

3

Documentation of CEDR Data

To help clarify the organization of data in CEDR, several terms have been defined. However, due to the diverse nature of the data in CEDR, these definitions are fairly broad. Some of the most commonly used CEDR terms are defined below. Other definitions can be found in **Appendix F**.

Data Definitions and Relationships

Worker Health and Mortality Studies

A *data file set* is a collection of logically related files and is the term used to refer to collections of any type of CEDR data files. For the Worker Health and Mortality Studies, however, CEDR files are grouped into working data file sets and analytic data file sets.

A *working data file set* contains files of data -- called working files -- that were collected and updated during decades of follow-up of worker cohorts.

An *analytic data file set* contains files of data -- called analytic files -- on which the analyses or results of a study are based. Researchers select data from a number of potentially diverse working files to form their initial analytic data file set. **Figure 3-1** shows the relationship between working and analytic data file sets in CEDR.

Working files may contain an assortment of data, including demographic, work history, industrial hygiene, vital status, internal dosimetry, and external dosimetry data, gathered from a variety of sources. Working files are more dynamic, with new, updated, and corrected information being added to the files. The level of

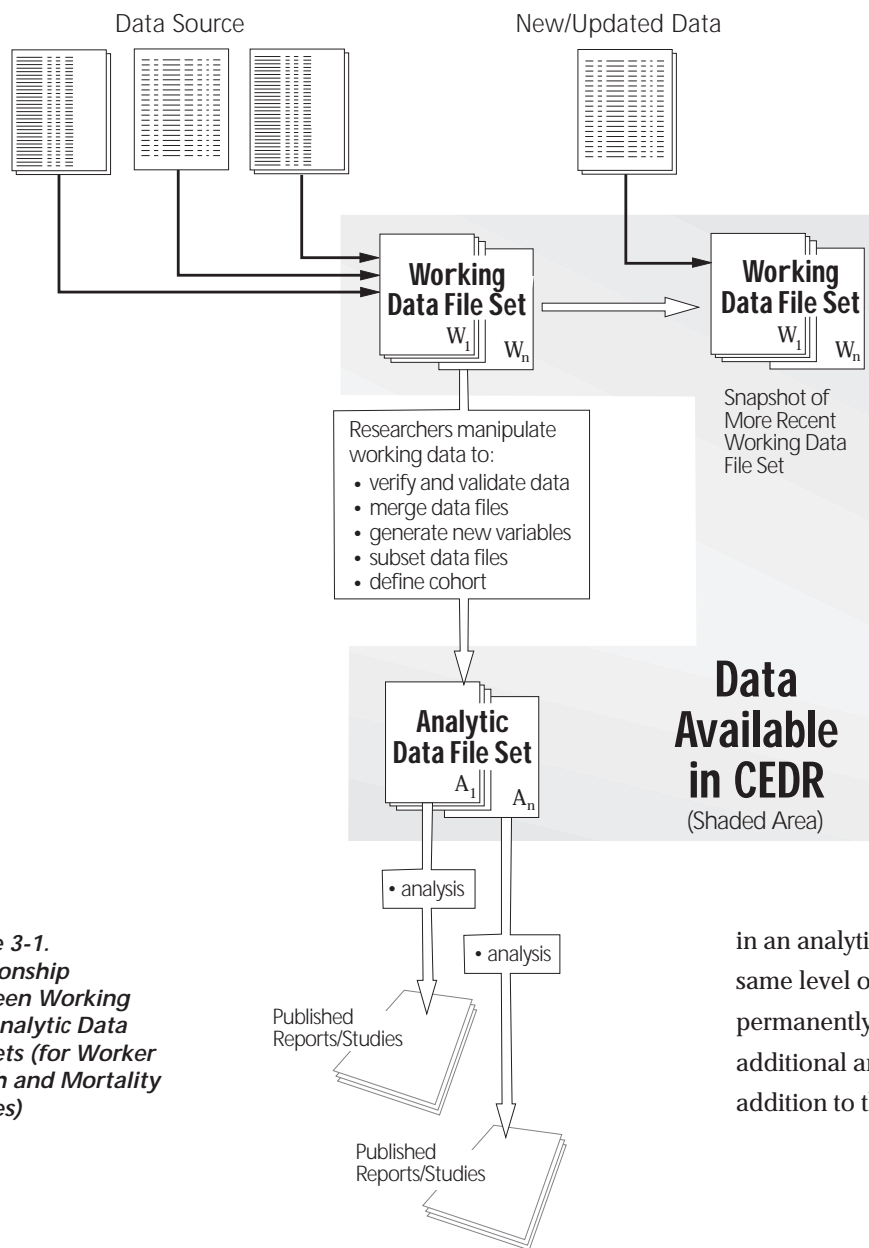


Figure 3-1.
Relationship between Working and Analytic Data File Sets (for Worker Health and Mortality Studies)

verification and validation applied to working data will vary by site and by study. For some studies, working files have been periodically updated in CEDR; older versions of these working files have been archived.

Analytic files contain data upon which a researcher more directly bases a study's conclusions or reported findings. Typically, analytic files are composites of data (derived from working files) formulated to meet specific cohort descriptions and study needs that represent the best data available at that time. To create analytic files, researchers may subset data to form a cohort, merge data from multiple working files, validate and edit the working data, and convert the data as necessary to conduct analyses. Analytic files generally represent data that have been extensively reviewed, validated, and verified by a researcher. However, the files may also contain additional data that the researcher collected but did not use in the final analysis or publications. If the study analyzed only a subset of individuals or variables contained

