DOCUMENT RESUME

ED 243 906 TM 830 844

TITLE Archiving Methodology. Volume III: Project-Level Documentation Standard.
INSTITUTION Leinwand (C.M.) Associates, Inc., Newton, Mass.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 31 Mar 80
NOTE 37p.; For related documents, see TM 830 842-845.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Archives; *Databases; Data Collection; Delivery Systems; *Documentation; Information Dissemination; Information Utilization; *Program Descriptions; Research Projects; Research Utilization; *Standards
IDENTIFIERS Contractors; Contrators; Data Management; *Secondary Analysis

ABSTRACT

This volume of "Archiving Methodology" provides contractors with standards for developing project-level documentation. This volume is designed to acquaint a potential data user with the research and/or evaluative effort that generated the archived data. It serves as both an introduction to and an overview of a collection of data files by providing background information on an entire project, including its original purpose, historical perspective, and enabling legislation. Descriptions of component data files, narratives on the background and significance of the data, and summaries of research goals and findings acquaint the reader with the key concepts of the data files. Beginning with sections that explain the purpose, background, and significance of the data files, subsequent sections deal with the study design, the sample represented by the data, the statistical analyses performed, the major findings of the study, and an overview of the actual data files generated by the study. (PN)

ARCHIVING METHODOLOGY

VOLUME III: PROJECT-LEVEL DOCUMENTATION STANDARD

Submitted to

National Institute of Education

by

C.M. Leinwand Associates, Inc.
March 31, 1980

CONTENTS

ACKNOWLEDGMENTS

These standards are based on the direct experiences of the staff of C.M. Leinwand Associates in archiving large numbers of data files and on the experiences others working in data archiving have shared with us. We have attempted to synthesize the efforts of many people working in the field into a single comprehensive document. A large portion of the file-level documentation standard is based on the work of Dr. Richard Roistacher and Barbara Noble of the Bureau of Social Science Research. Additional input was provided by Alice Robbin of the University of Wisconsin and by many other researchers active in the data archiving field.

We view these standards as vital archiving tools and we view archiving as a service which will stimulate future researchers to conduct secondary analyses, thus maximizing the potential use of existing data files. We hope that, as experience with both the documentation standard and with archiving in general grows, the standard will evolve to greater usefulness. We encourage users of the standards to comment on their experiences with the standards. Their comments will assist C.M. Leinwand Associates, Inc., in extending the applicability of the standard to all types of data files.

DOCUMENTATION STANDARD OVERVIEW

The following documentation standards were developed to provide data archivists with a consistent documentation standard equally applicable to small, single-survey data sets, as well as large evaluation projects involving hundred of related data files. The standards were developed to be all inclusive and to explain fully the documentation necessary to describe any collection of data files. In using the standards, data archivists will select those sections of the standard applicable to their specific documentation needs. In this way, consistent documentation can be developed for a wide range of projects. Two levels of documentation are described: project and file.

Project-level documentation is a single volume designed to acquaint a potential data user with the research and/or evaluative effort that generated the archived data. It does this by providing background information on an

entire project, including its original purpose, historical perspective, and enabling legislation. Additionally, project-level documentation describes each of the substudies undertaken as part of the project, detailing target populations, types of information contained in the data sets, and key findings of previous analyses. An interested researcher will read the project-level documentation first to form an initial opinion as to whether the data could serve his or her research needs.

File-level documentation may consist of many volumes, each providing detailed information on a data set obtained from a project. It specifies elements of the individual files (e.g., unit of observation, scale, data time frame) which can vary from file to file. File-level documentation presents additional information, such as codebooks, describing actual contents of the file, and detailed descriptions of the universe, target, and obtained populations. File-level documents describe the individual data sets completely and serve as the analyst's guide to their use. File-level appendices contain extended bibliographies of related documents, detailed coding and editing descriptions, and copies of the original data collection instruments. For some projects, a separate volume describes each file; for other projects, several files are described in one volume or grouped together by substudy.

PROJECT -LEVEL DOCUMENTATION STANDARD

I. INTRODUCTION

Traditionally, the documentation of machine-readable data files has been limited to "codebooks" or "record layouts" which provided basic information about a data file's format and content. Such documentation described each data field within the file, its location, and its values. For some time, this limited documentation was adequate for most secondary research purposes. The "codebook" and "record layouts" bear at least some resemblance to certain aspects of the standard proposed to create file-level documentation (Volume IV of this series). However, when researchers seek information about the study which has generated the data files, they are usually referred to the final report and possibly the original proposal. A very ambitious researcher can try to obtain other study documents if they exist, such as, quarterly or progress reports, memos, correspondence, and study notes. In addition, the original project staff can be interviewed when information gaps exist. With much effort, the researcher can piece together a fairly good description of the study and hopefully document any deviations from the original proposal over the course of the study.

