

Lucci, York; Rokkan, Stein; and Meyerhoff, Eric. A Library Center of Survey Research Data: A Report of an Inquiry and a Proposal. New York: Columbia University School of Library Service. June 1957. [Available in microform from the: Columbia University, School of Library Service, New York, New York 10027]

It has been twenty years since the Lucci and Rokkan report first appeared. In many ways the data archive movement may be said to have begun with its publication. Many ideas presented therein have been built upon by later archivists and scholars. It is particularly interesting that while the Columbia Library School received the grant for the report, libraries have, until quite recently, been reticent to become involved in data archive activities. [See review of Howard Dalby White's Ph.D. dissertation below.]

#### Background of the Report

Rapid growth of opinion polling and survey research following World War II made cross-national research a real possibility for scholars through the use of secondary analysis of comparable information collected in different countries. The difficulties of locating and obtaining even domestic researches, however, proved discouraging. The need for a library of survey research data which would assemble the more important survey data and make them readily accessible to scholars for secondary analysis prompted Professor S. M. Lipset then at Columbia University to propose that an investigation be made of the need for and the problems of establishing an international library center of survey research materials. In response to this proposal the Behavioral Sciences Division of the Ford Foundation awarded a grant in the spring of 1955 to the Columbia University School of Library Service to undertake such an investigation. York Lucci was given responsibility for the investigation in the U.S. and Stein Rokkan for Europe. In addition, Eric Meyerhoff served as a consultant on the more technical archival considerations.

The field investigators were charged with ascertaining and evaluating four areas: 1) Potential utilization of a center for research materials; 2) Availability of research material, its cost and difficulties in acquiring; 3) Adequacy of available data in terms of their potential for secondary exploitation; 4) Research and administrative requirements involved in such a library center.

The results of the Lucci U.S. investigation and Rokkan European investigation appeared as separate reports since the situations in the U.S. and Europe differed greatly with respect to the availability of survey materials and the extent of interest among scholars.

Part I. The Situation in the U.S. and Recommended ActionBackground and Purposes

Lucci discusses two chief reasons for a centralized archive: lack of funds for the collection of primary data and inadequate use of data already collected. He then outlines past efforts to collect and disseminate results of polls and surveys such as the 1938 Public Opinion Quarterly's AIPO poll results, various periodical compilations published for short intervals, and the Roper depository at Williams College. He concludes that published releases do not permit systematic analyses and that any archives that exist have been too limited and have failed to develop archival practices permitting usage by a wide community of scholars.

The "Loss of Data" receives much attention in the Lucci report. A salient example is that of the Canadian Institute of Public Opinion where all but eight of 212 polls between 1941-1951 had been destroyed. With other examples of data loss, such as the experiences of the Italian DOXA organization, the Swedish Gallup Institute, and the U.S. Office of War Information, Lucci builds a strong case for the need to undertake a strong preservation effort. Parallel developments in other disciplines of basic tools for research (cyclotrons for scientists, libraries for historians) are compared with the lack of basic preservation of the raw material for social science research. The Human Relations Area Files is discussed as the best known effort in the social sciences to assemble information from a variety of sources and make it available to a wide community of scholars.

One hundred and thirty seven social science researchers in the U.S. were contacted for ideas and reactions concerning a central data library for the social sciences. Their ideas are consolidated and set forth by Lucci in the following sections.

Sources of Data

The massive accumulation of data gathered by universities, the government, independent agencies, and foundations cannot be readily documented, but Lucci estimates that five to ten million interviews have been conducted annually in the post-wars years. One market research organization alone reported 25 million cards in storage. Though much of the data may be insignificant, if even five percent offer the possibility of scientifically useful secondary analysis, a quarter to a half-million cards a year might be stored. Urgently needed is some means of locating that fraction of theoretically significant material, assembling it, making its contents available to researchers, and encouraging its exploitation.

Queries to survey organizations and academic institutions produced an overwhelmingly positive reaction about sharing data. Lucci details at length responses from prominent centers, polling agencies, governmental agencies, and commercial research organizations and includes a section on "Conditions Imposed by Contributors." The main objections he found to a proposed center were feelings that one cannot work successfully with someone else's data unless there is

close liaison with the person or agency initially responsible for collecting the data. Lucci counters this by advocating complete documentation and noting that "if survey research is to make any claim to science it can hardly refuse, indeed it should welcome, having a set of facts subjected to close scrutiny and the possibility of alternate interpretation.

#### Utilization of an Archive of Survey Data

Uses of survey data for secondary research are discussed and six major points elaborated upon: 1) as the only primary source of knowledge about certain facts or patterns of behavior; 2) as a means of testing hypotheses; 3) the cumulation of cases; 4) to prepare for new primary research; 5) to assist the work of the historian; and, 6) to train students.

Objections to the proposal for establishing a center of survey research materials related to the extent to which utilization could be anticipated. Respondents to Lucci's questions felt that technical familiarity with methods of quantitative research was not widespread and that only large universities such as Michigan, Columbia, and Chicago could make use of such facilities. Lucci feels that availability will increase the skills of the entire social science community, and then reports comments of respondents, summarizing at length areas where respondents felt the center could be especially useful.