in an analytic file, not all variables or records in the file may have the same level of validation. Analytic files are static and are retained permanently in CEDR. If similar files with updated data are used for additional analyses, a new data file set is defined and stored in addition to the original analytic data file set.

Information about CEDR's Structured Documentation

- ❖ More information about the attributes of CEDR's structured documentation is contained in Guidelines for Data Documentation. Copies are available from the CEDR Program Manager or it can be accessed through the CEDR electronic information system.

Other Types of CEDR Holdings

Collections of electronic files pertaining to environmental dose reconstruction projects or other health-effects data constitute other types of data file sets that will be documented and permanently retained in CEDR. Some of these collections, such as the Radium Dial Painter Studies and the U.S. Transuranium Registry Data, contain data which are linked to particular individuals, and they are treated similar to the files relating to the Health and Mortality Studies.

Grouped data, which are not directly linked to particular individuals, appear in the data file set provided by the Radiation Effects Research Foundation. Dose reconstruction data, such as the data file set pertaining to the Nevada Test Site, consist primarily of environmental or other measurement data, which are also not linked to particular individuals.

Reports, such as epidemiologic surveillance reports and reports of record inventories, are static text files in CEDR that contain information also not directly linked to individuals.

Structured Documentation

CEDR also includes a large volume of *structured documentation* that describes the files and data in CEDR. This documentation provides an overview of the study and data file sets. More detailed information is also included in the structured documentation for each file and each variable. Taken as a whole, the structured documentation provides essentially the same information normally found in a researcher's code book. This detailed documentation supplied by each data provider maximizes the usefulness of CEDR data to researchers unfamiliar with the study.

How CEDR Data Are Collected

As researchers near completion of their study, they compile their electronic files, including working files, analytic files, and structured documentation, into data file sets that are submitted to CEDR. Although the data files may differ considerably in content and format, the accompanying structured documentation is processed into a standard format. This standard format ensures that the structured documentation explains the data as consistently and completely as possible.

