

Staff Paper

The Changing Relationship of Statistical Data and Analysis

James T. Bonnen

Staff Paper No. 97-17

April 1997



Department of Agricultural Economics
MICHIGAN STATE UNIVERSITY
East Lansing, Michigan 48824

MSU is an Affirmative Action/Equal Opportunity Institution

The Changing Relationship of Statistical Data and Analysis

James T. Bonnen
bonnen@pilot.msu.edu

Abstract or Summary
number of pages: 12

THE CHANGING RELATIONSHIP OF STATISTICAL DATA AND ANALYSIS

by
James T. Bonnen¹
Michigan State University

The term data refers to the direct product of measurement or counting -- i.e., of a collection process. Data are symbolic representations of concepts, quantities, actions, etc. Information is more. It is a decision input that usually integrates data from different collection processes and subject matters with various types of analytic interpretation. Interpretation may range from little more than an exploration of the characteristics of a data set, to editing, imputation or formatting of data for presentation, to estimation or to encoding in an index or scale, or to complex economic, engineering, and biological modeling. Interpretation can be used to assure accuracy, identify needed additional data or information, provide context for and thus relevance to a specific purpose, and generally add value to a data set by combining it with other data or to established theoretic or analytic frameworks. Information is data that are processed, organized, interpreted, and communicated ultimately to provide utility in a specific decision or problem context (Bonnen 1977).² Almost all statistical agency products involve some amount of interpretation or analysis, even though statisticians rarely think of much of this as analysis.

Changing Demand for Data and Analysis

The environment within which statistical agencies operate has changed greatly over the last two decades or so. Rather fundamental change has been taking place in the nature of the demand for data and information. Demand has been slowly shifting away from the relatively homogeneous, public and private, subject matter data sets of the past. The growing complexity of industrial societies has changed the agenda and processes of policy making and, thus, also changed the nature and often the value of information needed for decision. Since World War II, our society and economy have become very much more complex, specialized, and interdependent. Many markets are global. Various sectors interact, each sector creating many kinds of conflicts and impacts external to itself. In responding to these growing conflicts and external impacts, many groups in society have pressed government to intervene in a pervasive manner, and with immense impact. Public and private policy making has become far more extensive, interactive, and

complex. The distinction between public and private sectors has become blurred. As a result of this greater complexity and interdependence, policy decisions today are much more dependent on quantitative information to identify and understand very specific, complex problems, problems that have gotten beyond the capacity of "seat-of-the-pants" decision making. Thus, demand is shifting away from general subject matter information toward more decision-specific information crafted from multiple subject matters and sources. Private sector demand for data and information has, as a whole, increased over the last two decades, and with it the number of users and the value of information. The private sector role in producing decision specific information has grown with this change in the nature of demand for information.

The speed of interaction has increased along with scope and complexity. This means one must have near real time access to data and analysis capacity. Without such capacity in place the decision maker will not be adequately informed in a timely enough fashion to perform or compete well. In facing complex problems, ad hoc decisions, public or private, increasingly are prone to large errors. The short run demand for information is thus more inelastic, rigid -- large changes in the cost of information induce proportionately smaller changes in the quantity of information purchased. The compression of time for a decision also means that any information relevant to such a decision becomes "old news" faster and thus often loses its value more rapidly than in the past.

Several decades ago, most of the data used by both public and private decision makers were descriptive data drawn from single, general subject matters (e.g., housing, health, agriculture, education, energy, or subsets thereof); these data were combined with a modest amount of analysis to provide information for decision makers. This is multidisciplinary, subject matter information useful to a set of decision makers facing a common set of problems. Today, in a far more complex, interdependent economy and society, some information needs still may be served reasonably well by subject matter information. But a larger, growing part of a decision makers' information needs require more specific, more highly processed information products combining complex analysis with sophisticated, multiple subject matter data bases. The largest part of the private information industry has developed in response to *decision-specific* needs. At the same time statistical agencies are being pressed or expected to produce more highly processed *specialized problem-oriented* (not problem or decision specific) subject matter "analysis" to serve the

public good sector of statistically based information demand as well as to provide products for further processing by private sector firms into information for a specific decision (see figure 1). Today the decision maker needs more than specialized problem-oriented, subject matter information. Rather the policy or decision maker ultimately needs *prescriptive problem-solving* information designed to address a single, specific problem (Johnson 1986; Bonnen 1986).³

