

## PRIVATE AND PUBLIC SECTOR RESPONSIBILITY FOR THE COLLECTION, DISTRIBUTION AND ANALYSIS OF STATISTICAL DATA

**Joseph E. Kasputys**  
Data Resources Inc.  
Lexington, Mass.

Statistical information is a vital national resource. Quantitative data on who we are and what we do influence countless decisions in all parts of the Federal, state and local governments, in commercial and investment banking, in manufacturing, in service and retail trade, and in various nonprofit enterprises such as hospitals, schools and foundations. Such data also play a role in our personal lives, shaping education and career choices, personal finance and investment strategy, regional preferences and other lifetime decisions. The collection and distribution of statistics is an essential role of government.

Like most resources, statistical data is indeed a limited resource. The American public, in general, and business, in particular, as evidenced in the mounting concern over paperwork burdens, have a finite capacity to respond to the ever-growing demands for information from the government. It simply costs too much to respond to the total sum of information requirements arising from legislation, regulation, and other individually worthy program requirements. Once the data are collected, this proliferation of information can have the effect of turning "more into less" by making it extremely difficult to select relevant data from a myriad of conflicting sources, definitions, and time periods. At the same time that both the capacity to produce and the capacity to utilize statistical data have approached the limits of reasonableness, we have had the accompanying orthogonal concerns over the right of the public to have access to data collected by the government and the right of the individual person or business to receive appropriate protection from invasions of privacy.

For these reasons, a more comprehensive effort does need to be made to review and control Federal statistical policy. Current programs conducted by the Office of Federal Statistical Policy and Standards are certainly helpful to assure that maximum use is made of the existing statistical system and that new requirements are appropriately designed to impose a minimum burden on the reporting public while meeting, to whatever extent possible, any unfulfilled needs for data that affect a broader user community. In my own view, these current programs are not sufficient. Further, it is unlikely that any measures will be truly sufficient, including proposals to establish an Office of Federal Information Policy

or an independent Office of Statistical Policy in the Executive Office of the President, until both the Legislative and Executive Branches fully recognize statistical data as a scarce resource and begin to treat it accordingly. However, more central review and control should contribute to this realization and, therefore, I would support some form of increased coordination and control by the Executive Office of the President.

In considering what types of control may be appropriate, it is helpful to reflect on why the Federal government collects statistical information. I believe there are four major reasons:

1. **Regulation.** Regulation implies control, and control requires comparison of actual performance with a standard. This comparison requires measurement, which converts into a need for statistical data. The explosion in Federal regulatory activity has, in turn, generated enormous data requirements. Legislators have also learned that statistical data options can affect regulatory outcomes. As a result, legislation increasingly specifies the data to be used, which limits the flexibility that should be present in the design of the Federal statistical system.
2. **Program Operation.** Many programs require data in order to operate at all. Major examples include revenue sharing, local public works, and similar grant programs tied to specific formulas. Other programs, while not operated purely on a statistical basis, require data for evaluation of effectiveness and possible modification. As with regulation, those who design programs in the Executive Branch and enact them into law in the Legislative Branch have learned that the data used do have a material impact on program results. This is, of course, true with formula programs, where endless varieties of variables are tested in alternative formulas until one is found that provides the program designer with a distribution of funds that is intuitively acceptable. Again, the needs of the Federal statistical system are secondary in this process, which instead encourages the development of increasing amounts of data collection to provide statistics that are uniquely appropriate for each program.

3. **Policy Analysis.** Statistical data must be collected to provide information to policymakers in the Executive and Legislative Branches on economic and financial conditions that exist nationally and internationally. Data on the national income accounts, production, prices, population, trade, and similar items fall into this category. It is probably the area that requires the most difficult decisions and one in which an office charged with statistical policy can have the greatest influence. If the data needs of society could be correctly anticipated, and if proper discipline were exercised in program design, legislation, and regulation, practically all the statistical information needed for regulation and program operation would be regularly available through the thoughtful and coordinated development of the statistical system. While this may be an ideal that can never be reached, it is a worthwhile goal to keep working toward.
4. **Information Programs.** Certain government programs exist for the principal purpose of keeping the public informed. The intent may vary from stimulating technological innovation through the diffusion of knowledge to reducing market imperfections through improved information.

