PROJECT LINK-LINK: AN INTERACTIVE DATABASE OF ADMINISTRATIVE RECORD LINKAGE STUDIES

Jane L. Crane, National Center for Education Statistics

Douglas G. Kleweno, U.S. Department of Agriculture

Much information exists on linkage studies using administrative records and, in some cases, survey data. A database called LINK-LINK illustrates the electronic retrieval of linkage study information. This paper is a guide for a prospective user of LINK-LINK. It will briefly describe the database and potential uses of the system, explain how one searches the database for general or specific linkage project information, outline procedures for obtaining copies of the database and address the future direction of the project.

The database is the end-product of a pilot study by the statistical policy committee formed from the Matching Group of the Administrative Records Subcommittee, a standing committee of the Federal Committee on Statistical Methodology. The committee encourages use of the database and solicits comments and suggestions from all users.

A DESCRIPTION OF LINK-LINK

LINK-LINK is an interactive information database devoted to administrative record and survey data linkage studies. The initial database contains 30 studies which were selected for complexity, originality, and diversity of record linkages. Appendix A provides a list of these studies by title.

Information for each study in the database was obtained using a self-administered questionnaire. The questionnaire, designed by the statistical policy committee, was completed for each linkage study. Initial studies were selected for complexity, originality and diversity of linkages. Respondents for the pilot study were contacted by telephone and letter before receiving the questionnaire. After the information was collected, it was edited for clarity and completeness and then it was keyed into the database.

The database is comprised of a series of menu-type prompts to direct the inquirer during the interactive information search. The menu allows the user to choose the search category from a list that appears on the screen. There is considerable flexibility in the database because of a variety of search categories. In addition, the prompts also allow selection of particular area of user interest.

LINK-LINK was written using a dBASEIII software program. The database, which was developed on an IBM PC/XT personal computer, is on a 5¼" floppy disk.

Equipment requirements for LINK-LINK include: an IBM PC/XT or any other fully compatible personal computer with the MS-DOS or PC-DOS Version 2.0 or greater operating system; 256K minimum bytes of memory; two 360K floppy disk drives or one 360K floppy disk drive and a hard disk drive, and a printer with at least 80 column capacity.

Objectives for Developing LINK-LINK

The primary objectives of the database are as follows:

1) inform and educate data users about record linkage activities;
2) identify and describe major record linkage data files;
3) illustrate procedures to meet confidentiality requirements associated with a particular record file;
4) demonstrate linkage methodology including software limitations, data quality concerns and linkage solutions; and
5) identify a knowledgeable contact person for further linkage information.

Type of Information Available

Each study in the database can be referenced to obtain a broad spectrum of linkage study information including: the linkage purpose; linkage methodology including software used; linkage data files; methods used to meet legal requirements for matching; type of dissemination of the linked data; names of cooperating institutions and their contact person; and titles of supporting linkage publications. A more detailed description of the database contents is given in Tables 2 and 3.

Potential Uses of LINK-LINK

LINK-LINK is a reference source for people seeking information on record linkage studies involving administrative records and/or survey data. The database is a useful tool to:

1) identify new and significant linkage programs using administrative records and survey data, or discover the most recent research activity involving linkage of records;
2) identify the potential uses of linkages involving administrative and/or survey data records;
3) identify the complexity and limitations of data linkages as dictated by public policy;
4) keep abreast of research in administrative record and survey data linkages and avoid redundancy of research efforts; and
5) use as a basis for additional research.

LINK-LINK's MAIN MENUS

There are two main menus which provide the user with a large selection of information to investigate record linkage studies contained in LINK-LINK. It is possible to search the database to identify all linkage studies for a certain characteristics such as the linkage

purpose or linkage method. It is also possible to search a specific project for detailed study information. The logic of the system flow is from general categories to specific study detail.

Table 1 shows the system's two main menus with the initial user selection categories. Based on the user's interest, the appropriate menu selector value is entered.

Main Menu I is an exploratory menu to give the user a listing of linkage studies by general category. Main Menu II provides detailed data specific to a study in the database. A series of submenus direct the user to the appropriate information of interest within the main menu.

Main Menu I

The user, upon entering the database, keys "do explore" to display the Main Menu I selection categories. As the user responds to additional menu prompts, the search for information narrows until a list of record linkage studies are identified. The format for the list of studies is a five-digit database reference number, a project title, and a brief statement of the study description. The listing is displayed on the computer monitor and is also routed to a printer for hard copy.

Table 2 provides a brief description of the Main Menu I selection categories. For example, to obtain a list of linkage studies used for the construction of a sampling frame, the user keys a "1" in the Main Menu I and a "1" in the submenu. The end point of the Main Menu I is a list of database linkage studies satisfying the conditions as defined by the user in one or more menus.

At the end point of a path search in Main Menu I, the system prompts the user 1) to return for further exploring using major categories in the Main Menu I; 2) to request specific information for one or more studies listed using Main Menu II; or 3) to leave the system entirely with a series of "Q" or quit prompts.

Main Menu II

Main Menu II provides the user access to detailed information on a specific linkage study. The user must know the five-digit database reference number which is provided when the listing of studies is printed at the end of Main Menu I. Only one study can be searched at a time. The user can request information on additional studies by entering each reference number as requested. All information displayed on the monitor is again routed to the printer for hard copy.

The user will generally access Main Menu II after exploring for information in Main Menu I. The user simply enters Main Menu II with the five-digit database reference number for which additional information is requested. Table 3 describes the five selection categories available.