Government statistical agencies have long provided *subject matter information* (e.g. on housing, population, energy, transportation or agriculture) for various classes of decision makers. In some cases this has been supplemented by subject matter analysis. Statistical agencies normally do not provide highly prescriptive and specialized *problem-solving* information, even for most public policy makers (see Figure 1). In government this can be provided by analytical and some research agencies, but more commonly by action agencies, and policy analysis groups in a cabinet secretary's office such as the Office of the Assistant Secretary for Policy Development and Research in Housing and Urban Development. This type of problem solving or policy analysis has even been contracted from the private sector in a few cases. For private sector decision makers problem-solving information is provided by in-house analysts or private sector information firms, or consultants. If they are to protect their reputation for objective provision of data and analysis, statistical agencies must avoid this type of prescriptive policy analysis, which in government inevitably involves political value judgments. This does not justify avoiding formal analysis of any sort, a self-defeating mind-set that limits ones capacity even in data design and collection.

Growth of the Private Information Industry

A private sector information industry has evolved in response to the need for highly specialized, decision-specific, problem-solving information, which neither public sector agencies nor the typical commercial or manufacturing firm is prepared to provide for itself or for others. When a private firm or public agency needs information that cannot be generated by its own internal information system, it is often faced with high costs and may have limited technical capacity for retrieval and analysis of data from external sources. Specialized information firms have found multiple niches in providing data and analysis to both public and private sector institutions. These niches range from provision

of on-line, continuously updated data bases for modeling and forecasting, to specialized analysis by consulting, research, and management firms, to statistical design, collection and analysis of data, to various types of communication and publishing services. A large number of information-industry firms are in the business of further processing and analyzing government statistics to provide specialized forms of information, which they then sell, not just to private sector firms, but back to the government -- even sometimes to the agencies that collected or processed the original data; e.g., when the U.S. Commerce Department buys economic forecasts from Data Resources Inc. or Chase Econometrics.

Changes in Information Technologies and Statistical Knowledge

At the same time, improvements in information technologies, institutions and statistical methods have also led to changes in the demand for information as well as creating greater capacity to provide it. The growing complexity of the economy has led to increasing interaction and interdependence between different sectors of the economy and society with the consequent need in decision-making to integrate data for new purposes from quite separate data bases. This creates a large challenge since these data sources in their original form are often based on different and thus inconsistent concepts and measurement procedures and in a decentralized statistical system are lodged in many different organizations. The same forces of complexity are pushing the demand for information away from descriptive, even data-rich but broad subject matter information toward information that combines multiple subject matters with far more analysis, much of which is focused on a specific problem or set of problems. Increasingly decision-makers are not satisfied with the older, traditional subject matter data products.

To maintain a highly relevant public good role, statistical agencies must respond by 1) providing more decision or problem-*oriented*, objective analysis with their data products or 2) collaborate with analytical agencies to provide this analysis; otherwise, by default 3) they leave it to private sector analysts to fill a public good gap, often with access to information limited to a few clients or with fragmented and, for many users, quite unsatisfactory results. The private sector information industry has a growing role to play, but their potential for providing public good data and analysis is inherently limited. Indeed, the private sector information industry is growing rapidly in the provision of data and analysis directly tailored for *specific* decisions, while in many traditional areas the role for public provision of

information *direct to private sector decision makers* is shrinking. Increasingly public information is further processed by private sector information industry firms before going to final users, especially those users without significant internal analytic capacity.

Improved information technologies and statistical methods have increased the *potential capacity* of statistical, analytical and research agencies to respond to the growth in demand for more detailed information that is problem-oriented and richer in analytic content. Achieving this capacity requires considerable start-up investment of R&D time, training and budgets for newer technologies. Unfortunately in many cases the executive branch has not approved or the Congress has not appropriated the necessary funding. Thus, new technologies, statistical procedures and designs have been developed, but implementation by government continues to lag well behind potential. Since the government budget prospects remain bleak, agencies face very difficult priority decisions in providing more data and analytic rich intermediate products and to deal with the criticism and the pressures to provide adequate measures of the growing service sector, improve the accuracy of the CPI, trade statistics, or the national income and product accounts. All are used in macroeconomic policy decisions as well as much private sector decision making. Despite a growing role in analysis, providing more analytic rich data bases is not always something private sector information firms can do easily, especially when this involves analysis that requires “mining” the microdata files of statistical agencies all of which must be protected by the promise of confidentiality.