Quality assurance procedures have been implemented throughout CEDR processing to ensure the data, as presented in CEDR, are consistent with that submitted by the data provider. Various checks

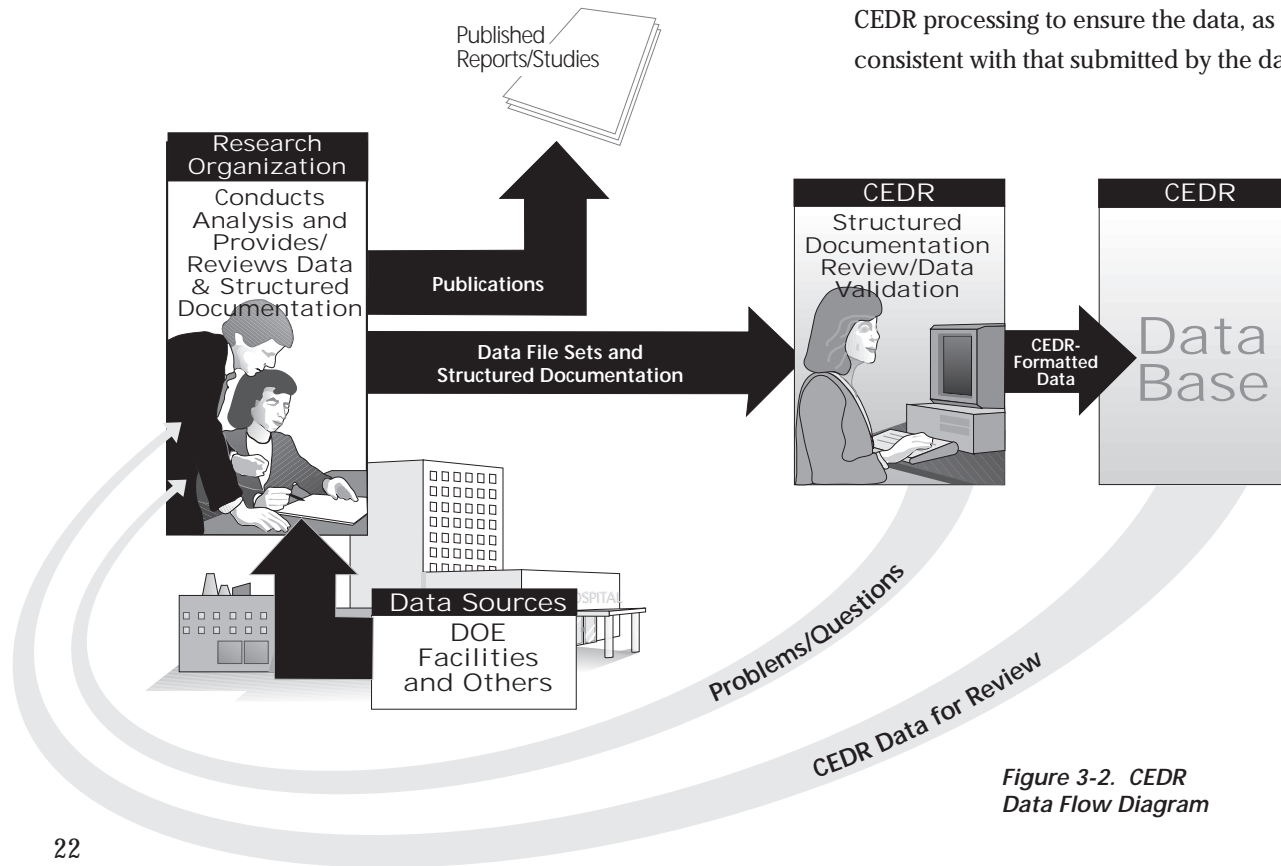


Figure 3-2. CEDR Data Flow Diagram

look for omissions and logic errors during data loading. In general, CEDR data are processed as shown in the data flow diagram provided in **Figure 3-2**. When CEDR data are received, the data files are examined and the structured documentation is reviewed for completeness. Data are not accepted until structured documentation requirements are satisfied. After acceptance, the data file sets and structured documentation are loaded into the CEDR system. The data and structured documentation are then reviewed by the researcher who provided the data for final verification and approval.