However, this method is time consuming and costly and can fail to provide all of the information needed especially if the researcher was relying only on the final report. Moreover, every researcher interested in a study and its data would have to go through the same process.

As time elapses after the completion of a study, the supplementary materials become scarce and the original contractors engage in other projects. Their memories of the earlier study fade and their patience with repeated questions from secondary analysts can wear thin.

The work of a secondary analyst can be aided if a single document already exists which accurately describes the research study. The quality of secondary analytic studies can be improved if researchers are spared the needless effort of reconstructing the original study.

This volume devoted to a Project-Level Documentation standard is a guide to the preparation of such documentation.

A. PURPOSE

Project-level documentation serves as both an introduction to and overview of a collection of data files. Descriptions of component data files, narratives on the background and significance of the data, and summaries of research goals and findings rapidly acquaint the reader with the key contents and key concepts of the data files. Readers will use the documentation to determine if information contained in the collected data files is potentially useful to them.

Project-level documentation begins with sections explaining the purpose, background, and significance of the data files. Subsequent sections deal with the study design, the sample represented by the data, the statistical analyses performed, the major findings of the study, and an overview of the actual data files generated by the study. A full description of data files and their characteristic in file-level documentation and is presented in a document separate from the project-level documentation. Volume IV in this series presents the standard for creating file-level documentation.

B. ORGANIZATION OF THE PROJECT-LEVEL DOCUMENT

The project-level document has three major sections: project overview, substudy description, and appendices.

- The title
- The project overview summarizes the most important facts about the project and its historical significance.
- The substudy descriptions include sections on substudy purposes, findings, samples, and information on the contents of the data sets in the substudy.
- The appendices include a glossary, bibliography, and other related materials.

II. PROJECT OVERVIEW

The first consideration to be made before constructing the project is to select a clear and meaningful title to identify the research project. The title may be identical to the analysis report title in many cases. However, if such titles contain acronyms, abbreviations, or highly technical terms, they should be renamed since such titles can mislead the reader.

The title, of course, appears on the title page and thus becomes the reader's first source of information about the documentation itself, the data being documented, and the originators of both. A critical issue in naming a document pertains to access. A title must contain sufficient information to permit librarians and archivists to catalogue the data and its associated documentation and to help researchers locate what has been archived.

The title of the project and its archive should be informative and concise, giving the researcher a highly compact description in a minimum of 25 words. Sometimes naming is easy, especially if a project title has been previously developed and consistently used in all project reports. In such cases, that title should also be used for the project-level documentation. If a title has not been created, then one should be developed incorporating the following components:

- descriptive words indicating content;
- geographic focus or unit;
- chronological year(s) of data target or data collection;
- source of data (e.g., court records);
- study or series number (important if data files are ordered).

Example:

" The American National Election Study
(CPS Study 495401 -- ICPR Study 7010)
Pre-Election Wave -- Sept.1, 1972 -- Nov.6, 1972
Post-Election Wave -- Nov. 7, 1972 -- Feb. 13, 1973"

Example:

"Change in U.S. Women's Sex Role Attitudes 1964-1974
#R01-MH25271
The National Institute of Mental Health"

Acronyms do not describe a data file and can confuse unfamiliar users--they should not be used in titles. For example, the "National Longitudinal Surveys of Labor Market History" and the "National Longitudinal Study of the High School Class of 1972" might both be designated "NLS," a name which lacks intrinsic meaning as well as precise meaning.

To facilitate filing and locating a title, begin the title with the most descriptive word, not a number or an article (e.g., "An," "The").

A prefered title for the National Institute of Education Safe School Study is:

"Safe School Study: Project-Level Documentation," not

"1. Safe School Study: .....," or

"The Safe School Study: ...," or

"Project-Level Documentation for the Safe School Study."

Like the title, the title page initially identifies the documentation. It elaborates slightly on the information conveyed in the title by providing:

- authorship;
- title of the document;
- title of the data set being documented;
- edition statement;
- imprint, including date(s), name(s), and place(s) of the documentor and data collector;
- series statement.

Figure 1 is a sample title page.