What statistical policy control and coordination will be effective given these sources of data requirements? For the first two sources, regulation and program support, it appears clear that statistical policy must be effectively incorporated into legislative proposals and regulatory actions. This can be best done from the perspective of the Executive Office of the President. Hopefully, with growing public concern over the paperwork and reporting burden, the Legislative Branch will become more sensitive to the need to make reference to statistical policy and precepts before enacting legislation and make increasing use of the capabilities of the proposed organization. In the third area, statistical data for policy analysis, I believe the existing highly decentralized Federal statistical system requires the improvement that can only be obtained through an organization that has broad perspective over Federal activities and sufficient clout to make a difference. This again argues for a stronger central role from a high level.

In the fourth area, government programs to provide information to the public are more controversial. Prior to the extensive development of publishing, media, and information industries, the government indeed fulfilled an important need by supplying information facilitating commerce and industry that would not otherwise be available. However, in 1980, we find a highly developed information industry, utilizing the latest technology, that has a strong capability to identify information needs, collect data, deliver results frequently tailored to the needs of specific clients, and even provide special analysis on the meaning of information for business decisions. In lieu of unilateral government determinations on the nature of private sector information requirements, market demand and the profit motive can now be used more extensively to govern statistical collection. Equally important, the reporting public becomes free to choose whether to respond to information requirements generated by the private sector, which can be expected to be in proportion

to the perceived value of the information. The efforts of a strengthened central statistical policy office should focus on transferring data collection and distribution of this nature to the private sector, while improving and rationalizing statistical operations in support of regulation, program operation and policy analysis.

I would like to encourage any new statistical office to remain separate from other aspects of information policy. Statistical policy is sufficiently unique and important to deserve separate attention. To be sure, statistical policy must be established with an awareness of telecommunications, ADP management, privacy, and related concerns; but should not be merged with these other functions which are more involved with regulatory concerns. Incidentally, with the reduced costs of computers and communications and rapidly rising costs of personnel, the emphasis on strong central control over the acquisition and use of ADP and related services is probably misplaced. More emphasis should be given to the improvement of productivity through the use of these capabilities than to elaborate restrictions on procurement.

Given that the Federal government does, and must, collect statistics, the government bears a major responsibility to make this information available to the public. Divergent interpretations of the term "making information available to the public" are possible. Interpretations of this term can range from placing information in a public reading room somewhere in Washington, D.C., such as the SEC public reading room, to the Federal government's placing all available statistics in a shared computer and aggressively marketing these statistics to private sector organizations. To the extent that controversy exists over the interpretation of the Federal role in information dissemination, I believe that much of it can be traced to two conflicting principles. On the one hand, we wish the Federal government to make available all information that is collected to the public, except where individual or business rights to proprietary and confidential information would be violated. On the other hand, an equally important precept in our society is that the Federal government should not encroach on activities that properly belong in the private sector. It is fortunate for our economic health that the Federal government has avoided entering into legitimate private sector enterprises where not needed because, despite this restraint on private sector encroachment, we have only seen the role of the government grow bigger over time.

It is clear that an extensive information industry has developed in the United States to take full advantage of available communications and computer technology. Some testimony to the existence of this industry is an article in the May 5, 1980 edition of FORTUNE (which incidentally is the 25th anniversary edition of the FORTUNE 500), entitled "Everything You Always Wanted to Know May Soon be On-Line." This article explores the rapid development of the on-line information industry, which includes both bibliographic and statistical databases. The on-line database industry includes firms that specialize in organizing and delivering just one or a few databases, along with those that provide almost encyclopedic coverage of all available quantitative data.

