ILLINOIS
AGRICULTURAL ECONOMICS
STAFF PAPER

SOME FUNDAMENTAL ISSUES IN THE DEVELOPMENT

OF COMMUNITY INFORMATION SYSTEMS

J. C. van Es

January 1978 · No. 78 S-3

Department of Agricultural Economics
University of Illinois at Urbana-Champaign
305 Mumford Hall, Urbana, IL 61801

SOME FUNDAMENTAL ISSUES IN THE DEVELOPMENT
OF COMMUNITY INFORMATION SYSTEMS

J. C. van Es

Introduction

Local officials, planners and change agents increasingly express the
need for local data. The need for data results from different pressures
on local leaders. Data are essential to social planning and social action.
This need for data is enhanced by the fact that many persons have drawn
the imperfect corollary that, if data are useful, more data must be more
useful. The multifaceted intrusion of large scale bureaucracies in the
local community such as Foundations or Agencies of State and Federal
government, have also created pressures to document requests for sup-
port or to provide evidence to show the ways in which local activities
meet community needs or national goals. Frequently the bureaucracies demand
that local data are provided according to a standardized format.

Collecting local data, especially in standardized format, is a difficult,
costly, and time consuming activity. Even when those data already have been
collected by others, the information usually is not available in an appro-
priate form to local users. With the development of computers it has become
feasible to centralize the collation   retrieval of large quantities of data.
At the same time such centers can maintain the contacts with data gathering
units (Census Bureau, State Government agencies, etc.) to allow for the con-
tinuing acquisition of large quantities of data. Computer technology allows
for fast retrieval of information and more advanced interactive networks
promise to allow for centralized data acquisition and storage to be combined
with decentralized use and interpretation of the data. In the remainder
of these pages I will look in some detail at the peculiar nature

of the practical difficulties that have surfaced in the development of these systems. In large part these problems are related to the conceptual basis for the operation of social data information systems. Many of us have looked to the social indicator movement to provide us with a useful conceptual basis. Given this background it may be useful to briefly acknowledge our roots and to review the social indicators phenomenon.

Social Indicators

Recently I received the following promotional message in the mail from the Smithsonian Science Information Exchange, Inc.:

Probably no area of social science research has experienced as large a recent growth as the "social indicators movement." There was a national need for a suitable social accounting framework, analogous to economic accounting which took dollars as a consistent unit of analysis. Social indicators are noneconomic measures of social conditions important to the social state of a nation. The potential of the use of social indicators in the formulation of public policy and the promotion of human welfare is recognized by policy-makers in government and the private sector.

The statement appears to capture well the expectations of organizers and funding sources for much of the information systems activity that we will be discussing here.

In their most general form social indicators are measures of social conditions (Land, 1975). However, little unanimity exists about what social conditions are, and even less unanimity on which of those social conditions should be measured. The latter question was initially (HEW, 1966) answered in terms of the policy utility of the social indicators for (1) the evaluation of specific policies and programs; (2) developing a balance sheet of social accounts, and (3) for setting policy goals and developing priorities.

Following the critiques on this approach (Sheldon and Freeman, 1970; Sheldon and Parke, 1975; Land, 1975), the social indicators movement has increasingly become more timid about attempting to attain these objectives.

The critics have challenged the immediate policy utility of social indicator research mainly on two grounds. In the first place, social indicator work has not been able to deal with the normative nature of policy evaluation. What are good or desirable policy outcomes for one set of eyes may be equated with calamity or perversion by another observer. Furthermore, these critics argue, at this point in time our understanding of the social system is so inadequate that we have neither the capability to design "social balance sheets," nor the ability to isolate the impacts of programs and policies.

Currently social indicator research appears to be heading in the direction of descriptive analysis (Stone, 1973). At a minimum these analyses have established descriptive information on current conditions, while other researchers have further attempted to disaggregate the variance in social measurements. In essence these latter efforts are analyses of social change. In some instances (e.g., Pampel, 1977) these attempts become both quite systematic, and problem-oriented. However, the selection of the variables in those models is no different from the selection of variables in other analyses of social change: a peculiar mixture of theoretical guidance, data availability, and researcher cunning.