Considerations in Use of CEDR Data

Confidentiality Protection Summary

- ❖ CEDR's goal is to protect the privacy of individuals while maximizing the usefulness of the data.
- ❖ CEDR's release policy de-identifies the data for use by interested individuals.

Data Format

Because the information used by various researchers was originally generated from different DOE facilities or other sources for different purposes and needs, there is considerable variation in the format and content of data files. Terminology used to name or describe variables is also inconsistent. There has been no attempt to standardize individual variables (field names) found throughout the CEDR data. Each investigator defined these variables as he/she wished. Users should pay close attention to documentation provided for each variable. Particular questions can be addressed to CEDR support staff.

Confidentiality Protection

CEDR complies with requirements of the Privacy Act of 1974 and the agreements between DOE and state health departments. Therefore, to protect the identities of individual workers, the following data release policy was implemented:

- Individuals are identified only by a number. Other identifying information, such as names and social security numbers have been omitted.

- The specification of race is limited to white, black, and other.
- The dates of birth and death have been truncated to the year.
- Other personal dates, such as date of employment, are truncated to month and year.

Cause-of-Death Data

Most DOE epidemiologic studies examined the mortality rate of the study population and ascertained the cause of death of the deceased workers. The cause-of-death information was obtained from death certificates. Each state has its own requirements and rules for releasing death certificate information. Three states require CEDR users to complete application materials and have these reviewed before cause-of-death information for these states can be released. Users should also note that, in spite of repeated requests by DOE, New York City Department of Health still will not permit release of cause-of-death information for deaths occurring in New York City.

4

Access to CEDR Information

Becoming a CEDR User

To satisfy the needs of diverse user groups and to fulfill confidentiality agreements, two types, or levels, of access to CEDR resources have been established. These access levels are described below and summarized in **Figure 4-1**.

General Access - Allows access to files of CEDR structured documentation, information about the CEDR Program, and various reports. These kinds of information are freely available for on-line access and browsing, and no registration or authorization is required. General Access also allows access to CEDR holdings that are not directly linked to an individual, such as grouped or dose reconstruction data, which do not require confidentiality protection.

Please sign our Guestbook. On-line registration is requested of those using CEDR. This registration is easily accomplished through CEDR's Guestbook. Please see "Sign our Guestbook" (<http://cedr.lbl.gov>).

Authorized Access - Allows access to data at the individual level, such as that included in the health and mortality data file sets (working or analytic) which require confidentiality protection. These CEDR users must complete the DOE confidentiality forms, as well as the forms required by the states of Alabama, Pennsylvania, and Rhode Island, found in **Appendix D** of this catalog and adhere to the terms indicated. Review of forms by these states usually takes about 3 weeks to complete.

Access Level	Information Available
General (No registration or authorization required. Please sign the CEDR Guestbook online.)	<ul style="list-style-type: none">• Structured Documentation• Reports• Grouped Data• Dose Reconstruction Data
Authorized (Confidentiality forms and registration online required.)	<ul style="list-style-type: none">• Structured Documentation• Reports• Grouped Data• Dose Reconstruction Data• Individual Data

Figure 4-1. Summary of Information Available for each of CEDR's Levels of Access

Methods of Accessing CEDR Data

Obtaining printed CEDR Documents

Interested persons may request printed CEDR materials by contacting the CEDR Program Manager or CEDR staff at Lawrence Berkeley National Laboratory (LBNL) (see Section 1 for addresses and phone numbers), or via the CEDR web site .

CEDRView: Digital Access to CEDR resources

The Internet-based World Wide Web is widely used to organize, search, and distribute documents electronically. CEDR provides access to its resources via World Wide Web, including:

- Browsing Structured Documentation
- Viewing Data Files
- Keyword searching Structured Documentation
- Downloading Structured Documentation and Data Files
- Viewing this CEDR Catalog online
- Downloading this CEDR Catalog to a printer
- Ordering a printed CEDR Catalog to be mailed
- Obtaining bibliographic information
- Signing the CEDR Guestbook
- Requesting to become an authorized CEDR User
- Sending comments about the current CEDR system (these comments will assist DOE in the evaluation and refinement of the CEDR information system)

This collection of services is called CEDRView.