Many of these firms indeed carry data that are collected and published by the Federal government. Since they are involved in assessing market needs and matching these needs to data availability, it is likely that most of the markets for statistical data, including data provided by the Federal government, that can be economically served either have been or will be identified. Products will be developed based upon available data by the private sector information industry to serve the needs of these markets.

It is important to emphasize that I know of no one in the industry who does not believe that the government should have access to all available technology to use in the dissemination of its data. Our principal concern is rather whether the government is competing with already existing private sector activities, which include computers, telecommunications and software that have been developed at considerable private expense. One consideration may be whether the existing industry efforts adequately serve all the markets for information that require such service. There is little concern over whether the FORTUNE 500 companies, major banks, or large research organizations are being adequately served with data. Questions have been raised from time to time over whether the needs of the small organization or the individual researcher in the university are being adequately met, and whether the government should not step in, using the latest available technology, to meet these needs. This is clearly a matter for public policy to decide. If such needs are not now being met by the private sector, it is likely that they cannot be met economically on a full cost recovery basis. At such time as the technology would permit such needs to be met economically, I am certain that the private sector would step in. Since it is unlikely that the government could perform these services any more economically than the private sector, we are then faced with a decision as to whether to subsidize information dissemination to these specialized markets. It is my own belief that in most cases, if the government makes information available through depositories and through statistical publications and reports, the small business or the individual researcher with an occasional need for information will be able to gain adequate access to the data and reports he or she seeks.

I would also encourage the government, as I mentioned earlier, to make full use of computer and communications technology in its own internal dissemination and utilization of statistical data. The government has been an advanced user in this regard. All private citizens would hope this will continue, for it should lead to the more effective use of statistical data for public policy development. It is axiomatic that data always receive less analysis than they deserve. New technologies available through the computer and communications networks should permit higher levels of analysis to be conducted, leading to better decisions. The only caveat to be considered is the classical "make or buy" decision that is observed within government, which centers around whether to develop an inhouse capability for the storage, distribution and analysis of such data or whether to obtain it from private contractors. Here again, I believe there is ample evidence to support the view that the private sector is considerably advanced in the distribution and analysis of quantitative data and has much to offer the government in this

regard.

In order to gauge the adequacy of the on-line database information industry, it might be useful to provide a few additional words of clarification on industry operations. DRI serves as a good example of this industry and indeed was cited in the FORTUNE article mentioned earlier as "probably the preeminent company in the on-line information field." DRI has spent many millions of dollars over the past 12 years in collecting, organizing and documenting data. Only a small portion of these data originate in the Federal government. While DRI does maintain data on the national income accounts, prices, population, housing starts and the like, much of DRI's data originates elsewhere, from such agencies as the OECD, IMF, World Bank, foreign governments; and from daily stock prices, commodity quantities and prices, market research surveys, and material published privately by trade associations. A representative listing of DRI databases is attached in an Appendix to this paper. DRI's current on-line storage capability for immediate access material is in the neighborhood of 36 billion characters and is scheduled to grow to 55 billion characters by the end of this year. We believe we have the largest collection of economic and financial information available anywhere. Many of these databases were originally issued in cross-sectional format. DRI has taken many periods of cross-sectional data and converted them into consistent time series. This often involves dealing with definitional changes from one time period to the next, changes in the data collection approach, or changes in the basic entities measured. Once organized in a consistent time series format, the data are described both in on-line documentation and in reference manuals. Mnemonics are assigned for easy use in referencing through software. DRI has developed software that permit any and all on-line data to be retrieved and manipulated with operations varying from seasonal adjustment to all forms of regression, to forming equations, to building models, to simulating the models, to producing reports and graphs for effective presentation of results. More importantly, it has been DRI's finding that on-line databases can only be efficiently and economically offered to markets with a good understanding of the applications to which the data will be put. Indeed, part of our product offerings include applications methodologies and consulting assistance that enable users to combine their own data on sales, production costs, and competition with the generally available data provided by DRI.