At this point in time, social indicators research appears to be a convenient label that is attached to research on aggregate level, noneconomic observations in society. Some of it is of high quality, some of it is not. Some of it appears to have utility to policy makers, although for much of the social indicator research the policy utility is only a promise. However, the initial excitement about social indicator research, combined with a time lag in the diffusion of the more sobering experience of the actual work, has created widespread expectations concerning the power of data systems. Many of us working in the area of community data systems have probably tried to

ride the wave of popularity of social indicators in order to get out activi-
ties off the ground. We may now be facing, increasingly, a period of dis-
appointment and cynicism about what we are able to do. It is therefore time
to take a closer look at some of the practical issues involved in providing
data for policy use in local communities. Specifically we will deal with
the following issues: different data users and their needs; the appropriate
level of technology; and the nature of the available data.

Needs of Data Users

At the University of Illinois we envisioned from the beginning two
classes of data user: researchers and local users. We anticipated that the
interests of these two groups would be different but compatible, if not com-
plementary. Researchers, as we define them, are generally characterized by
their need for data for many different units and by their desire to put
different types of variables into an explicit conceptual framework. Local
users are those users which have a particularistic interest in data about a
small number of communities or counties. These data are needed in order to
help solve a particular problem. Among local users we find personnel of
consulting firms, local planners, and professional change agents, such as
Extension Service personnel, as well as local citizen groups.

Researchers are not necessarily well acquainted with the nature and
existence of a wide variety of data, but they frequently have specific areas
of expertise and they are usually in a position to provide advice and feedback
on the contents of the data system. The researchers we have dealt with,
furthermore, almost without exception either were able to manipulate and
analyze the data themselves, or had access to people who could perform these
tasks for them.

Because of the nature of most academic research, researchers tend to be interested in generalizable relationships between variables. Consequently their analyses do not have the objective of reporting the particularistic information of great interest to local change agents and decision makers. However, we have found that frequently in the course of these research analyses intermediate results are obtained that are of specific interest to local persons. In other instances such information may be easily computed as part of a particular research project. Without a priori judging the utility of a researcher's analysis for local data users, we feel that the strategy of providing explicit information for local users has proven most successful in terms of making research complementary with local users' data needs.

Local users tend to share with researchers a concern about accuracy and currency of local data, two features over which the information system, unfortunately, exercises little control. A more difficult problem to deal with is the persistent need of local users for very detailed information on local conditions. While local users may be interested in comparing "their county" with others, for diagnostic and policy purposes local users often need data for small local breakdowns, such as townships, city blocks, urban fringe, etc. In most instances this data is simply not collected and reported for such units. Furthermore, for a state such as Illinois with 102 counties, more than 1240 incorporated places and ten SMSA's, there are real limitations on centralized handling of data on very small political or geographical units. Especially a centralized broad-based data system, that does not specialize in a small set of data, will find it difficult to cope with very detailed data unit breakdowns.

Local users also tend to be intermittent users. For most local users the need for and use of data relates to a particular activity they are engaged in. Grant applications, annual reports, and particular projects create the

need for certain data. To most users the search for secondary data
is a very minor part of their activities, and for most not a particularly
rewarding one either. That is one of the reasons for centralizing informa-
tion systems. However, for most local users data usage is so infrequent that
little "on the job" learning takes place. Since most local users have no
professional expertise in data handling and interpretation, the absence of
informal learning becomes a real hindrance to the further growth of local
usage.

Sporadic usage and its consequences appear to be especially prevalent for
rural counties and small towns, as well as nonprofessional users such as
small town political office holders and citizens groups. Centralized informa-
tion systems face a serious challenge of dealing with intermittent, nonpro-
fessional users if they want to avoid being mainly a support activity to
consulting firms and metropolitan professional planning and change agencies.

One solution to this problem lies in education, but it is not at all
clear what form such education is to take. In order to be useful, general
teaching about data usage would require more investment in time and effort
than most users will commit to it. More specific educational material aimed
at certain types of data appears to be ignored by most users unless they have
a need for that information at that time. At Illinois we deal with this situa-
tion essentially through over-the-phone consultation. This is time consuming,
has a very modest learning effect, and, more seriously, reaches only those
people who already actively seek information from us. Finding a way to help
the infrequent data user transform his ill-defined need for data into a re-
quest to an information system, will have to be increasingly a more serious
concern of all information systems.

