A NETWORK OF DATA ARCHIVES FOR THE BEHAVIORAL SCIENCES*

BY PHILIP E. CONVERSE

The rapid development of behavioral science research since World War II has produced a great accumulation of data and raised many problems about their usefulness and preservation for further research. One consequence has been a major cooperative effort to solve these problems and make the data available for secondary analyses.

Philip E. Converse is Senior Study Director at the Survey Research Center at the University of Michigan.

IN THE past decade there has been a marked growth of interest in the development of archives of behavioral science data cast in machine-manipulable form (punched cards or tape). This interest has several sources. It is stimulated, for example, by the rapid increase in the sheer volume of systematic information being collected in one form or another from historically significant populations. Formal government censuses have been increasing in regularity and scope, and an increasing number of governments are conducting them. The development of sample survey techniques has led to a wealth of new research institutions, and survey data are becoming increasingly available from all continents. And, too, a growing cadre of behavioral scientists is now engaged in systematizing and refining other kinds of information from the public record (such as the social characteristics of various elites) for immediate research purposes, yet is creating in the process bodies of information of great potential utility to other research ventures as well.

Another stimulus has come from the heightening interest in comparative cross-national research. Effective primary research of this scope is extremely expensive and time-consuming, putting a special premium upon its continued exploitation for secondary analysis. While this fact is apparent to all researchers interested in such work,

up until very recently there have been no mechanisms for the systematic assembly of comparable cross-national data.

Perhaps most important as a stimulus has been the amazing expansion of data-processing capacities in the past decade. The growth of social bookkeeping and conventional libraries in the nineteenth century produced a wealth of information, but even after the advent of more advanced techniques of statistical analysis, this wealth was a glut, for the task of reducing tomes of tabular material analytically by hand was virtually prohibitive. All this is now changed: machines can in seconds or minutes perform feats of data refinement and digestion which, for the man-years they would have consumed, would never have been attempted by the most devoted scholar in 1945.

Among the remaining bottlenecks to a more efficient harnessing of information production, perhaps the principal one is the current lag in the social organization of the research community, which leaves us with little in the way of institutional bases for the orderly accumulation of behavioral science data and for the broad facilitation of access to such material. One example of many that might be cited will serve to illustrate the point. For several decades historians and political scientists in this country and abroad have labored over aggregate voting records, painstakingly locating dispersed sources and hand-compiling data for this or that sequence of elections in this or that portion of the country. It was inevitable that duplication of effort was large, although given the medium in which researchers compiled their information, there was little remedy for the problem save for the occasional publication of some of the grosser returns. However, the great flexibilities of duplication and transmission of information permitted by the new technology can now justify the systematic archival accumulation of voting statistics in permanent and machine-manipulable form, on an expanding time base and in growing geographical depth (both cross-nationally and intra-nationally down to smaller and smaller civil subdivisions). Indeed, the Social Science Research Council has expressed interest in such a data-accumulation effort, and has granted funds to Walter Dean Burnham, now of Haverford College, to collect certain gross returns back to 1824, and to assess costs involved in retrieving and organizing the American materials in still greater depth. Yet one of the problems faced by any agency in promoting a general-purpose collection of this sort has been the absence of any obvious place to locate the materials once they have been organized. Naturally, there would be no dearth of institutions welcoming the gift of such a collection. But there have been no clearly appropriate institutions already "tooled up" to promote easy access to such materials on the part of any

interested member of the behavioral science research community. This is the organizational bottleneck and the problem of effective data archives.

This article is intended to reflect the thoughts and activities of one group that has recently been working on the problem. In June of 1962 the Inter-university Consortium for Political Research was constituted with a membership of twenty-one major American universities (now thirty-eight), and goals of speeding behavioral research in the policy sciences through intensified training and archival developments.[1] In connection with the launching of a Consortium repository, it became apparent that there was some danger of duplication of efforts with nascent repositories at other institutions. In order to consider the possibilities of increased coordination of efforts, the Consortium sponsored meetings of an *ad hoc* "Committee of Eight" during the winter of 1962-1963.[2] The Committee included both members of the Consortium Council and representatives of already-established data libraries that had indicated an interest in broader collaboration. Consensus was reached on a number of initial policy directions, and the Committee drafted a proposal for funds to support an eighteen-month period of exploration during which contacts might be broadened in the reasearch community and a viable organizational and technical design for inter-repository coordination might be drawn up and ratified. Such a design would form the core of a more ambitious subsequent proposal for funds to support the archive network.