Figure 1: Sample Title Page

JUVENILE DETENTION AND CORRECTIONAL FACILITY CENSUS OF 1971:
USER'S GUIDE FOR THE MACHINE READABLE DATA FILE <1>

Produced by

U.S. Bureau of the Census <3>
Washington, D.C.
1971 <6>

for

National Criminal Justice Information and Statistics Service <4>
Law Enforcement Assistance Administration
U.S. Department of Justice
Washington, D.C. 20537 <5>

Rev. LEAA 1975 ed. <7>
Revised by LEAA Data Archive and Research Support Center <8>
Center for Advanced Computation
University of Illinois
Urbana, IL 61801
(217) 333-3234

User's Guide Prepared by
LEAA Data Archive and Research Support Center <9>
(Under LEAA Grant 77-SS-99-6003)
December 1978 <10>

LEAA User's Guide 4th ed. <7>

The first section of the project-level document is an overview of the important facts of the study and is designed to give the reader a thorough grasp of its substance and evolution. These facts are organized thematically:

- Abstract
- Background - historical perspective and significance, issues addressed resulting in the undertaking of the study
- Research topics investigated

In making decisions about how much detail should appear in the project overview, it is helpful for archivists to remember that the overview is usually the first contact the potential user has with the data archive and, in most cases, with the studies themselves. The archivist should not assume that users have any previous knowledge about the study and should give the reader all information necessary for an understanding of the study's basic purpose and structure.

The intent of the project overview is to convey clearly and immediately the important elements of the study, how these elements were conceived, and how they articulate with each other. Thus, the overview not only describes the study's historical and theoretical background and the topics it investigated, but also clarifies the overall coherence of these elements, i.e., how they logically flowed together to form the study.

The length of the project overview varies with the complexity and scope of the study. It is recommended that descriptions of complex studies requiring a lengthy overview employ subtitles for organization and emphasis. (See example below.)

A. ABSTRACT

The abstract concisely summarizes the project's subject matter, research goals, data content, and major findings. In a page or less, it helps the reader to survey the contents of the study quickly and to ascertain whether the archive may contain data of interest. It should be short, informative, and easily comprehended by the reader.

EXAMPLE: Abstract

> The Safe School Study was undertaken by the National Institute of Education in response to Congress' request that HEW determine the number of schools affected by crime or violence, the type and seriousness of those crimes, and how school crime can be prevented. The study is based on a mail survey of over 4,000 schools and on an on-site survey of 642 schools, and case studies of 10 schools. Principals, teachers, and students contributed to the study.
>
> Risks of Crime at School
>
> Although school violence and vandalism increased during the 1960's, they have leveled off since the early 1970's, and there are some hints of a decline. Still, about 8% of the nation's schools (6,700) have a serious problem with crime. Secondary schools are more likely to have a serious problem than elementary schools.
>
> The risks of crime directed against schools are higher in the Northwest and West than in the North Central and Southern States, and tend to be spread throughout urban and suburban areas. The risks of personal violence are higher in junior high schools than in senior highs, and are higher in larger communities.
>
> Extent of the Problem: Personal Violence
>
> About 2.4 million secondary school students (11%) have something stolen from them in a typical month. About 1.3% of the students (282,000) report being attacked in a month. Relatively few are injured seriously enough to need medical attention.

> Among secondary school teachers, about 12% (130,000) have something stolen at school in a month's time. Some 52,000 are physically attacked, about 1,000 of whom are seriously enough injured to require medical attention. Around 6,000 have something taken from them by force, weapons, or threats.
>
> Young teenagers in cities run a greater risk of violence in school than elsewhere, except in high crime neighborhoods. There, schools are safer than the surrounding communities.

B. BACKGROUND INFORMATION

Central to this section of the project overview is the rationale for collecting the data being documented. After initially stating the purpose(s) for the data collection effort and the major research question(s), the historical context of the study is described.

Many projects are undertaken as part of an evaluation of a social program. In such cases, it is important that the reader understand the context in which the program was conducted, as well as the specifics of the data themselves. Therefore, the program is described--its goals, intended results, and the events leading to its implementation.

The significance of the study and its data can be evaluated by reference to the research community and existing theory. Theoretical significance can be shown by answering the questions,

- What makes the data contained in this study of interest?
- What is the relationship of these data to those of other studies examining similar questions?
- On what specific problems and issues did the study focus?

Finally, if the collection of these data was precipitated by some key event or action, such as a Congressional mandate, or in response to a riot, this seminal event or action is described.

EXAMPLE: Background

> The National Crime Surveys are conducted for the Law Enforcement Assistance Administration by the U.S. Bureau of the Census in order to obtain current and reliable measures of serious crime in the United States. The program is unique in that it turns to the public for information about crime rather than relying on law enforcement records. The primary impetus for this type of approach was an earlier victim survey which indicated that a considerable number of crimes were not being reported and consequently existing crime measures did not reflect the true rate of crime occurrence. In addition to providing an alternative or supplementary source of data for estimating the amount of crime, the surveys offer a level of descriptive detail about the victims of crime and their victimization experiences not previously available.