Further justification for the establishment of a center is given by a bibliography (pp. 134-38) appended to the report of published works based on secondary analysis, which represent different ways in which secondary analysis techniques can be used; expansion upon the use of secondary analysis by students and commercial organizations; and, the need by international organizations such as UNESCO for access to such a data center.

#### Conclusion and Recommendations

Lucci summarizes by noting, "On the basis of the present inquiry, it can be stated flatly that the data are available...and it seems fairly clear that the possibility of their use is present and can be developed further." Such a center would maximize the use of available social science talent (especially at universities with minimal facilities); maximize the use of research funds; promote the comparability of data collection; facilitate cross-national research; preserve data in a systematic manner; and, spearhead new studies where lacks are noted.

A "Specific Proposal" is outlined with functions of the center noted at length: systematic collection and preservation of data; an index to data stored; information activities about the data to be disseminated to the research community; promotion of the center's materials; provision of duplicate sets of data for researchers with their own facilities; provision of tabulations where appropriate; maintenance of records of secondary research activities and publications; and, promotion of training and standardization.

Each of these functions is discussed in great detail along with possible locations for the center, staff considerations, equipment and storage needs, operating costs and sources of funding.

Part II: A Review of the Situation in Western Europe and  
Some Recommendations for Possible Action

Rokkan sketches rough estimates of the data accumulation problem in Western Europe. He estimates that in the four major nations: France, Italy, the United Kingdom and Germany, half a million interviews have been generated in each nation since 1951. Among the smaller countries: Austria, Belgium, Denmark, Finland, the Netherlands, and Switzerland, the yearly production of interviews is on the average 75-100,000. Thus, the total interviews for Western Europe are between 3 and 4 million annually.

The problems of preserving, storing and choosing relevant data from this quantity is seen by Rokkan as directly linked to problems of social science utilization. Most polling agencies limit analysis to the presentation of overall response distribution, never analyzing the data as thoroughly as their methodological quality and their theoretical relevance seem to justify.

Surprisingly little has been done by Western European social scientists to avail themselves of these analysis opportunities. The lag in utilization reflects the conservative reluctance of great numbers of social scientists to accept the new techniques as scholarly and scientific, the uneven distribution of statistical analysis skills among academic social scientists, and the difficulty of access to the primary materials in the polling agencies.

During 1955-56 Rokkan discussed problems of storage and data utilization with directors and officers of survey research organizations and with university teachers and research workers in twelve countries of Western Europe, as well as at regional meetings such as WAPOR, ESOMAR, the Congress of the International Political Association, and the International Sociological Association, in order to elicit reactions to the possibility of a solution through the development of an international archive of survey materials.

The possibility of an international archive was presented to survey organizations by a questionnaire devised by Rokkan. The response was unanimous in emphasizing the need for systematic encouragement of secondary analysis, although opinions varied about the steps to be taken to facilitate secondary analysis of data already assembled.

The responses were capsulized by ESOMAR President, Leif Holback-Hanssen: Western European materials should be stored and classified in a separate facility located within the Western European region; development of the archive should be planned in close association with representatives of the survey profession; and, the archive should be built up gradually through a series of pilot analyses.

Rokkan's queries to academic social scientists are discussed in the context of the status of survey research in Western Europe: i.e., there was little being conducted. Most social scientists tended to favor independent gathering of data in order to build gradually to the need for a regional or international repository. Rokkan then discusses the format of pilot analyses and possible topics of these analyses.

Rokkan concludes: there is a need for fuller utilization of the bodies of information assembled by survey organizations through systematic analyses from the primary record; a need for the organization of a program of comparative secondary analyses of survey materials in selected substantive areas and, concurrently with them, the development of an archive of selected materials; and, a need for "groundwork" analyses and an archive within the Western European region. Cost estimates are given.

#### Appendices

As noted above, a selected bibliography of works done using secondary analysis techniques is included. Another appendix details methods for indexing and cataloguing archival materials.

#### Reviewer's Comments

The Lucci-Rokkan report is a fundamental document for the data archivist. The well formulated report gives the background of the data archive movement and predicts many of the developments we are now experiencing. This report is essential reading for all involved in data archives, because of the perspective it provides on the U.S. and Western European research communities and because of its visions of the future.

White, Howard Dalby. "Social Science Data Sets: A Study for Librarians." Ph.D. dissertation, University of California, Berkeley, 1974.

As noted in the review of the Lucci-Rokkan report above, libraries have not generally accepted responsibility for acquiring data in machine-readable form although a number of scholars and librarians have argued that machine-readable data is a logical extension of a library's collection. White explores the relationship between libraries and archives, giving a historical overview of arguments that advocate the placement of data archives within libraries. He compares library and archival functions and although he explains that many tasks are similar, he concludes that archival tasks confront librarians with much that is unfamiliar, such as "opening tapes"; making back-up copies; updating originals; reproducing and distributing tapes; providing programs; and, training users in computer techniques; thus, making the merger of the two unworkable.