Questions have also been raised on the appropriate role of the Federal versus the private sector in performing analysis of statistical data. It is my belief that the Federal statistical agencies should focus their analytical efforts on developing descriptive information from data that are collected. This involves organizing data to present it effectively, so that the users of the data can fully understand what the statistics mean. This is a critical role which will often stimulate policy action, since proper presentation will of itself indicate that such action is needed. Any analysis that goes beyond display or rearrangement of factual information should be limited to political appointees and policymakers, together with their supporting staffs, who should be kept separate from the statistical agencies and the information which they release.

This latter type of analysis would include forecasts and interpretations of the underlying meaning of statistical information. It is important that analysis of this sort be released separately from the factual statistical information provided by Federal agencies so that the public can readily tell fact from opinion.

Private sector firms, of course, should engage in any and all analysis appropriate for the markets being served. Indeed, this provides a plurality of information to the general public. The information industry has an important role to play in providing alternative interpretations of statistical data to businesses and industry, which can then be used to form opinions on the validity of the analysis being provided through the political system. This plurality of information, together with multiple sources of statistical information delivery, is a healthy aspect of our society and is essential to maintaining an informed public which can reach independent

judgments on major economic and financial issues.

The development of our national statistical system continues to be a matter of great importance to all economic and financial activity. The proper development of this system goes beyond narrow economic and financial concerns, and has direct influence in maintaining the principles of our democracy. The development and use of statistics should be a mutual undertaking, shared by both the Federal and private sectors. I believe these sectors have worked together very cooperatively in the past to give the United States clearly the best statistical data available anywhere in the world. We now face new challenges in how we can work together to fully utilize the advances that technology has given us. I believe that continuing dialogues among all concerned will contribute materially to our ability to face these new challenges and meet the needs of the general public and our private clients in the most effective manner possible.

#### APPENDIX: A GUIDE TO DRI DATA BASES

- |  |  |
|--|--|
| Age-Income Model                             | Forestry and Wood Service Data Banks         |
| Agriculture and Weather                      | IBRD's World Debt Tables Data Bank           |
| Automotive                                   | IMF's Balance of Payments                    |
| Best Executive Data                          | IMF's International Financial Statistics     |
| California                                   | Industry Financial Service                   |
| Canada                                       | Insurance Service Data Bank                  |
| Canadian Model Data Bank                     | International Energy Data                    |
| Chemical Data Banks                          | International Trade Information Service Data |
| Coal Data Bank                               | Japan  |
| Coal Model Data Bank                         | New York City Model Data Bank                |
| Commodities Market Data Bank                 | OECD Main Economic Indicators                |
| Compustat                                    | OECD National Income Accounts                |
| The Conference Board                         | OECD Trade Series A                          |
| Consumer Expenditure Survey                  | Cost Forecasting Service Data Banks          |
| Flow of Funds                                | Paper and Pulp Data Banks                    |
| Current Population Survey                    | European Model Data Bank                     |
| Platt's Data Bank                            | Target Group Index                           |
| Developing Countries Primary                 | European National Source                     |
| Source Data Banks                            | Data Bank                                    |
| PRO FORMA Data Bank                          | Transportation                               |
| DRI-FACS                                     | FDIC Data                                    |
| SITE II                                      | U.S. Central                                 |
| DRI-SEC                                      | FIEI   |
| Standard & Poor's Industrial Financial Data  | U.S. Macro Model Data Bank                   |
| Drilling Data Banks                          | U.S. Prices                                  |
| State and Area Forecasting Service Data Bank | U.S. Regional                                |
| Energy                                       | U.S. Weekly Banking                          |
| Steel Data Banks                             | Value Line                                   |