In December 1963 the initial exploratory proposal was granted funds by the National Science Foundation. By common consent, the Survey Research Center of the University of California (Berkeley), under the direction of Charles Y. Glock, was the formal recipient of the grant.

The purpose of this article is to aid in informing the research community about the activities of the Committee; to invite a broader base of discussion of policy alternatives; and to solicit the endorsements of the utility of the enterprise that will be necessary if ultimate operating support is to be achieved.

THE FORM OF ARCHIVAL DEVELOPMENT: CENTRALIZED OR DECENTRALIZED?

One of the primary points of data archives is to draw information into centralized stores so as to reduce the diffuse search time that

---

[1] For a more detailed statement of the Consortium structure and goals, see Warren E. Miller, "The Inter-university Consortium for Political Research," *The American Behavioral Scientist*, Vol. 7, 1963, pp. 11-12.

[2] Persons involved in meetings at one time or another are noted in the footnote at the beginning of this article. Befitting a social science gathering, the number "eight" represented a rather imprecise measurement.

investigators must now spend to assemble data. This goal of centralization, drawn to its logical extreme, leads naturally to the vision of a single national storehouse, completely catholic in its coverage of social science data. Indeed, it is just such an "automated" counterpart to the Library of Congress that is most frequently envisaged in pleas for repository development in this country.[3] And an example set by Soviet Russia has actually led, in the past year, to Congressional hearings on the possibility of a major national data center, although oriented primarily toward the natural sciences.

In the light of such possibilities, the activities of the Committee of Eight may appear to imply a judgment that a decentralized system or network of repositories is preferable to a single national repository. In point of fact, this is not the case, and some of the members feel that in the long run a single central archive will prove to be the most viable model.

The Committee's position is instead a pragmatic one. Even with reasonably good progress, it seems unlikely that a major national data center for the natural sciences will be a reality for several years. It is only realistic to double the probable delay where the social sciences are concerned. In other words, we may account ourselves lucky if anything substantial happens with respect to a major, nationally subsidized social science data facility within the next decade.

In direct contrast, what *has* already begun to come about is the establishment of a number of more modest data banks here and there in the United States and Europe.[4] All these facilities are deliberately limited in their conception along one dimension or another, be it projected size, the research clientele to be served, the geographical or topical scope of their holdings, or the character of the data they are designed to acquire (survey data as opposed to census data, for example). Typically, of course, they are limited in most of these directions at once. There do not seem to be more than three facilities at the moment in the United States that have both fair topical breadth

---

[3] The most thorough early discussion of the subject was York Lucci and Stein Rokkan, "A Library Center of Survey Research Data," New York, Columbia University School of Library Service, June 1957, unpublished manuscript.

For a more recent treatment, see Myron J. Lefcowitz and Robert M. O'Shea, "A Proposal to Establish a National Archives for Social Science Survey Data," *The American Behavioral Scientist*, Vol. 6, March 1963, p. 27.

[4] The Roper Center at Williams College was the first major development of its kind. For a description, see Philip K. Hastings, "The Roper Center: An International Archive of Sample Survey Data," *Public Opinion Quarterly*, Vol. 27, 1963, pp. 591-598, and "International Survey Library Association of the Roper Public Opinion Research Center," pp. 331-333 of this issue. More recently, major data holdings have been organized at Yale University, the University of California (Berkeley), and the University of Michigan. The first repository in Europe was the *Zentralarchiv* at the University of Cologne. Further data holdings are in various

and a deliberate capacity, already operational, to serve researchers at any distance from their own sites.[5]

Despite their numerous limitations, the "natural" growth of such installations is encouraging, for they are a witness to the prevalence of a felt need, and can be seen as the necessary prototypes of the kinds of facilties that must one day become available. Their growth is also forcing a wider circle of investigators to serious consideration of some of the more technical problems of large-scale information processing, storage, and retrieval.

Aspects of these facilities suggest that there are at least some virtues in the current *de facto* decentralization of repository development. These are most notable with respect to the data-acquiring or "eyes-and-ears" function of an effective repository. At the moment, facilities housing specialized types of data are developing at sites where there are concentrations of relevant specialists. That is, banks of sample survey data are organized most typically by experienced sample survey specialists. The systematic file of information on all the world's cities accumulating within the International Population and Urban Research Program at the University of California (Berkeley) has developed under Kingsley Davis and a staff of specialists in urban demography.

