-Energy of any papers submitted for publication from the data, and to cite both HEI-Energy and the PI in any publication, noting explicitly that the views expressed are those of the new analyst and not those of the PI, HEI-Energy, or HEI-Energy's sponsors.

10. *HEI-Energy Decision Making.* All requests for research data will be reviewed and decided upon by a Committee of the HEI-Energy Science Director, and the Chairs of the HEI-Energy Research and Review Committees, in consultation with both the research and review staff scientists responsible for the study in question. Any significant policy questions arising from a particular request will be considered, upon recommendation of the Committee and the President, by the Board of Directors.

The provision of data will not be simple to accomplish and will at times raise concerns and controversy from one or more parties; additionally, the best practices within the scientific community are rapidly evolving. HEI-Energy will attempt to provide data in a manner that, to the maximum extent practical, fosters an atmosphere of collegiality and mutual respect among all parties, with the aim of obtaining from the sharing of data the maximum benefit for science and for the quality of the public policy decision-making process.