C. MAJOR RESEARCH QUESTIONS

The purpose of this section is to outline the major research question(s) asked. Although this section could be as brief as a paragraph, it should not exceed three pages.

It is important to note that while every study is guided by one or more research questions, the level of specificity may differ. Studies often pose a general research question which is answered by posing multiple questions, each focusing on components of the larger question. This is particulaly true when overviewing a large-scale project which represents multiple substudies, each of which, in turn, entails several data collection efforts. Wherever possible, it is useful to choose a relatively specific set of questions for inclusion in the overview, since this type of question is more meaningful to the reader. However, in the

case of large-scale projects, this approach might result in pages of detailed questions; the archivist is then urged to choose a smaller number of more general questions.

EXAMPLE: RESEARCH QUESTIONS

The Safe School Study was designed to address the following research questions:

- How serious is the problem of crime and disruption in schools?
- How many schools, students, and teachers are affected, in what ways, and to what extent?
- When and where are the risks of crime and violence highest?
- Who are the victims and offenders?
- What are the attitudes and experiences of the victims?
- What factors are associated with violence and vandalism in schools?
- What measures are schools using to reduce or prevent crime?
- What measures do principals, teachers, and students recommend?
- What are the implications of this research for policy?

D. SUMMARY

The following example illustrates each overview component (abstract, background, research questions) discussed separately above. In this example, due to the complexity of the project and the importance of its historical context, an extensive background component was judged vital to an understanding of the project.

EXAMPLE: Project Overview

COMPENSATORY EDUCATION STUDY

ABSTRACT

The National Institute of Education's (NIE) Compensatory Education Study was a comprehensive study undertaken at the request of Congress. Its purpose was to gather information which would help in developing and considering future legislation. After issuing a request for proposals (RFP), NIE commissioned more than 35 substudies whose total cost was $15 million over a three-year period. Ten of those substudies investigated Title I funds allocation; two surveyed the nature of compensatory education services; seventeen researched student development; ten investigated the administration of Title I and state compensatory education programs. In addition to research in these four areas, NIE commissioned the development of a data archive. Its objective was to provide other researchers access to the compensatory education data for secondary analysis purposes. Six of the substudies were chosen for inclusion in the archive.

BACKGROUND

The compensatory education concept stems from the recognition that children from disadvantaged backgrounds frequently do not enjoy the same educational benefits as their more affluent peers. Many disadvantaged children attend schools in districts characterized by low overall revenues and/or high concentrations of disadvantaged families. Such circumstances place special strains on the schools and adversely affect the general educational development of pupils. Compensatory education is intended to ease these problems by providing disadvantaged children with additional services which will help them complete their education on more equal terms.

The Elementary and Secondary Education Act (ESEA) was enacted in 1965; Title I of this act targeted large sums of money to "meet the special needs of educationally deprived children." Many hailed this legislation as an effective and central part of the War on Poverty because they believed it would, for the first time, open equal educational opportunities to children handicapped by their material circumstances. In enacting Title I, Congress first sought to provide

additional financial assistance to school districts serving large numbers of students from low-income families and to schools enrolling the greatest number of such students. Second, Congress sought to fund special services for low-achieving children in schools with low revenue levels. And third, Congress intended Title I programs to contribute to the cognitive, social, and emotional development of participating students.

Although Title I allocated large sums of money, it also included direct provisions limiting school districts' use of this money and requiring districts to perform self-evalutions. These evaluations were to be monitored by each state and analyzed in terms of cost benefits. Thus, results could be fed into the United States Office of Education's (USOE) decisionmaking process and channeled back to the school districts.

In the years following the enactment of Title I, some of the early optimism faded because educators found weaknesses in certain aspects of Title I administration and because information about the program was generally unavailable. In practice, the unprecedented evaluation provision failed to produce the information needed to assist USOE in its decisionmaking. Federal and state enforcement of the evaluation was weak; schools resisted the requirement; the surveys performed lacked the hard data necessary to demonstrate success and were rarely read. Often, early national evaluations of Title I focused solely on children's academic performance, or, more selectively, on reading programs only. In doing so, these evaluations neglected the diverse programs also funded by Title I (e.g., math, science, health, counseling, and cultural enrichment). The scarce information concerning the nature of programs funded by Title I tended to describe only the incidence and distribution of various types of programs. Few studies provided data concerning the cost of operating compensatory education programs and whether or not the programs really worked.