The benefits of the closest possible working link between a corps of specialists and the development of a specialized data bank are several and quite obvious. The specialized facility is more likely to know of the existence of important bodies of data relevant to its specialties. Its personnel are best equipped to make judgments as to priorities in data acquisition, as well as to necessary quality controls on new data, and are in a position to provide good ancillary information on error margins. Finally, systematic quantitative data tend to yield a variety of second-order statistics (compounded indices, residuals from regression analyses, and the like) that are often of high interest in themselves to subsequent investigators. These products would tend to become an integral part of the data holdings most easily where the specialized library grows up in connection with an

---

stages of discussion or establishment in England, France, Holland, and Norway. Dr. Stein Rokkan of the Christian Michelsen Institute (Bergen, Norway) has, with the aid of UNESCO and the International Social Science Council, been particularly instrumental in developing and maintaining repository ferment in Europe as well as the United States. See Stein Rokkan, "The Development of Cross-national Comparative Research: A Review of Current Problems and Possibilities," *Information* (Bulletin of the International Social Science Council), Vol. 1, No. 3, pp. 21-38.

[5] It may be convenient for us to reserve the term "data repository" for these more general-purpose facilities, while using the term "data library" for the installation with a narrowly specialized collection organized primarily to service a local faculty or research staff.

active research program. Since concentrations of relevant specialists are necessarily scattered in pockets across the country, the decentralized development of data libraries has a great deal to recommend it.

Once beyond the data-assembly functions of a repository, however, the dysfunctions of decentralization begin to outweigh benefits. From the point of view of the individual user, a diffuse scatter of specialized data banks improves very little on the current data-search situation, except where the definition of the research coincides with one of the specialized holdings. Since data-monitoring functions are often best organized along lines of geographical jurisdiction, and many data libraries are already proceeding on a defined "area" base, decentralization and the consequent cross-repository search it requires will weigh most heavily on comparative, cross-cultural research.

Decentralization is wasteful as well in terms of the collective economics of archival development. Haphazard overlapping of topical areas from repository to repository is already leading to such wastes as duplication of effort, burdens placed on data-supplying outlets through multiple requests for their materials, and the like. Furthermore, communication has been sufficiently poor between data libraries that opportunities pass unnoticed to fill in a collection at one repository in a way that would establish a comparability (say, for another area of the world) between its holdings and those already in hand at another repository.

One of the most important dangers of the current decentralization has to do with the idiosyncratic languages that each repository may come to speak with respect to such things as schemes for data classification, the format of variables and descriptive materials, desirable computer hardware, and retrieval routines. This problem is particularly frustrating, for the situation has not yet gone beyond the point of no return. Most of the libraries are still making these decisions. Some of the decisions are large or have clear cost-efficiency answers; most are small and leave great latitude for arbitrary or accidental choice. Yet once this initial structure of decisions has been completed, the institution is thereby "baked into" a very ponderous cake of custom that is expensive to redo, quite apart from common human resistance.

It may be argued that these and other concomitant hazards spring less from decentralization itself than from a development that is not only decentralized but *uncoordinated* as well. We believe that the intrinsic merits of decentralization may be preserved, at least during the early years of these efforts, without all the ills of haphazard growth, if steps are now taken toward effective coordination. Developing the means for such coordination is the goal of the Committee of Eight. It would not be worth the effort even to address this goal if

the various repositories and libraries were bent upon going their own way. Almost without exception, however, the major developing repositories and those of the smaller libraries thus far contacted have expressed not only willingness but enthusiasm about collaboration. The economic savings in a pooling of resources toward collective technical solutions are too self-evident to miss, as is the desirability of establishing compatible links between one's own specalized collections of data and other adjacent materials accumulated elsewhere.

Hence our efforts are not intended to prejudge the relative merits of decentralized repositories as opposed to a major national repository. Instead, it is our belief that if experience makes clear in the next few years that complete centralization is a more feasible alternative, the coordination of existing repositories will have paved the way for such a step by stimulating the volume of use of these facilities (the presence and nature of demand is difficult to convey in the abstract to more remote decision makers), and by guiding collections to a common form in which they could be coalesced very economically at a later date.

KEY OPERATING PRINCIPLES OF A REPOSITORY NETWORK

In considering the problems of coordinating repositories, most of our discussions have returned again and again to the absolute necessity of providing means of facilitating access to any holdings of any repository in the network, and presumably to any lying outside as well (e.g. abroad).