A Brief Guide to Browsing, Searching, and Downloading in CEDR

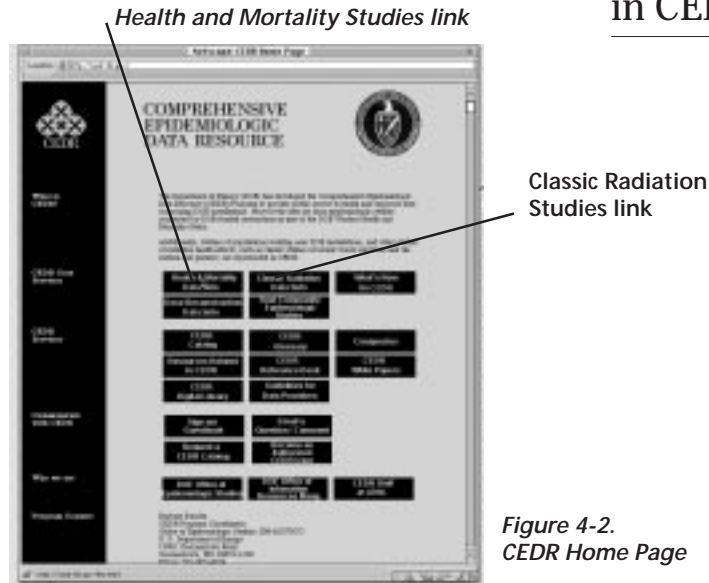


Figure 4-2.
CEDR Home Page

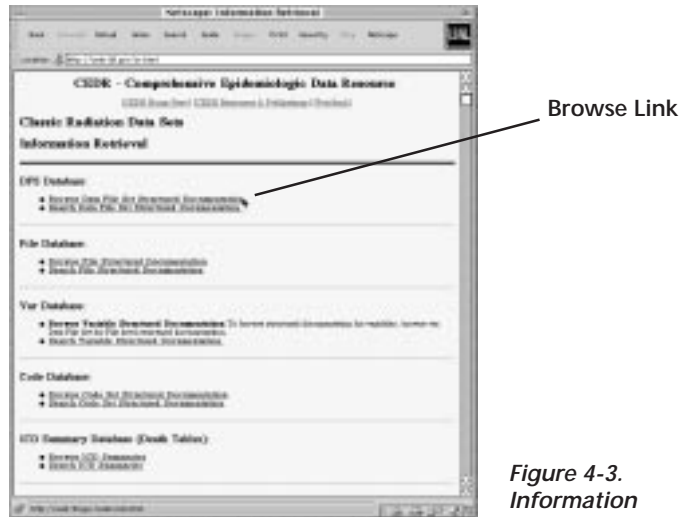


Figure 4-3.
Information
Retrieval Page

Structured Documentation

Structured Documentation is browsable (i.e., a complete ordered listing is presented) and keyword-searchable. Documentation describing CEDR Data File Sets is structured hierarchically, by Data File Set-level (overview), file-level, variable-level and code set-levels. For example, variable-level structured documentation is accessible from file-level documentation, which is in turn accessible from the Data File Set level.

To browse through these levels of CEDR Structured Documentation:

- Point your browser to the CEDR web site (<http://cedr.lbl.gov>)
- Select either Health and Mortality Data Sets or Classic Radiation Data Sets (figure 4-2)
- On the Information Retrieval page (figure 4-3), select Browse or Search according to the Data File Set of interest.



Data File Set link

Figure 4-4.
Browse Page

- If browsing was selected above, an ordered listing of Data File Sets will be displayed (see figure 4-4) . Select a link to display Data File Set-level Structured Documentation.



"See Files" link

Figure 4-5.
Data Files Set level
Page

- In the Data File Set-level display (figure 4-5), activate the link "See files associated with *Data File Set* name".