In "Data Archives as Publishers," (Chapter III) White discusses the most significant distinction between libraries and archives: major archives "publish" data sets. In his discussion, White gives a wealth of detail about the process of publication within data archives--material that is unavailable elsewhere.

Rather than merge libraries and archives White argues for an alternative: using libraries to make information about data more widely available by acquiring from publisher archives the human-readable materials by which the content of data files can be known. Suggestions for better bibliographic control over machine-readable data through traditional library catalogs are given. Because all social scientists are not employed at institutions with data archives, White argues that there is a need for an organization with cross-disciplinary responsibilities--a library or social science information center--to function as a locale for code-book browsing or searching.

Part II of White's thesis, "Buying Patterns at Two Publisher Archives," is an analysis of the transaction records of the International Data Library and Reference Service of the University of California, Berkeley (IDL&RS) and the Roper Public Opinion Research Center in Williamstown, Massachusetts. The analysis is meant to serve as a basis on which libraries may predict the potential need for codebooks and data sets. White's intensive exploration of the type of materials bought (human-readable vs. machine-readable) also presents an interesting picture of the buyer of the data bases who he hypothesizes tend to be "high-status" persons (professors and researchers rather than students); from "high-status" departments (e.g., among the top ten in their fields); male; and geographically "near" to the archive from which data sets are purchased. After subjecting his data to statistical tests, White finds that only sex is a predictor and that although those close to IDL&RS buy more, the reverse is true for buying patterns at Roper.

### Implications for Libraries

The major implication for librarians is that human-readable data sets are valued by social scientists in their own right (White demonstrates that information about data is as often purchased as the data themselves) and their purchase could be delegated to libraries. The question of the prestige attached to codebook purchases, as indicated by White's exploration of who buys, is a fascinating insight into the sociology of the social sciences, and White has brought out a unique factor in acquisitions policy that libraries might well heed.

Reasons for putting human-readable tools, such as codebooks, in academic libraries include 1) introducing them into institutions that are more numerous and visible than existing data archives; 2) assuring that the tools are accessible to all disciplines and professional schools on campus; 3) deploying them with the other bibliographic tools and substantive literature available to social scientists; 4) improving bibliographical information on codebooks by bringing catalogue cards on them together with cards on associated monographs; and, 5) improving bibliographical information on data access tools generally by making them discoverable through the local catalogue, and possibly through union catalogues up to the national level.

The final chapter of White's dissertation, "Acquisition of Data Access Tools," concentrates on guidelines for librarians interested in buying data access tools. Appendices include correlation matrixes describing buyers at Roper and IDL&RS and a substantial bibliography on data archives and librarians.

Writing this review from the academic librarian's viewpoint, one is impressed with the wealth of information marshalled to provide a case for library purchase of human-readable access tools. If all areas of library buying had as much organized data as Dr. White provides in this thesis, libraries would move from the realm of somewhat subjective buying to a more scientific policy. Thus White's conclusions act at once as a model for acquisitions policy at a universal level, as well as in the specific case he addresses.

Data archivists, especially those working in isolation from traditional libraries, will want to consider the implications of White's work for their field. Greater acceptance by libraries of access tools to data sets will generate a broader base of potential archive users and, since White argues cogently for the continuing housing of the actual data sets in archives, there is no reason to fear that the very specialized services of archives will be subsumed by libraries without staff expertise to promote their exploitation. We hope that Dr. White will consolidate his groundbreaking findings into articles to be disseminated in the library and data archive press. The thesis itself is a mandatory purchase for library schools and a critical accumulation of data for archivists. We hope that Dr. White's thesis is the first of many dissertations which will explore in greater depth the problems and characteristics of archives and their users.

#### QUANTUM COMPLETES SURVEY

Quantum members in Germany have completed a survey of completed, ongoing, and planned research projects in quantitative history. This survey has just been published as Volume I in a new series of Klett Verlag, Stuttgart, the HSF, which will be concerned with quantitative social scientific analysis of historical and process-produced data. The book entitled, The QUANTUM DOCUMENTATION, is available from Ernst Klett Verlag, Rotebühlstrasse 77, Postfach 809, D-7000, Stuttgart 1, at a cost of DM 39. The full bibliographic citation is:

W. Bick, P. J. Müller, H. Reinke. QUANTUM DOCUMENTATION: Quantitative historische Forschung 1977/Quantitative History 1977. Stuttgart: Ernst Klett Verlag, 1977.

#### NEW ORGANIZATIONS / REORGANIZATION

#### BASS HOSTS CESSDA MEETING TO CREATE IF-DO

Belgian Archives for the Social Sciences Hosts Committee  
of European Social Science Data Archives to Create  
International Federation of Data Organizations

On the invitation of the Committee of European Social Science Data Archives (CESSDA), representatives of data organizations from the United States, Canada and Europe gathered together at Louvain-la-Neuve, Belgium on the 20th and 21st May, 1977 to discuss and formulate further means of mutual cooperation in the area of social sciences data archiving services.