The notion of "access" has deceptively many facets. A decade ago, the key question was whether or not the person or agency that had gathered or systematized the information would be willing to donate it to the public domain after the initial period of private exploitation of the materials was completed. Under an older set of mores, any expectation that subsequent investigators might be able to use privately gained information for later reference often was considered presumptuous. This age is largely past within the academic research community, although access to materials often remains restricted where the data are gathered under commercial or certain kinds of government contracts.

Nonetheless, there is a wide gulf between the "access in principle" that pertains when the originator of the data is not reluctant to open them to unrestricted use and access in a much more practical sense of the word. Genuine access presumes that any far-flung investigator who might have reason to use particular data could be expected to know both *that* they exist and *where* they exist, and presumes as well that he can receive necessary descriptive information and the data them-

selves at minimal cost and time delay. It presumes, finally, that the data are his to keep and to duplicate for assistants or other persons as the situation may warrant. In other words, the goal of genuine access in a practical sense refers to the day when an investigator at one of the far-flung California campuses can, in an afternoon's work at a satellite of his local computer facility, learn just what data exist in the total network bearing on the hypothesis he wishes to check out, order by telecommunication either a statistical analysis on those data from an East Coast repository (or the raw data themselves if he has more extended use in mind), and have his output in time for dinner, all at minimal immediate cost.

It goes without saying that a certain amount of unfinished business lies between our current situation and this ideal. However, we are impressed by the degree to which the purely technological problems —those not incumbent upon social scientists to work out for themselves—are already solved. Indeed, social scientists have scarcely begun to capitalize upon this technology, for it is perhaps a bit difficult psychologically to adjust to the admirable flexibilities of information cast in this form. Compared with books as a medium, for example, the new information is not only exquisitely manipulable, but duplicable and transmissible as well.

Thus most of our unfinished business is either economic or organizational, or some blend of the two. On the economic side, access of the ideal sort described above comes with a substantial price tag, a fair portion of which must be borne by collective subsidy if any very broad access is to be achieved. Probably more important, however, is the unfinished business in the development of the appropriate social organization to harness the new potentials for the common weal.

We might note that such an access problem is not peculiar to a decentralized network of repositories. It is unlikely that there will be any great enthusiasm about a single massive national data center located at Washington or Chicago on the part of researchers in California, Texas, or Florida if the holdings can be used effectively only through a personal visit to the site. In the long run, it will be infinitely more pleasant, not to mention more economical, to shuttle information rather than researchers around the continent. In the short run, until the access problem is more fully solved, decentralized repositories (as opposed to a single central one) have all the more to recommend them, provided these repositories can at least "talk" to one another through rapid interchanges of data.

However, concern over practical access remains of fundamental importance to any plans for a decentralized but coordinated network. Part of this importance is quite frankly political. As long as any re-

searcher or institution is grossly penalized in efficient use of holdings as a simple result of physical distance from the site, there is likely to be fierce competition over the location of repositories and who gets to store what. Furthermore, most large granting agencies with the kind of resources necessary to support major archives are chary of grants that would grossly advantage one set of major institutions over another, and quite rightly so. Except as there is promise of progressive solution to the problem of access-at-a-distance, such agencies are reluctant to give much support at all.

In the measure that the access problem is solved, having any particular body of data at a local repository becomes less and less of a plum. Indeed, the possibility of a major local repository responsible for servicing a wide geographic clientele may properly come to be thought of more as a chore than a prize, for the operation of a service repository is something less than romantic. It requires, in the first instance, a concentration of computer and attendant human resources that only a limited set of institutions in the country enjoy. And it imposes fiscal and administrative responsibilities that many of these institutions might wisely eschew if they can "let George do it" without loss of significant research capability.

Therefore, as a prime operating principle, we have become deeply committed to the early and heavy allocation of resources toward solving some of the economic and organizational problems that currently make genuine practical access at a distance impossible. This commitment is a rather rigid one, for so much depends on it. The remainder of this article is devoted to other policy directions which now seem feasible in the light of the practical situation confronting us, yet which are subject to modification as the situation changes.

THE DIVISION OF LABOR

One of the first policy questions for such an enterprise has to do with the content scope of the total holdings within the operating network. At this point, we see no reason for establishing in principle any definitional boundaries short of the universe of data that have some interest for behavioral scientists. In practice, of course, this is a totally unmanageable order even for the long run, and tactically we would expect to concern ourselves with cultivating the small acorns on hand rather than being alarmed about the prospective size of the oaks.

