REDEFINING PROGRESS

BRAD CRABTREE, JILL CHOPYAK, AND CLIFFORD COBB

THE GREAT DATA PARADOX —

Threats to the Integrity of the Federal Data System in the "Information Age"

DRAFT COPY — NOT FOR CITATION

MAY 1999

REDEFINING PROGRESS
ONE KEARNY STREET, FOURTH FLOOR
SAN FRANCISCO, CA 94108

1.800.896.2100

TEL 415.781.1191
FAX 415.781.1198
E-MAIL info@rprogress.org
WEBSITE www.rprogress.org

TABLE OF CONTENTS

I. INTRODUCTION

In 1996, an open assault was launched in Congress on the Bureau of Labor Statistics (BLS). Newt Gingrich, then Speaker of the House, threatened to cut the BLS budget if that agency did not modify its procedures for preparing the Consumer Price Index. Although that particular episode was resolved relatively gracefully (with BLS making minor and legitimate changes on its own to the CPI), it served as a reminder of the vulnerability of the federal data system.

The purpose of this paper is to examine the value of a reliable, consistent public source of statistical information, to analyze some points of vulnerability in the system providing that data, and to propose ways to preserve and improve it. We start with the premise that a neutral process of collecting, processing, and disseminating data is an essential ingredient in providing a common stock of knowledge that helps to bind a democratic society together.

This paper is also an exploration of some fundamental paradoxes in public life. Although pundits have repeatedly announced that our society is at the dawn of the "Information Age," support seems to be weakening for federal spending on information, even though it claims far less than 1 percent of the federal budget. Although access to information is the life-blood of democracy, the federal government sometimes sells it for a price out of reach of ordinary citizens. Although devolution of federal programs to states and local governments involves massive data requirements, little attention has been given to problems associated with tracking and monitoring the transition. Although the public trusts statistics disseminated by the government more than the numbers produced by private interests, Congress is tampering with the credibility of public data by overtly politicizing them. In an age of supposed enlightenment, Congress and the American public are allowing the erosion of statistical capacity that is needed to produce intelligent policies.

This paper seeks to alert citizens to the importance of the federal data system, to make them aware of some potential threats to it, and to present some steps that might be taken to preserve and strengthen it. Because the perils faced by the system are subtle rather than dramatic, they might be ignored for years before the cumulative effects of decay become obvious. Like the rot inside an old tree that looks healthy on the outside, vital information systems may be allowed to deteriorate for a long time before the sickness is recognized.

We do not claim originality for most of the insights in this paper. We have relied extensively on groups in Washington, D.C. that are actively tracking and responding to the threats to the data system.[1] The usual disclaimer applies: we are responsible for errors of interpretation.

The value of this paper lies primarily in its breadth and scope. To our knowledge, this the first occasion on which anyone has attempted to evaluate the overall health of the federal data system, particularly from the perspective of users.

WHY STATISTICS MATTER

As Bill Gates understands, whoever becomes the gatekeeper for the flow of information in the future will have a commanding position in a wide range of markets. That is the real significance of the ongoing antitrust suit against Microsoft. The control of information systems has become one of the most important ways of gaining monopoly power in commerce today. If the strength and credibility of the federal data system are undermined, this will eventually increase the relative power of those who control private databases.

This is not to suggest that there is a conspiracy to dismantle federal offices devoted to data collection and publication. If the statistical system were directly under attack, it would be easier to defend. Instead the forces whittling away the strength of the data-gathering system are subtler. Indifference, stemming from a general lack of awareness of the value of this public resource, is the primary culprit. In an age when citizens increasingly believe the proper role of government amounts to nothing more than providing people with direct services like paving streets or fighting forest fires, gathering and publishing information appears increasingly difficult to justify.