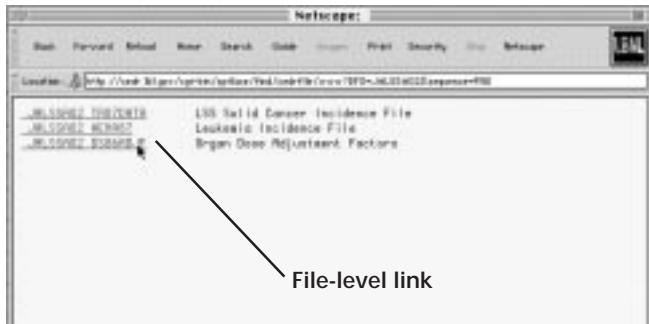


Figure 4-6.
Browse Page

- A list of files for the selected Data File Set is displayed (figure 4-6). Activate the link to display the file-level Structured Documentation for the desired data file.

- A file-level display is shown in figure 4-7.

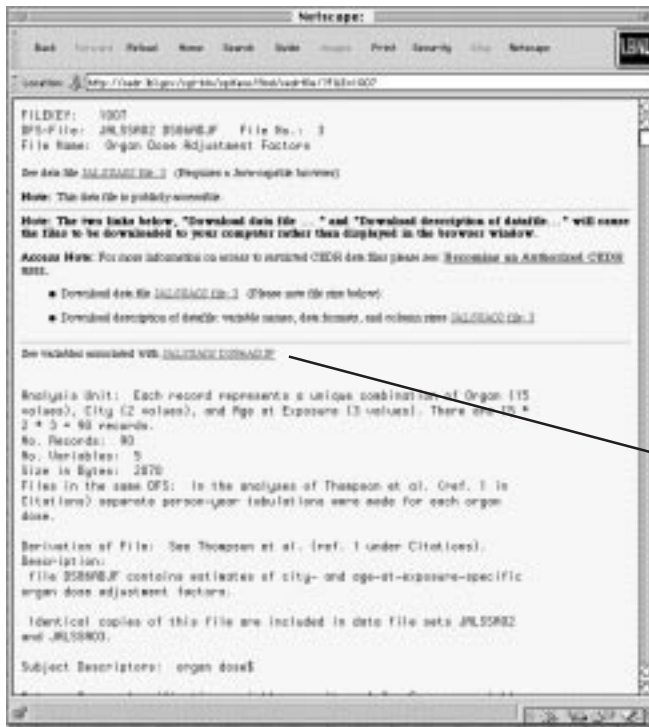
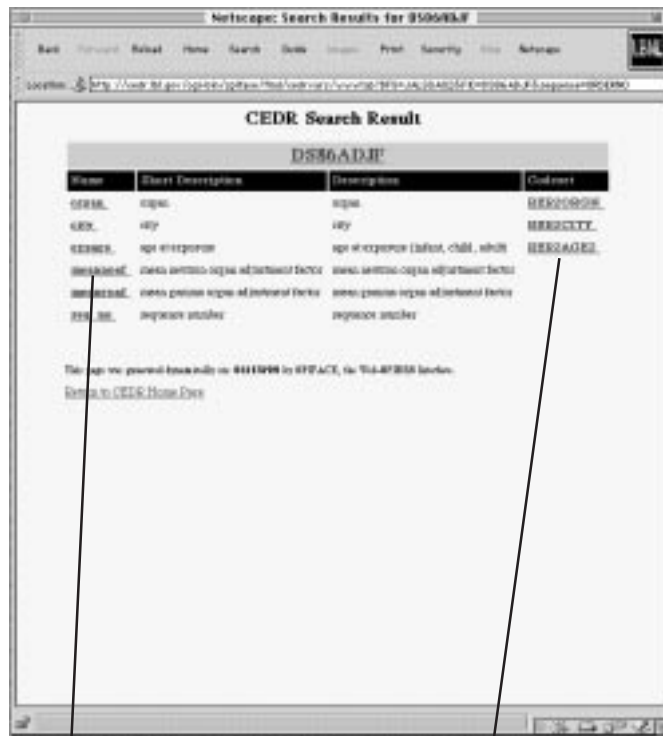


Figure 4-7.
File-level Structured
Documentation Page

From a file-level display, one can display more detailed variable-level Structured Documentation and view the corresponding Data File (see CEDR Data Files below).