The consequence of all of this is that statistical agencies cannot escape responsibility for greater integration and coordination of their data today with other data and with analysis. Statisticians often think of this as “processing” of data the agency has collected, not so much as analysis until this subsequent processing becomes formal modeling or becomes highly judgmental. We must face the fact that it is no longer meaningful to think about statistical production processes as separate from analysis in any fundamental sense. Statistical agencies today must have the subject matter analytic capability necessary to integrate the data and needed analysis or to collaborate with outside analysts or analytical agencies without “losing the ball” in the hand-off. There are consequently several areas of concern that need the attention of statisticians and cooperating analysts.

1. Changes in the nature of demand for information clearly suggest the need for statistical agencies to examine their information output for the integration of related data (in time and scope) for user convenience and relevance both within agencies and between. Clearly too, more analysis needs to be packaged with data products. This is a frontier that statistical agencies share with analytical and research agencies as well as some of the action agencies in government for whom they provide data.

The growing complexity of public and private decisions also requires greater integration of data from different sources that are typically processed and published today as separate series. Developments in environmental and natural resource policy, for example, have led to a need to integrate public data from across several cabinet agencies--a very difficult challenge. Even individual statistical agencies could produce more value for public and private decisions from their existing data, if the microdata from different series were exploited to provide single, more user-friendly reports bringing together for specific users otherwise disparate series in a common time frame and focus. Statistical and analytical agencies might profitably examine the packaging of their information products for opportunities of this sort both within their agencies and between. This collaboration is already a highly productive one in many places. In DOT the integration of data on the different modes of transportation holds substantial potential.

2. Some of the same changes in user demand suggest both statistical and analytical agencies or groups should explore the question of what additional *objective analysis* should be undertaken to make their output more relevant for use by decision makers in specific problem areas. A clear sense of who the most important public and private users are necessary. Statistical agencies often need to think about developing more subject matter analysis capacity or making greater use of the analytic capacity of government research organizations, or more likely, both.

Almost all US statistical agencies, need to provide more analysis in their statistical reports than they currently do. They preside over microdata sets that hold significant, but unutilized potential for various sorts of analysis. Analysis, however, takes endless forms and has many different purposes. These differences need to be understood and handled carefully. For example, different forms of analysis are necessary to:

- a) Understand and explain the significance of published data, or changes in data, for the public,
- b) Identify strengths and weaknesses in data for advising users on appropriate uses, and to feed back to statistical design and collection,
- c) Develop subject matter understanding of various problems faced by society and its decision makers,
- d) Decide what should be measured to represent accurately a concept or dimension important to analyzing a problem,
- e) Identify problems created by changing economic, social and demographic structures and societal behaviors, and
- f) Objectively identify alternative policy options and analyze their pro's and con's for addressing an important problem.

Whether statistical agencies should do all of this or not depends on the presence or absence elsewhere in government of well organized and closely coordinated research and analysis capacity. In some statistical systems the last several types of analysis listed above are done by analytical and research agencies. In other systems it is left to a statistical agency, if it is to get done at all. However, any statistical agency producing good quality statistics must at a minimum invest in the type of analysis described in 2a through 2c above. Failure of government to provide all these types of analysis will lead users to view statistical agencies as unresponsive or statistical product as lacking in relevance. This responsibility of statisticians for analysis, however, does not include direct policy analysis, which is problem solving and prescriptive in nature and involves making or supporting political assumptions or value judgements that will erode one's reputation for objectivity (see Figure 1). Given the current structure of DOT, the BTS has an opportunity to provide the analysis necessary to integrate the disparate modal data bases where that is needed to serve cabinet, White House and Congressional decision needs.

3. Many users, especially the growing number of new users, are often less than expert in interpreting statistical reports. Statistical agencies issue monthly or quarterly reports of long historical standing, often without much

explanatory text. In the current context of the greater complexity faced by decision makers and especially the growing number of users, a lack of interpretation or analysis can lead to user misinterpretation, confusion and misuse of data, and eventually to user distrust of government statistical products. Statistical product that is abused or questioned by users soon begins to lose value as information. Thus, the increase in demand is not just for data but for data plus analysis sufficient to the needs and understanding of the decision maker.