In addition to this paucity of documentation on program design, effectiveness, and cost benefits, a growing debate surrounded the use of poverty measures as a criterion for Title I eligibility. Advocates of a change from poverty measures argued that since the goal of compensatory education is to raise achievement levels, the most appropriate criterion is one which targets money directly to low-achieving children. Defenders of the status quo contended that the achievement criterion neglected the poor

and subverted the fundamental intent of Title I: to raise poverty levels by providing equal educational opportunity to the disadvantaged. Other criticisms of Title I were that districts misused their funds, either stretching them too thin or treating them as general aid and not really providing compensatory education services. Some critics charged the State Educational Agencies (SEAS) and the USOE Title I offices with inadequate monitoring of districts. Title I services were considered by some a mere substitution for state compensatory funds. Others blamed the federal guidelines for restricting districts from providing effective programs. A final criticism raised doubts about the basic intent of Title I: namely, that the relationship between poverty and educational performance had been misinterpreted and misapplied.

The Education Amendments of 1974 (PF 93-380) reflected these dissatisfactions and sought to obtain information on ways to reauthorize Title I. The amendments contained provisions which mandated NIE to

1. examine the fundamental purpose and effectiveness of compensatory programs;

2. analyze ways of identifying children in greatest need of compensatory education;

3. develop alternatives for meeting these children's needs;

4. consider the feasibility, costs, and consequences of alternatives for distributing compensatory education funds.

NIE was directly responsible to Congress for the design and execution of the Compensatory Education Study. The Study's framework was intentionally broad to encompass three elements common to all federal education programs:

1. Congress' objective to improve education for children;

2. funding and allocation procedures which Congress establishes in pursuit of its objectives;

3. operation of federal, state, and local agencies as Congress' agents in using federal funds to obtain its objectives.

RESEARCH QUESTIONS

To fulfill Congress' mandate, NIE commissioned a total of 35 substudies, including a National Survey of Compensatory Education, special demonstration projects concerned with alternative funds and allocation procedures, and a number of detailed case studies of particular aspects of the Title I program. The 35 substudies represented four foci of research activity.

1. Fund Allocation

NIE's mandate to examine this topic was defined in clauses (2), (4), and (5) of Section 821(a) of P.L. 93-380. This section called for investigation into the costs and feasibility of using alternative eligibility criteria for allocation of Title I funds; current allocation practices; and assessments and simulations of the effects of alternative criteria. Experimental projects in which school districts selected their own eligibility criteria were also conducted.

2. Administration of Compensatory Education

Clause (6) of Section 821(a) of P.L. 93-380 most clearly expressed this area of research activity. A survey of state regulations supplementing USOE Title I regulations was conducted, as well as case studies describing the implementation of regulations in a regionally representative sample of states and districts.

3. Services to Students

The federal mandate to investigate this topic appeared in clauses of Section 821(a) of P.L. 93-380. Subsequent substudies included a survey of all Title I programs to determine the nature and types of services presently provided.

4. Student Development

The NIE mandate to examine this area of compensatory education was contained in clauses (1) and (3) of Section 821(a) of P.L. 93-380. Research efforts consisted of surveys of compensatory activities to describe goals and methods employed by districts. Data were synthesized and re-analyzed using materials from education agencies, the federal government, education organizations, and independent researchers.

III. SUBSTUDY DESCRIPTIONS

After the project as a whole has been described, the archivist provides a brief overview of the relationship between the substudies comprising the project. The goal of this section is to inform the reader of how these substudies come together within the archive. Before this goal can be achieved, the archivist must first organize the project into substudies.

Since the definition of what constitutes a substudy is not always obvious, we offer three guidelines to assist in this organizational task.

1) Certain large-scale research projects actually consist of multiple independently-conducted studies; each independent study is a substudy. Usually, each of these substudies is undertaken by a different contractor (although this is not an essential defining characteristic); each has a different focus, and each attempts to answer a specific set of research questions.

EXAMPLE: WHAT IS A SUBSTUDY? (1)

> The Compensatory Education Study data archive consists of six independent substudies conducted by six different contractors. Each substudy focuses on a different aspect of Compensatory Education: Census Tabulations; Achievement Measures; Demonstration; State Administration Survey; National Survey; Instructional Dimensions Survey. Each also generated multiple data sets.

2) In some studies, data are collected from different target populations and generally analyzed separately. Each target population's data is a substudy, since a separate study design was the basis for selecting each type of target population and analyzing its response.

EXAMPLE: WHAT IS A SUBSTUDY? (2)

The Education Voucher Demonstration data archive consists of data collected longitudinally on students, teachers, parents, and schools. Separate surveys were administered to each of these target populations, and, while data from different files were sometimes merged, analysis activities generally focused on each group as an independent entity. Therefore, the data would be organized into four substudies.