- To display a list of variables in the file, activate the link “See variables associated with *Data File Set Filename*”.

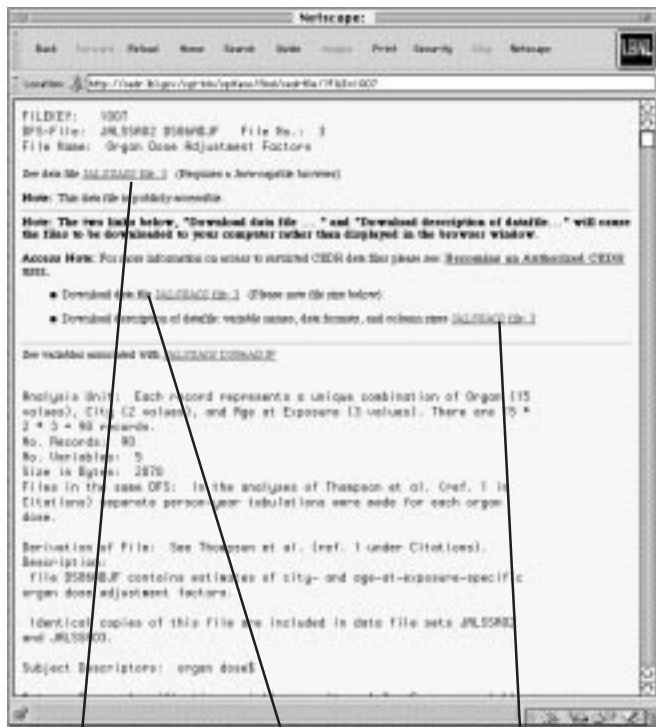


Detail about a variable

Code Set detail

- The variable listing will display a table with variable names, descriptions, and links to code sets (figure 4-8).
- For further information about a specific variable activate the variable name link; or for a display of the code set values and their interpretations, activate the hyperlinked codeset name.

Figure 4-8. Variable listing



View Data via
Java applet

Download
Data File

Download Data File
characteristics file

Figure 4-9.
File-level Structured Documentation

CEDR Data Files

CEDR Data Files are accessible via online Structured Documentation at the file-level, (please refer to figure 4-7) in two ways:

- View Data Files

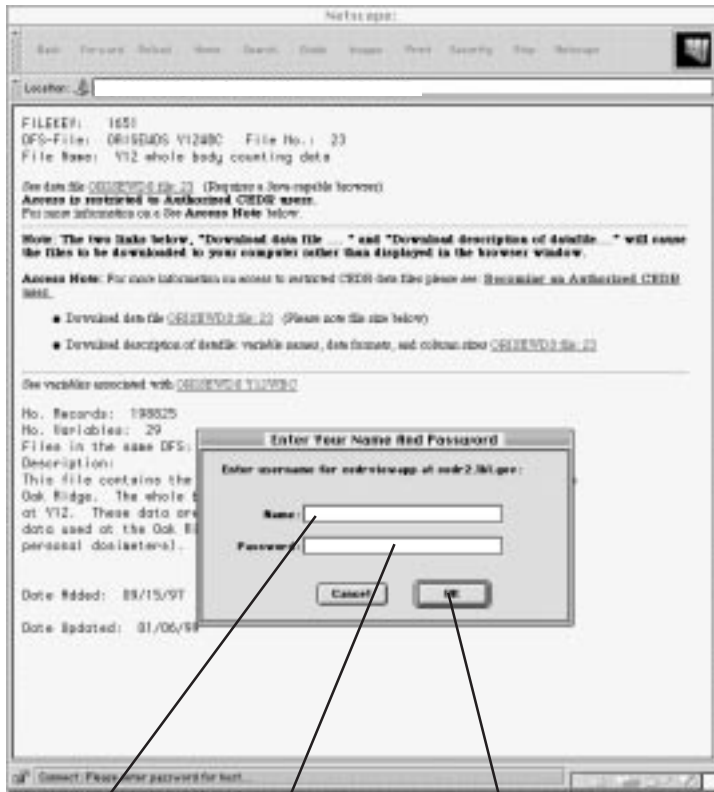
From the file level record, activate the link “See data file *File Identifier*” (figure 4-9). This Java applet method facilitates browsing the Data File (figure 4-11) without downloading (most files are more than one megabyte).