Closely related is the need to provide users and analysts with a clear documentation or description of the methods involved in collecting, processing, revising and analyzing statistical agency information products. This is an even more important need today in facing the growing complexity of information products

4. Transactions between statistical agencies and the research or analytical agencies and “policy shops” in a cabinet secretary’s office (or in action agencies) where the statistical products are used have long exhibited a weakness that is critical in an era requiring greater integration and coordination of decision systems. Analysts and researchers need to understand the capability and weaknesses of any data they use. Statisticians must understand the expectations of their statistical products generated by the data specifications for any model, or mode of analytic use. Serious mismatches between the characteristics of the statistical products used and the data requirements of analysis result in low quality or badly flawed data for analysis and, thus, often flawed information for decisions. Statistical agencies need to have in-house the analytical skills in the subject matters for which they collect data. This is necessary, if they are to understand their data sets and the analytic demands on that data well enough to communicate effectively with analysts and to understand and anticipate changing user requirements.

On the other side of this transaction, analysts and their agencies often lack the level of statistical expertise to understand statisticians and statistical processes well enough to communicate intelligible analytic needs to statistical agencies. In general, very little thought is given to how prone to error this transaction is in many environments. Statistical knowledge needed for data design collection and processing is not the same as that used in various analytical techniques. Analysts also often fail to do sufficient data analysis before using data, and sometimes do not invest enough in subject matter analysis and problem analysis (see 2a through 2f above). We need to think more carefully about what is necessary to make this transaction between statisticians and

analysts an effective one for users of information. Requirements will vary greatly depending on the sources of the data, the nature of the subject matter and the specific decision uses involved.

5. With the growing concentration and vertical coordination and integration in many markets, the policy goal of a “level playing field” becomes more important, but also more difficult to achieve. In attempting to deal with the conflicts and dilemma’s involved, statisticians cannot afford to yield to the private information industry’s ever present pressure to publish only data and leave all interpretation and analysis to them. This may add to private profit but it trivializes and undermines the integrity of the public information function and contributes to destructive market concentrations and less efficient markets. The changed nature of demand for information makes such concentration more likely and more destructive of broadly based competitive markets. In addition, analysts and statisticians need to be careful about satisfying only the needs of the most sophisticated user, which can increase the disparity between sophisticated and new, relatively less sophisticated user’s ability to extract value from statistical products.

Questions to be Answered about BTS

1. What is the scope of opportunity for BTS to produce a) statistical products; b) analytical products; and c) problem oriented information products?
2. To what extent does existing analytical capacity in the different organized modes of transportation preclude much opportunity for 1b and 1c above? (The answer turns more on the quality of the analytical capacity already present in these modes of transportation than on physical capacity.)

References

- Bonnen, J.T. "Assessment of the Current Agricultural Data Base: An Information System Approach." A Survey of Agricultural Economics Literature. Vol. 2. eds. Lee R. Martin and George G. Judge, at al, Minneapolis, MN: University of Minnesota Press, 1977, pp. 386-407.
- Bonnen, J.T. "A Century of Science in Agriculture: Lessons for Science Policy." American Journal of Agricultural Economics Vol. 68, No. 5, 1986, p. 1066.
- Johnson, G.L. "Ethics, Economics, Energy and Food Conversion Systems." Chapter 7 in Food and Energy Resources eds. David Pimentel and Carl Hall. New York: Academic Press, 1984, pp. 147-180.
- Johnson, G.L. Research Methodology for Economists. New York: Macmillan, 1986, pp. 11-28.

NOTES

1. Professor of Agricultural Economics, Michigan State University. This paper was prepared in draft form in June 1996 for the Panel on Statistical Programs and Practices of the Bureau of Transportation Statistics sponsored by the Committee on National Statistics of the National Academy of Sciences.
2. This distinction between data and information holds only in a process and product sense. Epistemologically data and information are identical; in both cases, the inductive content implies a deductive prior and vice versa. Also data as collected can have some utility in decisions. The common example occurs where most of the utility goes to whoever acts first on some new data on sensitive markets.
3. It is useful here to distinguish three or four general types of information: i.e., disciplinary, subject matter problem oriented and problem solving. Disciplinary information is the theory, empirical measurements and/or measurement techniques and methods explaining a fundamental class of phenomena such as physics, statistics or economics. Subject matter information is multidisciplinary in nature and is useful to a set of decision makers facing a common set of problems. Problem oriented information is multidisciplinary subject matter information useful to a single decision maker facing a specific problem--or a set of decision makers facing the same specific problem. Problem solving, however, requires that one reach a prescription for action and thus combines information with a judgment that one ought or should act on that information (Johnson, 1984, pp. 155-157).