The Appropriate Level of Technology

Our experience at the University of Illinois has lead us to the conviction that at this point in time we are best served with a technologically simple system. More precisely, we have rejected the development of an interactive network which would give decentralized users direct access to our data system. We have made these decisions on two grounds: 1) the nature of the users we try to serve, and 2) the cost involved in an interactive network.

I already discussed the fact that many of the users of our system have very little expertise in terms of data handling and data analysis. For many users even defining what their data needs are is a challenging task. In addition, as I pointed out, most of the users will need access to the data only intermittently and therefore they will not develop the routine expertise to handle the system efficiently. In some cases a sufficient number of local users may be concentrated in one locale to allow a support infrastructure to develop, but in our judgment many communities will not in the foreseeable future develop the critical mass necessary for such a development.

Related to the above is the matter of expenditure. Decentralized access requires high expenditures in programming and hardware alike. These expenditures vary from system to system, of course. Software and support services are very much a function of the scope of the program to be offered. Highly specific limited data systems obviously do not need the same kind of programming support as a program with widely varying options.

The cost of hardware is even more dependent on local conditions. In situations where the users already have access to the necessary equipment, additional information system programming can be added at relatively little cost to the user. Such additional opportunities may in effect contribute to the cost effectiveness of the equipment in use.

As I pointed out before, at Illinois we have not decentralized the
access to the data for the users, at least not through the use of advanced
computer technology. We have instead adopted a system where the accessing
of the data is centralized. This allows for human interaction between the
user and the data system, which provides the user more help in approaching
the data. In addition, we are using low level technology in the form of
printed material to provide potential users with information and to maintain
contacts with them.

The Nature of the Available Data

Because of the costs involved information systems typically do not gather
their own data. In practice they collate and retrieve data which have been
gathered by others.

Figure 1 (adapted from Bonnen (1975)) indicates the nature of the data
gathering and processing that takes place. On the left side of the Figure 1
we find the data gathering process for the primary user, such as the Census
Bureau, or a State Agency. These users determine the nature of the data
gathered. If their measurement, or operationalization, are inadequate, so
will be the quality of the data that is collected. Many of these data have
well-known data quality problems. The U.S. Census underenumerates certain
population groups; the 1974 Census of Agriculture also suffered from questionnaire
return problems. Many State Agencies find it difficult to collect and process
data in a solid fashion. Finally these data are collected when the original
user decides to, so these data are frequently no longer current, or suffer from
lack of continuity.

Less tangible, perhaps, but more important are the theoretical concepts that shape the data collection. By theoretical concepts we mean the related abstract notions that help us classify the empirical world. Family and household are such concepts, but so are farm, community, and health. Much data are gathered without explicit and adequate knowledge of the theoretical concepts underlying them. In many cases data are gathered because of tradition or more elegantly "to maintain continuity." Obsolescent data are gathered which appear no longer relevant to the empirical reality. Bonnen (1975) has pointed out that some data collected on agriculture are of higher quality now, than they were in the past. At the same time significant elements of agriculture's empirical reality (e.g., vertical integration) are not adequately represented in current data for the lack of theoretical concepts included in the data gathering procedures. Flora (1974) and others have argued that in some respects the agricultural census is not just inadequate, but in effect misleading. While we have little control over these matters, I cannot help but feel that those of us who are operating information systems have not taken these problems seriously enough. In ignoring them we probably should accept some responsibility for cluttering the landscape with invalid or unreliable material.

The right side of Figure 1 indicates how the data are used by the secondary user. The broken line indicates that the secondary user does not make direct observations of the empirical world.

The content of the boxes on the left side and the right side of Figure 1 will correspond closely when the secondary user uses the data for the purposes for which the primary user collected them. Local users, as compared to researchers, are less likely to stray from the original intent of the data

collection. However, even local users use variables as "proxies" for the empirical phenomena they want to observe.