- Download Data Files

From the file level record, activate the link “Download the data file *File Identifier*” (figure 4-9). When downloading a data file, users should also download the description of Data File characteristics by activating the adjacent link “Download description of datafile: variable names, data formats, and column sizes”.

Figure 4-10 illustrates a dialog box which permits CEDR Authorized Users to view or download CEDR Data Files from Health and Mortality Data Files Sets.

In all cases, assistance from CEDR project staff is available at LBNL (address and phone numbers are provided in Section 1).

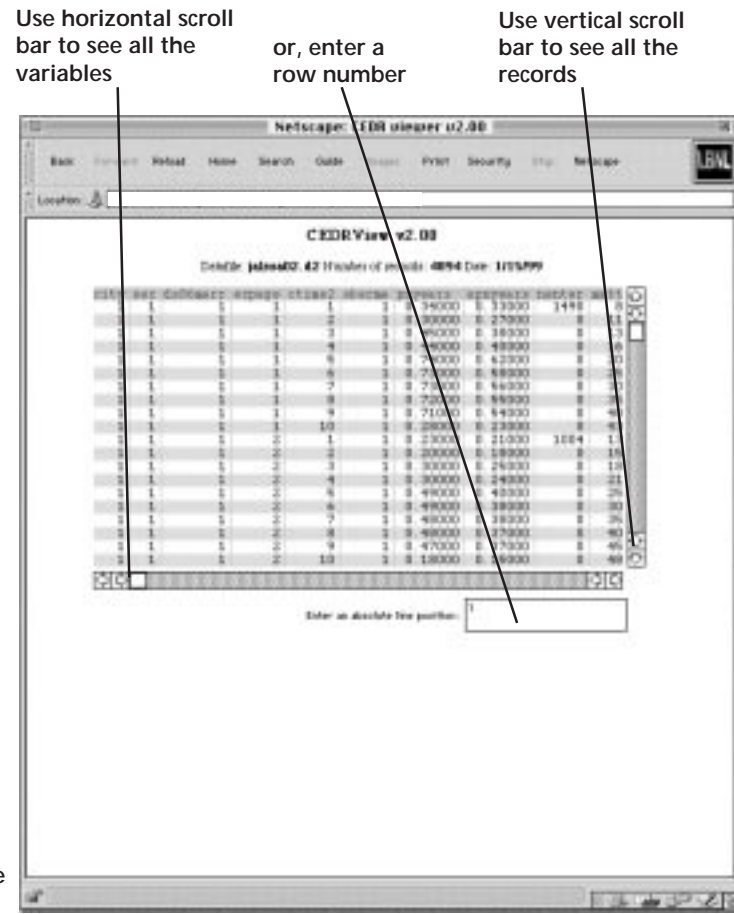


Enter
CEDR ID

Enter
Password

Then click "OK"

Figure 4-10
CEDR Authorized User
ID and Password dialog



Use horizontal scroll
bar to see all the
variables

or, enter a
row number

Use vertical scroll
bar to see all the
records

Figure 4-11
Viewing a CEDR Data File
using a Java Applet