STATISTICAL INFORMATION AND ANALYSIS
IN DECISION MAKING

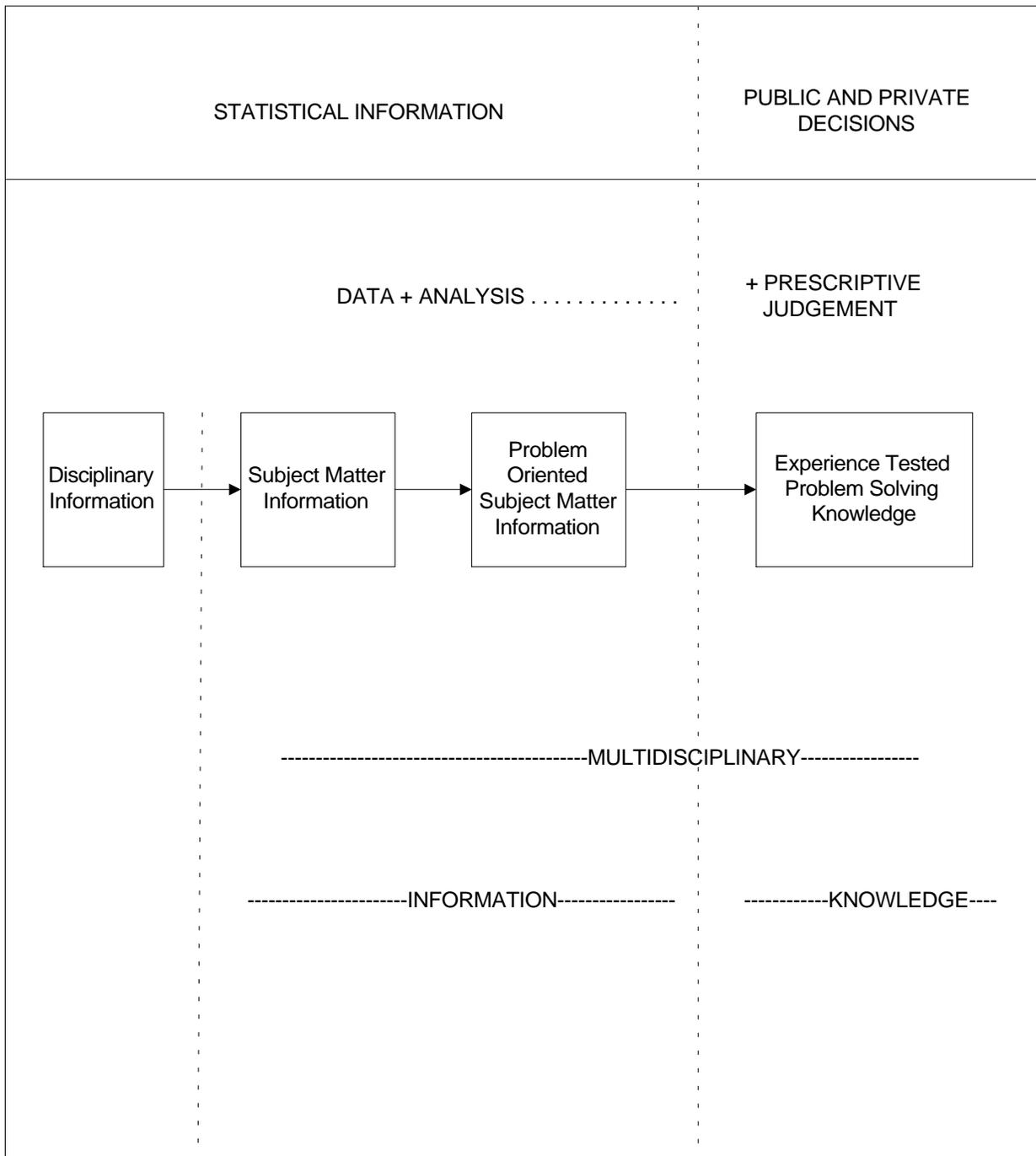


Figure 1