FEDERAL DATA: THE HIDDEN RESOURCE

Yet even if the federal government privatized or devolved all of its other functions and was left with only the barest essentials (national defense, patents and copyrights, standarized weights and measures, the federal judiciary, and certain types of

---

1 These groups include the Taxpayer Assets Project, OMB Watch, the American Library Association, the Bauman Foundation, the Association of Research Libraries, the Congressional Accountability Project, the Council of Professional Associations of Federal Statistics (COPAFS), and the Consortium of Social Science Organizations (COSA). Gary Ruskin 202-296-2787) We would like to thank the following people for assistance in preparing this document.:

Tony Black, Census Bureau; Curt Davies, Environmental Working Group; Margaret Dunkel, Inst. for Educational Leadership; Henry Griggs, Census 2000 Initiative ; Shelly Kossak, Population Resource Center; Ron Kutscher, BLS (Associate Commissioner); Steven Landefeld, BEA; James Love, Taxpayer Assets Project; TerriAnn Lowenthal, Census 2000 Initiative ; David McMillen, House Government Operations Committee; Martha Riche, Census; Peg Seminario, AFL-CIO; Howard Silver, Consortium of Social Science Organizations; Edward Spar, COPAFS; Matt Stagner, Office of Planning and Evaluation, HHS; Cynthia Taeuber, Census Bureau; Kathleen Wallman, OMB, Chief Statistician of the United States.

research), providing data about economic, social, and environmental conditions should remain one of its primary duties.

There are three main reasons it should maintain responsibility for being the nation's chief information assembler. First, it is far cheaper for the federal government to collect, analyze, and publish data than for smaller jurisdictions to do so. State and local governments rely on federal data, and they would find replication of that data extremely difficult if each jurisdiction had to provide it. Thus, when questions are removed from the Census, for example, no other level of government takes over that responsibility. The information is simply lost. Second, there is considerable value in maintaining standardized statistics that permit comparisons across states. Third, the federal government has developed a positive image as a data manager that operates professionally and with considerable autonomy from political pressures. Businesses, newspapers, academic researchers, and advocacy groups that want to analyze economic, social, and environmental trends depend on the government's neutrality in the collection of statistics. For these reasons, the federal government needs to continue playing its long-standing role as a primary collector, analyzer, and disseminator of information, particularly statistical data series.

Because the federal government's role in providing data to the public is seldom mentioned explicitly in the media, this service is likely to be undervalued. It might seem to the casual observer as if statistics just exist on their own. When a newspaper article refers to the number of people who drive to work each day or death rates from cancer, it seldom gives the reader any sense of how such numbers are derived. In fact, a large federal apparatus is required to uncover those basic facts.[2] A few billion dollars is spent each year to supply the statistics that reveal conditions and trends.

If we want to know the unemployment situation in each city and county or how fast the cost of living is rising, there have to be data collectors paid by the federal government to ask people questions about jobs and prices. If we want to keep food on the table into the next century, we not only need farmers plowing fields; we also need information systems about soil erosion, the financial condition of farmers, and the likely effects of climate change on the grain belt. If we want companies to offer life insurance, the government will need to keep collecting the demographic and other

---

2. In the United States, there are four statistical agencies that process economic data: the Census Bureau and the Bureau of Economic Analysis in the Commerce Department, the Bureau of Labor Statistics in the Labor Department, and the Internal Revenue Service in the Treasury Department. There are also a number of specialized statistical offices that deal with health, energy, crime, education, transportation, agriculture, and other issues. (See Appendix A for a detailed breakdown of the various statistical agencies.) Because authority for collecting, analyzing, and publishing statistics is dispersed among all of these agencies, there is little coordination among them. They use different reporting formats, computer programs, and identifiers, which makes sharing of databases difficult both for the agencies themselves and for public users of the data. The problems associated with dispersed authority arises at several points below, but it is not a central theme of this report. Appendix B discusses a number of failed efforts over the past fifty or so years to better coordinate federal data. While that remains a perennial issue, there are more immediately pressing issues now.

information necessary to predict changes in life expectancy. The availability of data on hundreds of topics is an essential, though almost invisible, part of our society.

Like filling in potholes on city streets, the work of statisticians is a continuous maintenance task of great importance. Statistical work permits problems to be identified and good management decisions to be made. Statistics enable governments to make budgets years in advance, preserve environmental resources, protect the public health, reduce crime, and make intelligent choices about investments in public enterprises. Without the sort of knowledge that statistical agencies provide, achieving collective goals would be difficult or impossible.

THE VALUE OF INFORMATION AS A SPOTLIGHT

There are two distinct ways in which data can be of use: directly as a tool for shaping social behavior and indirectly as a tool for making better decisions.

The first type of information involves publicizing socially harmful behavior in order to discourage it directly. This might be called deterrent information.

