

---

## Chapter 5: Confidentiality and Researcher Access

When producing the data files, grantees must decide how much data can be released to researchers without disclosing individually identifying information, including potentially sensitive information about welfare receipt, family income, and family well-being. The Work Group did not develop specific guidance relating to confidentiality concerns, which often are specific to the state laws, interagency agreements and data sets of each grantee. Concurrent with the Work Group deliberations, however, ASPE staff worked with grantees and with ORC Macro staff to develop an overall approach to researcher access that addresses many confidentiality concerns. This approach and related issues of ongoing file maintenance are described below.

### 1. Researcher Access

The **level of protection within the data file depends heavily on the issue of access.** If data sets are at a secure site with restricted access and signed agreements are in place that protect confidentiality, there is little need to strip data of potential identifiers, such as geography, etc. If the data are on a CD-ROM and distributed to anyone who signs a data-sharing agreement, the work of stripping and masking personal identities becomes much more difficult and depends heavily on population and sub-population size. ASPE recommends that grantees adopt the former approach, which is to **place the Welfare Outcomes data files in a secure site with restricted researcher access.** However, grantees also have the option of producing “public use” data sets that have been stripped of identifying information and would be made available with fewer restrictions on researcher access.

#### Restricted-Access Data Files

ASPE has entered into an interagency agreement with the **National Center for Health Statistics (NCHS)** to facilitate placement of Welfare Outcomes files at the **NCHS Research Data Center.** To gain access to the data, a researcher will have to apply for the data, sign an agreement not to disclose individual identities, and either come to the Research Data Center in Hyattsville, Maryland, or use a remote access (electronic) system. Research use of the data will be monitored to prevent breaches in confidentiality and inappropriate disclosure of data. For example, the NCHS Data Center does not allow researchers to print out individual observations or cells with fewer than five observations.



Chapter 1 and Tip 1.2 show how to replace Case ID numbers with unique identifiers.

Grantees planning to place their files at the NCHS Research Data Center are expected to **strip names, addresses, social security numbers, and Case ID numbers from the files** (replacing the Case ID numbers with unique identifiers as described in Chapter 1 and Tip 1.2). Most of the rest of survey data files, and large elements of the administrative data files, should be retained in the data file, following the production and documentation procedures outlined in the first four chapters of this guide. Some data might not be able to be released, even through the NCHS data center (e.g., extremely sensitive data such as child welfare administrative records or other records that cannot be released under state laws or state interagency data sharing agreements). In general, however, the restricted-access data files can be quite rich in detail, providing more information (e.g., county of residence) than would be retained on the files if they were released more publicly.

### Public-Use Data Files

A few grantees have expressed a **preference for retaining control of the release of their data**, and for taking on the task of producing public use data files by stripping data sets of geographic identifiers and other elements that could be used to disclose individuals. Copies of the resulting public use data files will be sent directly to researchers who apply for access and sign agreements to respect the confidentiality of the data. ASPE proposes that the **NCHS Data Center be used as a central point of access**, through which researchers may request copies of the data sets. In addition, grantees may also make the data available themselves.

**Grantees planning to produce public-use data files are responsible for protecting the confidentiality of clients and respondents** in their administrative and survey data files. This is likely to involve further manipulation of the data sets beyond what has been outlined in the first four chapters of this guide, though the recommendations of the Work Group are still generally applicable. Some data elements will probably have to be collapsed further, to avoid identifying of individuals (for example, the aggregation of counties into state regions and the aggregation of individual ages into age groups). Extensive literature exists on methods to prevent disclosure of individual identities; and a bibliography of resources related to confidentiality and preventing disclosure of individual identities is provided in Appendix C.



See Appendix C for resources related to protecting identities and confidentiality.

---

## 2. Ongoing Support for Data Files

Once the grantees have produced data files, documented them, and determined the level of researcher access, they can turn the file over to ASPE. **ASPE is willing to take primary responsibility for ongoing support for the data files.** This final section discusses the submission of files to ASPE, procedures for handling future technical questions, and researcher fees to support ongoing maintenance of the files.

### Submitting Files to ASPE

When grantees submit the data files to ASPE, they should provide a written statement specifying whether their files are to **be restricted-access data files or public-use data files.** Grantees should also indicate their **desired level of future involvement** with applications for use of the file, which can range from review of applications to no involvement. Finally, grantees should provide **contact names of individuals** who can assist ASPE and NCHS staff in responding to technical questions about the administrative and survey data files (see Tip 5.1).



Tip 5.1 outlines issues that grantees should address when submitting their data files to ASPE.

---

## ✓ **Tip 5.1: Submitting Files to ASPE**

Grantees should address the following areas in a cover letter accompanying the files:

### **Restricted Access or Public-Use Files**

- Are the files to be kept in a secure site with researcher-restricted access? Or are they to be made available as a public-use file?
- Grantees may choose different levels of access for survey and administrative data files.

### **Desired Level of Future Grantee Involvement**

- Do grantee personnel want to review applications for access to data, receive notification of such applications, or have no further involvement?
- Include contact information (name, address, e-mail, telephone number) of person(s) to be contacted regarding applications for access to data.

### **Contact for Technical Questions**

- Provide name and e-mail address of person(s) who can respond to technical questions about the administrative and survey data.
- Researchers will not contact grantee personnel directly, but will send e-mail to NCHS and ASPE staff, who will screen questions before forwarding appropriate questions to grantees.

### **Special Conditions for Use**

- Any additional conditions of use?
- For example, do grantees have data-sharing agreements that must first be signed by researchers, in addition to agreements developed by NCHS and ASPE?

Grantees are encouraged to discuss their proposed plans with ASPE staff prior to submitting the data files. Grantee should also talk with ASPE staff if they have questions or concerns about access to data files.

---

## Responses to Technical Questions

Under the agreement with the NCHS Research Data Center, a **centralized e-mail account** will be established through which researchers can pose questions that they may encounter while using the grantees' data files. Questions will be answered, to the extent possible, by NCHS staff, who will consult with ASPE staff as needed and refer researchers to grantee documentation as appropriate. Questions that cannot be answered by NCHS, ASPE, or ASPE's technical assistance contractor will be forwarded to the listed grantee contact(s) on technical questions.

## Researcher Fees

To support the ongoing maintenance of the restricted access files at the Data Center, NCHS charges modest a **researcher access fee** to researchers who use the files. Currently, the NCHS Data Center charges \$1,000 per week for on-site access, or \$500-\$1,000 per month for remote access, depending on the size of the files. Nominal fees, if any, are charged for distributing the public-use (general access) data files.

In order to alleviate burden to researchers and encourage access to the Welfare Outcomes data files, **ASPE plans to subsidize up to 75 percent** of the researcher fees, up to a set amount per project and a set amount of projects. Further information about how to apply for access to the Welfare Outcomes data files, including applications for subsidy of researcher access fees, will be posted on the ASPE-sponsored web page on Leavers and Diversion studies <<http://aspe.hhs.gov/hsp/leavers99/index.htm>> as it becomes available.