3) A project operating under a single study design usually does not have component substudies, even if multiple data files were generated. The key determinant is whether the different data files were used to answer substantially different research questions or to answer the same questions from different perspectives. Sometimes, this distinction is quite difficult to make, since all projects have an overall research goal governing the collection of each data set. The issue is whether substantively different research issues were analyzed using each set of data.

EXAMPLE: WHAT IS A SUBSTUDY (3)

The Safe School Study collected data from principals, teachers, and students throughout the United States. Different data sets were collected from each group, e.g., teachers and students provided information on questionnaires and in follow-up interviews. However, all of the data sets were utilized to develop estimates of crime levels in schools in the U.S., to support various hypothesis of the causes of crime, and to determine the most effective crime prevention measures. In this study, each data set was analyzed in light of the same research goal, so only one substudy existed.

Each substudy is then detailed. Each substudy description consists of the following components:

- title,
- purpose,
- design,
- sampling procedures,
- data analysis,
- major findings,
- file descriptions.

A. TITLE

Substudy titles need not be as detailed as project titles. Usually, they are compact and contain no more than 15 words. However, some of the conventions for the project title are followed in devising substudy titles, notably, the use of descriptive words, the placement of the most important word first, and the avoidance of acronyms.

EXAMPLE: SUBSTUDY TITLE

Study Title:

Compensatory Education Study;
National Institute of Education

Substudy titles:

1. Fund Allocation within Districts, Demonstration Studies
2. Instructional Dimensions Study, 1976-1977

B. SUBSTUDY BACKGROUND AND PURPOSE

The description of the substudy's purpose features the same components as the description of the project's purpose, except that it explains the relationship of the substudy to the overall purposes of the project. It is usually one to two pages in length.

EXAMPLE: SUBSTUDY BACKGOURND AND PURPOSE

> The Instructional Dimensions Study was funded by the National Institute of Education (NIE) as part of the Compensatory Education Study, a comprehensive research project conducted in response to a mandate by the U.S. Congress in the Education Amendments of 1974. The Instructional Dimensions Study is designed both to gather data on program effectiveness in reading and mathematics and to compliment the NIE National Survey of Compensatory Education, which describes the ways Title I funds are used nationally.
>
> The Instructional Dimensions Study is an in-depth assesssment of the relationships between selected instructional constructs and students' achievement. These constructs are briefly described below....

C. SUBSTUDY DESIGN

The major question in this section is, What was investigated? It is answered in a one- to four-paragraph description.

EXAMPLE: SUBSTUDY DESIGN (1)

> In order to assesss the validity of our attitude measures and the idea that structural positions influence attitudes, we first investigated the cross-sectional associations between personal characteristics and attitude responses. Thirteen socio-demographic variables relevant to women's sex role attitudes were seen as falling into four categories. The four categories were then cross-tabulated with the socio-demographic variables.

EXAMPLE: SUBSTUDY DESIGN (2)

> The effects of changes in demand for organizational services on various personnel categories were examined. While the approach is intended to be general, the specifics of this theory are generated by characteristics peculiar to the organization under study, namely, school districts. After discussing the empirical findings, implications of the model for the study of administrative intensity were drawn.

A secondary question in the study design is, How was it investigated? To answer this question, important variables in the study are identified, or unique indices, scales, or measures created or used in the study are noted.

EXAMPLE: SUBSTUDY DESIGN (3)

> The analysis focused on the independent variables of marriage and husband's characteristics. The dependent variables were all cultural variables concerning sex role ideology.

EXAMPLE: SUBSTUDY DESIGN (4)

> The Orshansky index was used to determine poverty cutoff points.

D. SAMPLING PROCEDURES

A good sampling description gives readers a precis of the sample design and helps them to decide whether data in a substudy might be appropriate for their particular research purposes. Therefore, sampling procedure is one of the most important elements for a researcher to examine in substudy documentation.

Each description of sampling procedures includes descriptions of the universe, target, and obtained samples, and the data collection method. If appropriate, multiple sources of data and multiple surveys within a

substudy are also briefly described. A detailed discussion of the actual sampling technique or specific sampling design is unnecessary, since detailed documentation is provided in the file-level documentation.

1. Target Sample and Universe

This section describes the population of respondents or subjects within the substudy data files. A brief description of the universe of subjects is given first, followed by a description of the actual sample obtained for the data files. The reader is usually most interested in exactly who or what the data represent and the factors limiting the sample's use in analysis activities. This is an especially important consideration in large national sample surveys to prevent researchers from making unwarranted assertions based upon the data. Therefore, the archivist must cite any known limitations the sample imposes on potential analysis.

In addition, this section briefly describes the sampling method (random, stratified random, etc.), response rates, and special procedures used to improve response. If multiple sets of data were collected for a substudy, each sample is described.