A recent example of this is the Toxics Release Inventory (TRI), a regularly published database of the amount of toxic chemicals released into the air by private companies and government facilities. As a result of TRI, polluters have voluntarily cut emissions by far more than they might have under a system of legal enforcement. The only sanction in this case was the fear of a bad public image. (See appendix C for more on the TRI.)

Reporting of political campaign contributions also evens the political playing field to some extent. In the absence of legal limits on the size of donations, the disclosure of large contributions may limit them if a legislator does not want to appear too closely tied to a special interest group. When large contributions are made, the mere act of reporting them counter-balances the value of the additional funding to the campaign. In California, at least one statewide initiative may have been defeated by voters because of the appearance that the tobacco industry was trying to "buy" the election.

The State Department's publication of human rights violations in its Country Reports is another example of information influencing behavior directly. The availability of information makes human rights an issue in a way that would not be possible if the data were unavailable. The threat of showing up on that list encourages countries to avoid mistreatment of their citizens. In addition, the reports have made human rights a more prominent issue in the minds of State Department officials (Innes 1990, 15, 16, and 19).

In a fourth case, some school districts have become more fully integrated as a result of filling out a standard survey of the federal Office of Civil Rights. In those cases, the survey sensitized district officials to their inadequate integration policies and to the potential for either legal or political action against them if they failed to

take decisive action. The availability of the results of a district's survey to the general public and to local civil rights groups promotes compliance with the spirit of the law as much or more than overt legal action would (Weiss and Gruber 1984)

THE VALUE OF INFORMATION AS BACKGROUND

The second type of information has a value that is less immediate and striking, but far more common. Most of the data gathered by the Census Bureau, the Bureau of Labor Statistics, or dozens of other agencies provides background knowledge that enables the public and policy-makers to make better decisions. For example, knowledge about the distribution of poverty and homelessness within the population permits better targeting of services to those most in need. Information about the use of highways and transit during commute hours enlightens the debate over how transportation funds should be spent.

The value of much of the information collected by the federal government is not obvious because it may be used by only a small number of researchers and analysts. If only five hundred people use a database that costs fifty million dollars to compile, it might seem a rather extravagant use of public money to continue to conduct the surveys. That view is naive, however. The small number of analysts who use the data provide a service that can benefit the entire population.[3]

One of the most difficult tasks of statistical agencies today is to explain to the public why data collection remains an important task, even when the data are not of immediate relevance to the average person. The connection that a planner sees between demographic projections and budgeting for new school buildings, roads, or other infrastructure may not be obvious to the average citizen. Perhaps tracking data on communicable diseases seems unimportant in a society that takes public health measures so much for granted. More generally, the public will not recognize the value of statistics if it does not comprehend the most basic services that government provides to maintain society intact.

---

3. An analogy may help to explain why statistics used only by specialists are of value to all citizens indirectly. Imagine that the companies making air conditioning units that cool large buildings suddenly stopped publishing maintenance and repair manuals. Only a few thousand people--those who work on the machines--would be directly affected. Ultimately, though, millions of people would be affected as offices and stores became uninhabitable. This is just one of many examples in our society of specialists serving as intermediaries for large numbers of end-users of ideas or products.

Specialists should not be given the power to run society. Instead, to continue the analogy, they might be compared to a maintenance crew that keeps an air conditioner in working order. It is up to the crew to give the people in the building the chance to control the internal climate. In the same way, specialists can provide citizens the technical means with which to control the 'temperature' of society.

STATISTICS AND ACCOUNTABILITY

In a system of representative government, citizens need to feel connected on an ongoing basis to the activities of government. Voting once every two years is not sufficient. Providing citizens with information about economic, social, and environmental conditions and about government responsibilities is one way to maintain that sense of connection. The collection and presentation of statistics allows citizens to evaluate government policies and hold leaders accountable.

Consider statistics about unemployment. Before figures based on sound methodology were published in the late 1940s, the national employment situation could be judged only on the basis of impressionistic evidence, which was subject to speculation and controversy. The unemployment rate, as defined and measured by the Bureau of Labor Statistics, now provides a consistent and widely accepted gauge of the health of the economy. It also focuses attention on a problem that most Americans feel government is responsible for managing.

Government could be made even more accessible and accountable if more such performance measures were established. Thus, the statistical agencies in a democratic society need to supply at least some data that can be understood by the general public, not just technical experts.