This tactic has many specific applications. We may imagine a multidimensional space or grid of cells within which any particular body of data may be located in such terms as topical content, geographic area represented, nature of population of reference (e.g. elite vs. mass), nature of the methodology generating the data (e.g. enumera-

tion, sample survey, historical content analysis, etc.), degree of aggre-
gation of the data (observations on the individual, as opposed to
parameters of collectivities), and the like. The partitions between the
cells correspond to the kinds of lines which up to now have typically
been chosen for self-limitation by nascent repositories and libraries.
They form as well reasonable lines for the division of labor in any
network, since they represent the kinds of substantive and methodo-
logical expertise available locally. This does not mean that we would
envision a repository for each possible cell. Whereas some small and
highly specialized local library might fill only a single cell, any reposi-
tory large enough to take on the burdens of servicing an outside
clientele would normally accept as a jurisdiction a much larger "row"
or "block" of cells (e.g. sample survey data of *all* kinds of content
for sub-Saharan Africa) or even, perhaps, some gerrymandering of the
space.

In the near future, at least, we would feel no necessity to allot all
of such a space to one repository or another. Instead, the first steps
toward a division of labor will be no more ambitious than the ascer-
tainment of what portion of the space each of the current major
repositories will accept as a responsibility, along with the necessary
adjudication to solve any overlapping claims. Such an initial effort
will leave very many empty cells along every dimension save probably
the geographical. This, as it seems to us, is no cause for concern. We
feel that growth of holdings must be dictated by some joint function
of research demand on the one hand and acquisition costs on the
other. In this light, unassigned cells or blocks of cells signify nothing
more alarming than exorbitant information costs or a research demand
as yet insufficiently crystallized to warrant investment.

Initially unassigned jurisdictions will also reflect the fact that there
is room for the network to grow, and that it is in no sense a closed
corporation. Whether initially unassigned space is later undertaken
by an existing repository or becomes part of a block of cells accepted
by a new repository will depend on the immediate situation. None-
theless, from a starting point of three or four major repositories well-
distributed geographically, some expansion is to be expected.

It would seem likely, furthermore, that close working relationships
would tend to develop between major repositories and institutions or
faculties wishing to maintain smaller and more specialized data li-
braries in connection with local research and graduate programs. In
such cases, it could be mutually beneficial for the data library to
assume the "eyes and ears" or data-assembly function for the larger
repository within its chosen cell. By feeding duplicates of its data
collections to the repository as they are organized, it could build its

own library and contribute to the larger enterprise at the same time, without becoming obliged to establish the administrative and technical plant necessary to service continuing requests from a large outside clientele.

COORDINATION AND REPOSITORY AUTONOMY

At this point the strength or degree of coordination that will come to seem desirable between the repositories is not at all clear. On the one hand, some collaboration would seem essential. A minimal level of collaboration would probably entail (1) joint work on some of the technical problems of data format and retrieval procedures, leading toward increasingly compatible holdings; (2) a routinely high level of communication between repositories as to existing holdings and future plans; (3) acceptance of major jurisdictional obligations; (4) joint search for ways to speed up interrepository data transmission; and (5) some central secretariat or clearinghouse equipped with at least gross information of specific holdings at each major repository.

On the other hand, it is neither expected nor desirable that participating repositories should give up significant portions of their "private lives" as autonomous facilities. Those repositories now operational have grown up with their own individual financial bases and their own executive bodies, in response to unique constellations of demands that they will naturally continue to serve. The primary change that membership in the network will mean is that some functions—usually ones already performed—will be voluntarily taken on in a more formal way as obligations to the collectivity, and that, pursuant to the securing of collective funds, the repository will collaborate in a joint assault on problems of distribution, access, and the like. This in turn means, for example, that if a repository wishes to store large amounts of data that are by collective definition part of the jurisdiction assigned another repository, it will as always have every right to do so. Part of the challenge to the collectivity will be to make sufficient progress on the facilitation of interrepository transmission of data that such overlapping storage will come to be increasingly pointless in every light, so that repositories will be content to store no more than the materials for which they have primary responsibility.

It is likely that the particular balance struck at any time between the "private lives" of the repositories and the portion of their efforts that are direct responses to mutual decision of the collectivity as a whole will depend on the ratio of private to collectively secured funds. The latter kinds of funds would be used, among other things, to "even out" the technical progress and substantive comparability of data holdings across repositories. Thus, for example, one repository

on its own funds may have allocated a great deal of effort to the sheer acquisition of data and relatively little to the systematic assembly and cataloguing of materials that increase their accessibility. Another repository may have reversed these emphases, having narrow but highly accessible holdings. Collective funds could be earmarked in such a way as to induce the first repository to intensify its data-organization work and the second to intensify its rate of acquisition. Similarly, collective funds could be channeled to aid in the acquisition of a relatively expensive body of information at one repository to provide a match for some comparable body of data of particular theoretical significance from another part of the world that is already stored in another repository.