2. Data Collection Methodology

The method of collecting the data is explained here: mail survey, administered interview, observation, etc. Special characteristics entailed in the data collection process and the impact of these characteristics on analysis are briefly described. This section also notes when the data were collected.

EXAMPLE: SAMPLE PROCEDURES (1)

> The primary data of the present study are from a questionnaire survey of 1,079 social scientists in the disciplines of anthropology, economics, political science, and psychology. The total

population was all social scientists; the target sample consisted of the academic members listed in the most recent directories of the major professional associations representing the four disciplines.

A random sample of 500 was drawn from each discipline directory, and an initial 12-page form was sent in December, 1973. Four months later, after a second mailing of the questionnaire and a final follow-up letter, 1,079 usable responses had been obtained. The return rate was 54.0%, a fairly common response level for college faculty.

EXAMPLE: SAMPLING PROCEDURES (2)

1964 NORC College Seniors Study: Conducted by the National Opinion Research Center at the University of Chicago, this mail questionnaire survey was based on a national probability sample of approximately 42,000 women and men graduating from American arts and science colleges and universities in June, 1961. First questioned in 1961, the sample was remeasured for a second time in the fall of 1964. The analysis used a randomly-selected 10% subsample of the total sample remeasured in 1964. The attrition rate for the 1964 data was 85% of the 1961 respondents. The response rate in 1961 was 95%.

EXAMPLE SAMPLING PROCEDURES (3)

Fifteen-Year Follow-up Survey 1970: Conducted by Bruce Eckland at the University of North Carolina, this survey sent mail questionnaires to a sample of women and men drawn fifteen years earlier by the Educational Testing Services. The original respondents were high school sophomores and seniors at the time. Although the sample was national in scope, it was not a probability sample of all high school sophomores and seniors. It can be viewed as either a probability sample of such persons with major exclusions (such as large cities) or as a national nonprobability sample. About half of the original respondents to the Educational Testing Service study responded with usable questionnaires to the Eckland follow-up (7,500 out of 15,000).

E. DATA ANALYSIS

The data analysis section includes a brief explanation of the statistical methods used in the study. Knowledge of the data analysis techniques used will enable future researchers to decide if additional types of analyses are indicated. However, detailed descriptions of the analyses performed in the original study are of relatively limited value to future users of the archive. If these users desire more information about the statistical procedures used in the study, they can obtain the final report of the original study which is cited in the bibliography.

EXAMPLE: DATA ANALYSIS (1)

> Commonality analysis was used to attribute portions of variations in achievement gain to the various study model elements and to pretest performance. Costs were also related to achievement gain.

EXAMPLE: DATA ANALYSIS (2)

> Achievement eligibility rates were estimated using regression procedures corresponding to each of two models (Rank-9 and Rank-8). The independent variables included proportions of children in each location category, racial/ethnic group, and level of property status.

EXAMPLE: DATA ANALYSIS (3)

> Two kinds of analyses were performed in this substudy. Summary and breakdown tables (descriptive statistics) were used to describe the effects of the current poverty formula (e.g., its distributional impact). Simulations using the Federal Education Finance (FEF) model were conducted to analyze the effects of the alternative poverty formulas.

F. MAJOR FINDINGS

Only the major findings of the substudy are listed and summarized in one to three paragraphs. This summary includes a statement of whether

or not the substudy supported the hypothesis and a discussion of the relationship of the findings to the research questions described in the substudy purpose and design sections (B., D.). The summary should be a straightforward account of the study findings and should not attempt to discuss the significance of these findings, or their relationship to theory. nor to quote percentages, tests of significance, or other measures. Readers requiring such specific information can consult the final report cited in the bibliography.

EXAMPLES: MAJOR FINDINGS

> There is support for the interchangeability principle.
>
> Most of the combatant nations in the study experienced substantial increases in their rates of homicide. The increases were pervasive and occurred after large and small wars.
>
> The model accounts for a large proportion of the variance in recidivism among first offenders.
>
> The number of friends who use marijuana is the best single predictor of whether the respondent will use marijuana.

G. FILE DESCRIPTION

This section briefly describes each data file within a substudy. It informs readers of the type, scope, and scale of data in each file. Each file description tells about

- the type of data in the file, alterting the reader to unexpected data and highlighting important or unusual contents (e.g., this file is the only known source of national-weighted data on violence in schools broken down by location within school);

- the data collection instrument used to create the file;
- the number of data items per subject;
- the number of subjects.

It is important to emphasize that the file description which occurs in the project-level documentation is a very brief overview of the data files associated with a substudy. The file descriptions should not be confused with file-level documentation which appears in the separate file-level document. Recall that project-level documentation is researchers' first look at a study; they are deciding whether or not to use the data and the brief file description completes the picture. File-level documentation is read when the researcher has decided that the data meet his/her needs and, at that point, require more detail in order to design their secondary analysis.