It is difficult to understand why this complex issue of "proxy" variables, "meaning transference," or "conceptual equivalence" has attracted so little systematic attention. The state of current practice borders on sheer anarchy. Recently I reviewed a paper for a scientific journal, in which the authors used one primary variable to represent six different secondary variables! In psychological measurement and educational measurement extensive systems have been developed to ascertain the validity and the relia- · bility of measures. While some of the advanced statistical techniques used in secondary data analysis are beginning to address these issues, it is clear that only minor effort has gone into developing the systematic evaluation of data quality.

Much of the effort to improve the quality of secondary use of existing data will depend on a well-articulated theoretical framework which will give researchers and users the conceptual tools to address issues of data validity and reliability.

A well-articulated theoretical framework will also be the only effective deterrent to that great seducer of secondary data users: doing the most we can with the data we have. Secondary data users have been much too glib in substituting available primary data for those needed. Great ingenuity has been devoted to rationalizing why a particular variable was a justifiable measure of a particular concept, when the only real reason appears to have been that it was the only piece of available data that even vaguely resembled the concept being measured.

The following quote from Howard Phillips may illustrate the point I am

trying to make:

> As I review literature concerning decision making within the small
> community, my concern is that rural development researchers have
> become mass secondary data manipulators. This approach is useful on
> the national or regional level, but there are still decisions made
> at the local level and we do not have much research-based information
> that recognizes this. What we are doing in using the secondary data
> manipulation approach is measuring input. We are making all our
> judgments on the basis of what we put into the system for making our
> decisions rather than on results. For example, it is well known that
> rural educational systems are inferior to every educational system.
> This judgment is based on data such as the number of books in the
> school library, not on whether anybody reads the books. Or it is based
> on the salary scale of teachers without consideration for differences
> in the cost of living in different areas. This is the kind of research
> information we have today; it is input research.... (Phillips, 1974).

I think it is fair to say that when it comes to using secondary data, we

need more careful thinking in much that we are doing, and less infatuation

with manipulating and crunching existing data.

Conclusion

The preceding paragraphs may leave you with the notion that community

information systems are beset by problems. As a matter of fact you are right,

they are. However, as any good preacher will tell you, only the faithful

give him or her the opportunity to warn about damnation in glorious detail

and, I might add, most treat the preacher's word with benign neglect! For

most of you, your attendance here indicates your interest in information

systems and I'm sure you will place my words in proper perspective. More-

over, many of my fellow speakers today and tomorrow have faced the issues I

have discussed and in many cases they will challenge the points I have made.

The need for information systems is real and the conceptual and techno-

logical developments will continue to make such systems more effective and

more available. At the same time I feel strongly that the period of research

and development of information systems will be very lengthy. I hate to be

platitudinous, but the road ahead will be long and the pitfalls are many.

Truly satisfactory systems may not develop for some time, and may have to

wait until decentralized information accessing becomes an integral part of

our society. I don't know if we should have the audacity to consider our-

selves to be pioneers in this field, but I do think we share with real pioneers

not only the frustration and sheer drudgery of the task but also the challenge

and the excitement. And if we don't experience success all that often, we

will learn to settle for the satisfaction of having tried at least.

References

Bonnen, James T.

1975 "Improving Information on Agriculture and Rural Life," *American Journal of Agricultural Economics*, 57:754-63.

Flora, Jan L.

1974 "Research on Declining Communities: The Great Void," pp. 140-44, in Larry R. Whiting (ed.), *Communities Left Behind: Alternatives for Development*. Ames, Iowa: Iowa State University Press.

H. E. W.

1966 *Toward a Social Report*. Washington, D.C.: J.S. Government Printing Office.

Land, Kenneth C.

1975 "Social Indicator Models: An Overview." pp. 5-36 in Kenneth C. Land and Seymour Spillerman (eds.), *Social Indicator Models*. New York: Russell Sage.

Pampel, Fred Carroll, Jr.

1977 "Social Indicator Models of Changes in the Status of the Aged: Labor Force, Occupation and Income, 1947-72." University of Illinois unpublished Ph.D. dissertation.

Phillips, C. Howard

1974 "Mass Secondary Data Manipulators." Pp. 139-140 in Larry R. Whiting (ed.) *Communities Left Behind: Alternatives for Development*. Ames: Iowa State University Press.

Sheldon, Eleanor B., and Howard E. Freeman

1970 "Notes on Social Indicators: Promises and Potential." *Policy Sciences* (April):97-111.