It is possible, if the database reference is known, to skip Main Menu I and go directly to Main Menu II by keying "do lnktomn2." This command will place you at the beginning of Main Menu II where you will be asked to select from the categories identified in Table 3.

THE FUTURE OF LINK-LINK

At this time, the future of LINK-LINK is uncertain. The Matching Group of the Administrative Records Subcommittee is searching for an individual or Agency to assume responsibility for the database. Because the current version of LINK-LINK is a pilot effort still in the development stage, an evaluation of the database design is in order. In addition, the mechanics for updating current linkage studies and adding new studies to the database must be addressed. It is also necessary to support users who request a copy of the database.

Copies of the LINK-LINK database may be obtained by mail. Send two formated floppy disks for each copy of the database requested and a pre-addressed mailer to return the disks. Specifications for the floppy disks are:

> 5¼" flexible disk
> Double Sided
> Double Density
> 40 tracks

Send correspondence and floppy disks to:

> Fritz Scheuren, Ph.D.
> Chairperson, Administrative Records Subcommittee, Federal Committee on Statistical Methodology
> c/o Statistics of Income Division
> Internal Revenue Service D:R:S
> 1111 Constitution Avenue, N.W.
> Washington, D.C. 20224

ACKNOWLEDGMENTS

Table 1: Main Menu Selection Categories in LINK-LINK

Menu	Selector	Category
MAIN MENU I	(1)	Identification of Linkage Purpose
	(2)	Restrictions on Access of Files for Linkage Purposes
	(3)	Linkage Methods and Related Issues
	(4)	Data Files Used in Linkages
	(5)	Subjects and Respondents on Files
	(6)	Title and Short Description of Linkage Project
	(7)	Type of Dissemination
	(8)	Documentation of Linkage Studies by Title and Author
MAIN MENU II	(1)	Access to Files for Linkage Purposes
	(2)	Linkage Methodology
	(3)	Data File Description
	(4)	Titles/Authors of Written Documentation
	(5)	Contact Person for Study Information

Table 2: Description of Selection Categories for Main Menu I

Selector Category	Description of Contents
1. Identification of Linkage Purpose	Ten linkage purposes are identified. The user selects a category of a list of studies.
2. Restrictions on Access to Files	A submenu with two options are available to the user to identify general study safeguards: 1) studies where access to linkage records is permitted when respondent permission is obtained, and 2) studies where agency policy or legal authority restricts disclosure (general or specific statutes).
3. Linkage Methods	Four options in the submenu permit the user to investigate how database study files were linked: 1) software used for data preparation; 2) software used for matching; 3) data quality problems; and 4) linkage problems. Each submenu prompts the user to select a category of interest.
4. Data File Used in Linkages	Datasets used in all linkage studies contained in the database are listed. Number and title of a study are listed first, followed by the dataset(s).
5. Subjects and Respondents	Four general categories of subject/ respondent interest are available.
6. Title and Description	List of linkage studies with database reference number, title, and study description is available.
7. Type of Dissemination	Four dissemination categories in the submenu are available for the user to obtain a list of linkage studies: 1) released in aggregate form; 2) public use microdata file; 3) restricted use microdata file, and 4) no dissemination.
8. Documentation of Linkage Studies	List of linkage studies with any published documentation by author, title, and date is available.

Table 3: Description of Selection Categories for Main Menu II

Selector Category	Description of Contents
1. Access to Files for Linkage Purposes	Specific information on: parties to the transaction; incentives; how legal requirements were met; how records were obtained; procedures to protect identifiable records during linkages; type of dissemination if any; and steps taken to prevent disclosure after records have been linked.
2. Linkage Methodology	Specific study information on: software used to prepare data files and to link records; problems in data quality; and problems encountered during the linkage process are listed.
3. Data File Description	Specific linkage study data set names and key variables are listed from each data set.
4. Titles/Authors of Documentation	References of publications by title, author, and date for specific linkage study are provided.
5. Contact Person for Study Information	Specific linkage study resource person including individual's title, employer, address, and telephone number are identified.

Appendix A: Database Studies by Title

o Tax Year 1979 Sole Proprietorship Employment and Payroll
o Residential Energy Consumption Survey
o Developing A Sampling Frame Of Petroleum Sellers
o IRS/Census Direct Match Study
o Tax Year 1979 Partnership Employment and Payroll
o Employer Reporting Unit Match Study (ERUMS)
o SRS/ASCS Data Exchange
o Intergenerational Wealth Study
o Enhancing Data From the SIPP With Economic Data
o IRS 1979 Occupational Coding Study
o Linked IRS-SSA Data File
o Updating of the SSEL
o IRS 1982 Estate Collation Study
o Deriving Labor Turnover Rates From Admin Records for U.S. and 30 States
o Mail List Development for 1982 Census Of Agriculture

o High School and Beyond--Third Follow-up Student Financial Aid Record Component
o National Health and Nutrition Exam Survey, Epidemiologic Follow-up Study
o Census/IRS Link Study
o 1982 Partnership Employment and Payroll Link Study
o 1982 Sole Proprietorship Employment and Payroll Link Study
o Continuous Wage/Benefit History Project
o IRS Mortality Statistics Study
o Current Population Survey/ National Death Index Match Study
o Forward Trace Study
o Continuous Work History Sample System
o Wage and Tax Statement Extract
o Information Returns Program Match
o IRS/SSA/DOD Match
o Special Frame Study
o Master Employment List-Unemployment Insurance Records of Texas and Pennsylvania