Citizens will regard government data as trustworthy only if statistics are collected impartially--in ways that do not merely serve partisan political ends. Historically, Americans have been able to rely on the consistency and integrity of the agencies that publish statistics.[4]

By remaining impartial, statistical agencies can provide information that enables citizens to hold government accountable. To avoid even the hint of bias, those agencies have shied away from measures that would involve subjective judgments (such as quality-of-life indicators).

Maintaining consistent sources of information is essential in a democracy. If citizens believe they can rely on professional statisticians to provide reliable data, regardless of which party is in power, there will be less danger that people will be swayed by demagoguery. Accurate information from a disinterested source allows citizens to judge the wisdom of government policy and to distinguish between necessary and arbitrary government action.

For example, the public became alarmed in 1980 about the dangers to health from toxic wastes buried in the ground based on newspaper accounts of a few extreme cases (Love Canal and Times Beach). The Superfund program, established at that time, has led to billions of dollars being spent by the federal government and private parties.

---

4 . On the occasions when there have been efforts to politicize their work, statistical agencies have been able to resist the intrusion. Their survival has been assured by their remaining scrupulously neutral and therefore relatively invisible to the public.

Sometimes, this money has been spent on relatively minor risks. If better data, revealing the true magnitude of the problem, had been available at the time, a more carefully targeted policy might have resulted. In general, the absence of reliable statistics about the magnitude of a problem allows rumor and anecdotal evidence to cause an overreaction.

Congress and the public need to know more, not less, about the challenges facing the nation. We need more information and less ideological posturing about the economy to explain why national savings are in decline and deficits persist. We need more detailed information about the environment to determine which habitats are most threatened, which chemicals are most dangerous to health, and which methods of forestry management are best. We need more information about the conditions under which people successfully break the cycle of poverty or overcome addiction to drugs.

Holding government accountable with statistics does not mean that government should try to remedy every ill. Intelligent forms of accountability would judge when government was doing too much, not just too little. There are many problems which government cannot solve, but which might be better understood with statistics provided by government.

The failure to distinguish between the need for more information and the need for government spending or regulation has created an untenable situation. In order to block programs they do not like, some politicians now oppose data collection on issues related to those potential programs. They fear the statistics will be used to promote excessive government interference in private markets. Yet, holding the growth of government activism in check should not entail cutting off the flow of information about the condition of American society.

II. EROSION OF THE STATISTICAL SYSTEM

The information services provided by the federal government are simply taken for granted by most citizens and most members of Congress. They assume the data currently collected by the federal government will continue without interruption.

That is a false assumption. In a climate of fiscal austerity, statistical agencies are likely to lose ground. They lack strong political support networks, and their work rarely rouses passions. People in specialized fields may be aware of the importance of statistics in their area of interest, but only a handful of professional statisticians show concern about the health of the statistical system as a whole.

There is a danger that in the next decade or two, much of the data collected by the federal government will decline in quality and quantity. If that happens, judging the harm or benefit of changes in government policy will become more difficult. The capacity of statistical agencies to function as a central nervous system of government will be impaired. Unless the importance of data gathering is recognized by enough members of the press, the public, and elected officials, the humdrum work of collecting and processing statistics might deteriorate. The freely available knowledge-base that makes democracy possible could be eroded in the name of frugality, deregulation, devolution, the efficiency of the private sector, and the need to protect privacy.

Federal statistics require years of advance preparation, like planting a forest for the next generation. Agencies cannot simply gather information at a moment's notice. If the statistical apparatus is allowed to deteriorate, the costs to Americans will be felt for years to come.

Already federal agencies have begun to cut data series, publications, and analytic reviews. Furthermore, new initiatives are not being undertaken as much as in the past. Young people are not being recruited into statistical agencies fast enough to replace those who are retiring. In these and various other ways, this vital organ of self-governance is dying by degrees.

The various threats to the statistical system are discussed in this section of the report. Funding cuts, privatization (monopoly control and pricing), devolution, concerns about privacy, and politicization are all likely to contribute to the weakening of the statistical system.