EXAMPLE: FILE DESCRIPTION (single-data file)

> Semi-structured questionnaires were administered to a national random sample. The file contains responses from 1,504 persons on 316 variables. There were four decks or cards per respondent.

EXAMPLE: FILE DESCRIPTION (multiple file substudy)

> Data for the longitudinal study comprise 37 separate files. Thirteen instruments were used to collect data for year one; identical instruments were used for year two. Eleven of these instruments were used again in year three; however, the Math and Language Arts Program Description Instruments were not used in the third year. A brief discussion of each Demonstration file follows.
>
> Classroom Roster
>
> All homeroom teachers in third- and fourth-grade classrooms in each sample school were asked to use a self-administered form to describe the characteristics of students participating in compensatory activities. These "rostered" students

represent the universe of students in the Demonstration. Three files, one for each of the Demonstration years, are contained in this archive.

1. The first year consists of data for 51,029 students on 38 variables.

2. The second-year file includes 38 variables for 47,859 students.

3. The third-year file contains data for 45,529 students in 38 variables.

Principal School and Background Information

All elementary school principals in the Demonstration districts were asked to complete a self-administered questionnaire to describe the compensatory program(s) operating in their schools. These district/school-level data are contained in three files.

1. The first-year responses of 596 principals to 348 background variables are included in this file.

2. The second-year principal responses representing 152 background variables on 474 schools reside in this file.

3. The third-year file contains 284 variables from 367 principals about their schools' programs.

Principal Supplement 1, Title 4, E.S.E.A.

These data are the self-administered questionnaire responses of all principals of schools receiving Title I. The data contain descriptions of Title I services provided in each school. Principals were interviewed in each of the study years; thus three files for each administration have been archived.

IV. APPENDICES

A. BIBLIOGRAPHY

Two types of bibliographies, each divided into two sections, may appear in the appendices. The first type is an overall project bibliography; the second type is substudy bibliographies. Each of these bibliographies may be broken down into two sections. One section lists reports generated by the study or substudy, that is, reports that were products of primary or secondary analysis of the study's data. The second section contains references to reports, articles, and papers which provide background information, such as enabling legislation or discussions of the issues or topics being researched.

B. GLOSSARY

The purpose of the glossary is to demystify project terms for those persons who will conduct research long after the original research team disbands. Secondary analysts using data from a project may come from a wide variety of academic disciplines, private research organizations, or policy groups. In many projects, a number of key terms are used again and again, and take on new and special meanings that are eventually taken for granted by persons working on the project. Such terms are a familiar shorthand for those directly involved in a project; to a secondary analyst, however, they may be vague, meaningless, or likely to be misconstrued. For this reason, a glossary of terms used in each substudy is incorporated in the appendices to the project-level documentation. The glossary is alphabetized, sometimes

several pages in length, and includes acronyms, abbreviations, and special terms referring to

- government agencies, research consortiums, or research organizations;
- enabling legislation, administrative units, and programs administration;
- study variables;
- data analysis models;
- other concepts.

The glossary also includes more familiar words whose meanings are "new" because of the way they are used in the project.

Example: Acronyms for government agencies, research consortiums, or research organizations.

1. AAAS. American Association for the Advancement of Science.
2. APA. American Psychological Association.
3. DHEW. Department of Health, Education and Welfare.
4. NIE. The National Institute of Education.

Examples: Acronyms and terms referring to enabling legislation, administrative units, and program administration.

ESEA. The Elementary and Secondary Education Act (ESEA) of 1965.

C.E. Refers to the Compensatory Education Study as a whole.

Section 821. An educational amendment instructing the National Institute of Education to conduct studies evaluating the ESEA Title I compensatory education programs.

Examples: Acronyms and names for variables in the study; operational terms.

IDS. Instructional Dimensions Study, a substudy of the Compensatory Education Study which assessed the separate and combined effects of five sets of classroom processes: opportunity, individualization, instructional events, motivational processes, and teacher background.

Matching by mastery. In the Instructional Dimensions substudy, the presence and use of curriculum-provided method to assess mastery.

Motivational processes. In the Instructional Dimensions study, the measures of classroom climate and the incentives teachers provided students for learning.

Examples: Terms for data analysis and models.

D. The index of dissimilarity which measures segregation. Otis and Beverly Duncan use the term to refer to ecological segregation.

Pullout instruction. Refers to the teacher actually taking the student out of the classroom setting for individualized instruction vs. regular instruction where the student is given instruction with the combined class.