Sheldon, Eleanor B., and Robert Parke

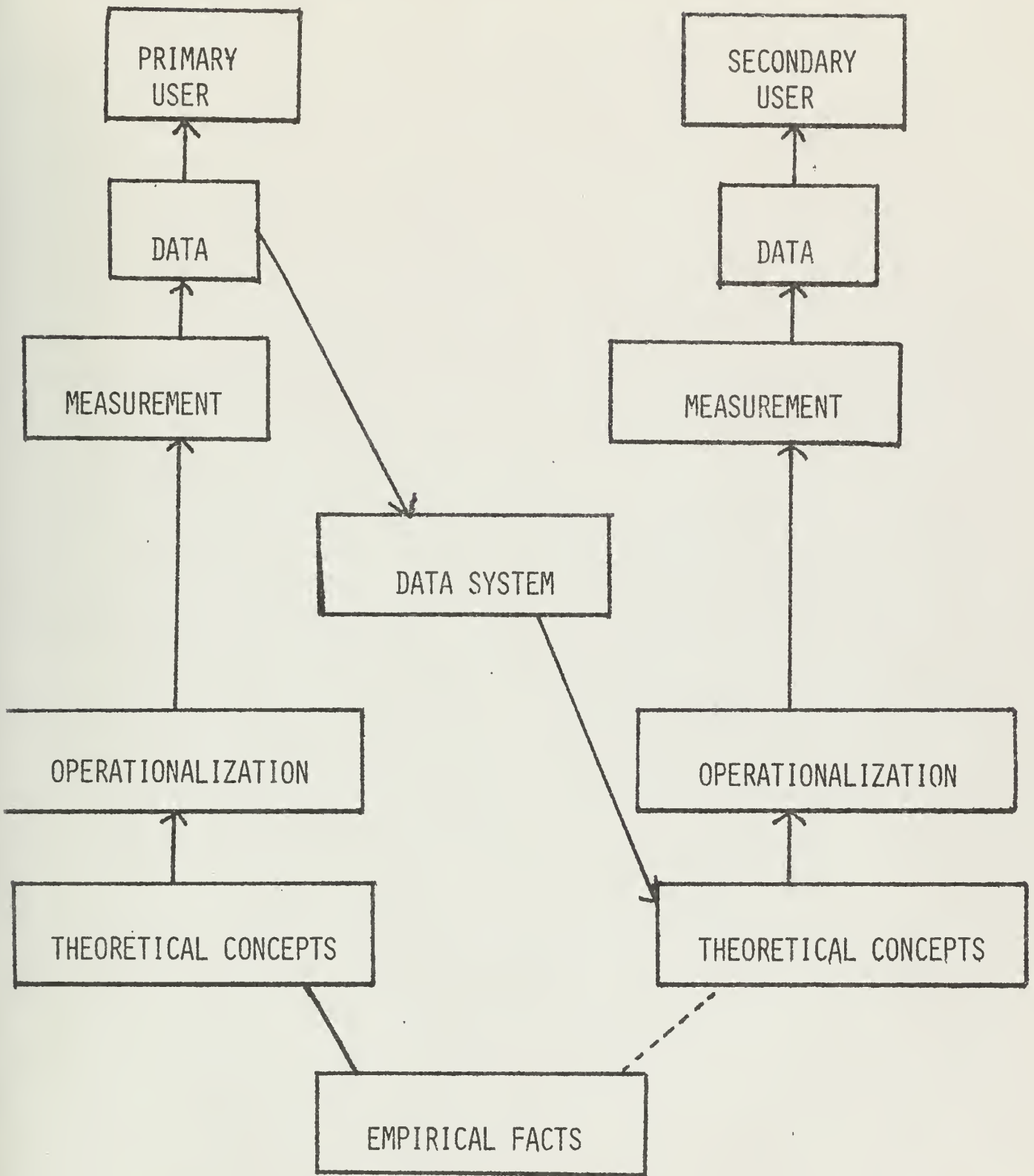1975 "Social Indicators," *Science,* 188: 693-99.
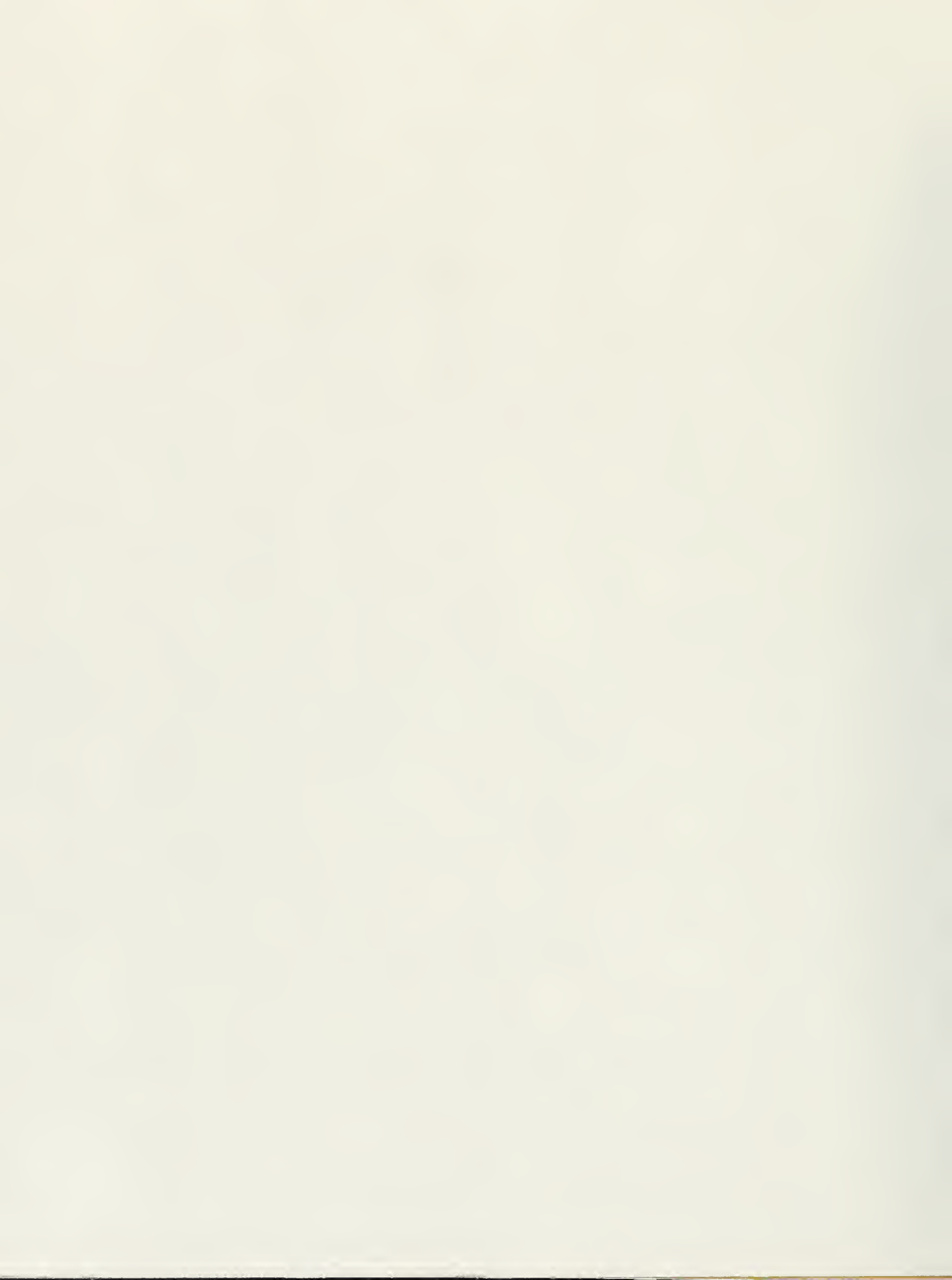
Sofranko, Andrew J.

1974 "The Computer and the County Agent," Journal of Extension 12 (Winter: 29-37.

Stone, Leroy O.

1973 "Some Principles for a System of Socio-Demographic Statistics," Socio-Economic Planning Sciences 7:661-80.

Figure 1. PRIMARY AND SECONDARY DATA USAGE