FISCAL PRESSURES - LOSS OF DATA

Perhaps the biggest problem facing the federal statistical system is the willingness of Congress to treat the creation and maintenance of data series as potentially expendable. Even though data collection and processing are relatively cheap activities and even though the accuracy and comprehensiveness of data are crucial to both commerce and policy-making, many members of Congress fail to recognize the special significance of the federal statistical system.[5] When the budget knife is being wielded, statistical agencies are highly vulnerable.

The Vulnerability Of The Statistical System

The statistical system is vulnerable to budget cuts for several reasons.

It is a relative political orphan, with few friends outside the world of public policy, research and academia. Only a handful of associations such as the Council of Professional Associations of Federal Statistics (COPAFS) and the Consortium of Social Science Organizations (COSA) represent broadly the interests of the federal statistical system.

Most associations, companies and other interest groups have shown a tendency to fight each other for scraps of funding for data, rather than mount a concerted defense of the whole statistical system. Only new political constituencies, willing to support statistics as a public good, can make a difference. One of the best hopes lies with the growing ranks of citizens across the country who use federal data sets to develop their own economic, social and environmental indicators at the community level.

What little attention statistics receive often comes in the form of public mistrust, or a curious mix of official disregard and hype. Also, the media continually reports conflicts over the meaning and accuracy of government figures, lending unspoken credence to the cynical belief that statistics are crassly political.

The dispersed structure of the system (see Appendix A) leaves vital statistical functions tucked away in every nook and cranny of the federal government. Some long-time observers of data issues have argued that consolidation of agencies and programs into a larger institution would enhance the visibility and clout of government statistical functions. Indeed, even old hands can find the system byzantine and difficult to navigate. Hence, important programs, surveys and data sets may face reduction or elimination long before the larger consequences are well-publicized or even understood.

---

5. There are now a handful of staff members or senior committee members in Congress with jurisdiction over statistics who truly qualify as experts on the federal statistical system.

Finally, the nature of the congressional appropriations process forces statistical agencies to compete with substantive programs in unrelated areas. Subcommittees of the House Appropriations Committee are granted fixed amounts of federal discretionary money to allocate within their areas of jurisdiction. For example, the Commerce, Justice, and State Subcommittee, which is responsible for the Census Bureau, receives a fixed amount for all programs. Consequently, the census must vie not only with other Commerce Department functions, but with increased security at U.S. overseas facilities or with putting more police on the streets.

The source of pressure: cuts in all discretionary items

The most recent threats to the statistical system arose in the mid-1990s from pressure to balance the federal budget. Data collection is especially vulnerable to across-the-board cuts. Under those circumstances, it is politically very difficult to argue that data are of greater value than direct services.

Yet there should never be any need to try to balance the budget by cutting statistical budgets. The federal statistical system accounts for only a tiny portion of the overall federal budget. The approximately $2.5 billion currently allocated to all federal data and statistics amounts to less than two-tenths of one percent of total federal spending. Fiscal austerity for federal statistical programs registers scarcely any impact on the budget. Cutting the information-gathering activities of the government in order to fight a deficit is extremely short-sighted.

Nevertheless, the desire to balance the overall budget provides political cover for efforts to dismantle the federal knowledge base. A member of Congress who offers to save taxpayers $300 million can sound like a fiscal hero to the uninformed, who do not understand the consequences of diminishing the quality of the Census or forcing another agency to rely on its old computer equipment.

Budget deficits do not stem from spending on the routine tasks that government has always performed. The tide of red ink has flowed from the unrestricted growth of entitlements (Social Security, Medicare, Medicaid, and federal pensions) and interest on the national debt.[6] If current trends continue, by the year 2012, interest payments plus entitlements will consume all federal taxes (Bipartisan Commission 199x, 11). At that point, the only way to balance the budget would be to close down the rest of the federal government (defense, air traffic control, housing assistance, job training, national parks, environmental protection, the federal courts, and so on).

---

6. Entitlements are payments that Congress must make unless the laws are changed. The rest of the budget is called discretionary because Congress has the power to control it. Real spending (in 1995 dollars) on entitlements (including net interest) has grown from $264 billion in 1970 to $986 billion in 1995. Meanwhile, real defense spending declined from $350 billion in 1986 to $272 billion in 1995. By piling up debt, the fixed portion of federal spending dedicated to interest payments has grown from 7.4% in 1970 to 15.2% today. For price deflators and defense spending, see Council of Economic Advisers (1997, Tables B-3 and B-78). For growth of entitlements, see Fleenor (1995, Table C-12).