Meanwhile, the private research of individual investigators would go on as before, save at those points where it would be positively facilitated. That is, an investigator who wishes to conduct a major piece of primary research in some topical area covered by a distant repository would naturally do so. In the initial stages, he would be helped by access to more systematic information on prior work in the area, and could readily provide himself with background data in whatever depth he desired, through the relevant repository. After he considered his primary research completed, it would be customary for him to contribute duplicates of his data to the appropriate repository. Undoubtedly, over time, data-gathering agents would be willing to pay increasing attention to suggestions from the repositories as to coding conventions and formats for various types of data which would make the data as multipurpose as possible and which would, through their prior standardization, greatly reduce the cost to the repository of assimilating them within its holdings.

PARALLEL DEVELOPMENTS IN EUROPE

As intimated above, parallel ferment toward the establishment of data repositories is occurring in Western Europe, with aid from organizations such as UNESCO and the International Social Science Council. In the European setting, the development of a multiplicity of major repositories, each covering a national or a linguistic area, is more of a foregone conclusion than it need be for the United States. At the same time, it is likely that these several repositories will grow up in close collaboration with one another, so that the end product will be an integrated network of repositories very similar in structure and intent to what we have sketched above.

Fortunately, close contact is being maintained between the American and the European developments. As we write, there is every reason to hope that the two networks will take shape with compatible

data currencies and easy access to one another's holdings. This does not mean that American repositories will not want to store substantial quantities of European data, and vice versa. However, it is unlikely that either side would wish to store information relevant to the other in the same intensive detail that may be sought for native research materials; and as data interchanges are facilitated, the point in doing so will be reduced as well. Thus the basic costs of data assembly for European materials will become less on our side, and it is likely that, in time, other divisions of labor with respect to information concerning other areas of the world may be achieved as well.

A LONG-RANGE PERSPECTIVE

It is quite obvious that limits on the rate at which repositories can assimilate new bodies of information mean that, against the total universe of significant behavioral science information, network holdings in the first few years are likely to seem both puny and fragmentary. Yet the holdings can be expected to increase steadily, and with this trend the possibilities for fruitful secondary analyses based on archival data should increase exponentially.

At the same time, the sheer bulk of the holdings will pose deepening problems of financial support and adequate service personnel. At the close of its eighteen-month period of exploration, the Committee of Eight assumes that a collectively endorsed request for major funds will be presented to some agency to support the venture over an extended period of time (at least for three years, perhaps for as much as ten). Yet it does not seem reasonable to expect to proceed over any very long period of time, working from special grant to special grant. Once the enterprise has had time to build its holdings and demonstrate satisfactory rates of use, it will be important to switch from "innovative capital" to more stable bases of subsidization.

Similarly, over the initial periods of archival development, social scientists will be obliged to devote major amounts of administrative energy to its nurture. Judging by all examples to date, they will do so only begrudgingly. And, of course, the enterprise in question is nothing more than a library, albeit one in rather unconventional dress.

While some of the earliest agitation for data archives proceeded on the assumption that it would take only a little prodding to touch off interest on the part of conventional librarians in the new horizons represented by archives of machine-manipulable data, up to now we have found little resonance even from advanced "research libraries" that enjoy some passing acquaintance with the use of data-processing machines for documentation searches and classification routines. Nonetheless, it may well be to conventional libraries that our path sooner

or later must lead. Both as a means of establishing claim to stable subsidization and as a means of turning all or much of the operating weight of such an enterprise over to relevant professionals, it will be important in the developmental years to cultivate the interest of personnel trained in library service. Ultimately, the definition of a research library must be expanded to keep within hailing distance of research needs.

<div align="center">CONCLUSION</div>

We have used this opportunity to publicize steps now being taken both to stimulate the development of repositories for behavioral science data and to put them on a more coherent footing. We welcome discussion and criticism. We would be particularly appreciative of any information concerning actual or projected data libraries that could be relevant to such an enterprise. Finally, readers sympathetic with the goals of the repository enterprise could make a signal contribution by forwarding a note expressing that sentiment: progress in these areas may well come to hinge on testimonials to potential demand, and we would appreciate endorsements we might use toward this end.

Please address all communications to Professor C. Y. Glock, Director, Survey Research Center, University of California, Berkeley 4, California.