Statistical programs are part of discretionary spending, which means they are in that part of the budget that is being squeezed by the growth of entitlements. If the federal budget slips back into a deficit, statistical funding will again be jeopardized.

Even in the midst of budget deficits, however, cutting spending on statistics is a mistake. It is equivalent to eating the seeds during a famine that would have been sown the next spring or selling off the library of a major university to finance a short-term deficit. An act of desperation in the present will cause even greater suffering in the future.

Thus far, budget-cutting has not been sharp for most statistical agencies, with the exception of the Energy Information Adminstration and the Economic Research Service (Department of Agriculture). From fiscal year 1994 to 1997, the average decline in the inflation-adjusted budgets of the largest ten statistical agencies was about 5%.[7]

For some agencies, these cuts follow two decades of growth in both budgets and demands. For example, the budget for BLS has roughly tripled in real terms. But that sort of growth was not matched by Census and BEA, the other two agencies that constitute the nerve center of the federal statistical system. Exclusive of the decennial census, the former grew by about 30% in real terms, while the BEA budget remained relatively flat until it grew by almost 50% during the Bush Administration.[8]

In addition to the termination of data series, an equally serious type of cutback is the reduction in explanation of the data that continue to be published. For experts in a field this may not cause too much of a problem. The public loses, however, when federal agencies collect only the raw data and cease providing an evenhanded analysis of the statistics they produce.

Unless the users of statistics in state and local government, citizens groups, universities, and industry speak up, the data they have come to rely on may simply stop being collected, one series at a time. A war of attrition on the data system has already been declared. The effects have already begun to be felt. The future promises to be far worse, however, unless a concerted effort is launched to maintain this basic function of government.

---

7. That figure was calculated from Table 1 assuming cumulative inflation from mid-1994 to mid-1997 to be 6%. If EIA and ERS are excluded from the calculation, the decline would have been only 1.3%.

8. These estimates of long-term growth are based on calculations made by David McMillen, minority staff, House Committee on Government Reform and Oversight. The figures for the Census exclude the rapidly escalating costs associated with administering the decennial national census.

Agency officials sometimes assert that statutory requirements placed on statistical agencies have grown even faster than budgets, but this claim has not been well documented. Given the emphasis on budgets and mandates, the absence of such analysis is striking.

TABLE 1: NOMINAL BUDGETS BY STATISTICAL AGENCY, FY 1992-1999

Fiscal Yr.	1992	1993	1994	1995	1996	1997	1998	1999
Census	$125.3	$127.4	$133.8	$136.0	$133.6	$135.0	$136.7	$146.1
BLS	300.6	323.9	343.0	351.3	343.1	360.8	380.5	398.9
BEA	34.7	41.7	39.8	42.2	40.5	40.9	42.5	43.1
SOI	26.8	27.2	23.9	28.8	26.1	24.3	25.8	28.8
NASS	82.6	81.0	81.9	81.0	81.1	82.7	81.8	80.4
ERS	58.9	58.9	55.2	53.9	53.1	53.1	71.6	65.8
EIA	76.3	82.3	86.6	84.5	72.3	66.1	66.8	70.5
NCSH	79.7	80.4	83.4	81.4	77.5	86.0	84.6	94.6
NCES	77.2	77.9	77.9	80.9	78.9	79.7	91.0	104.0
BJS	22.1	21.4	20.9	21.4	21.4	21.4	21.5	25.0
BTS	---	6.8	15.0	17.8	19.4	24.8	31.0	31.0

Source: Council of Professional Associations of Federal Statistics.

Note: The figure for the Census Bureau above is for current programs and does not include costs for the decennial census.

Some Specific Cuts

The Bureau of Economic Analysis (BEA), which generates the primary economic data for the U.S., has begun making cutbacks. Some of the series that used to be contained in the National Income and Product Accounts (of which GDP is one), have been discontinued.

- BEA has eliminated its regional economic projections program used by state planners for tax purposes, public construction, and other long-term decisions. Further cuts will mean the loss of the regional economic data series themselves, which are used to allocate $100 billion in federal spending and may determine state funding levels under future block grant programs (Landefeld 1996).
- Estimates of pollution control and abatement costs have been cut. The Bureau made its last round of estimates in September 1996. Begun in the 1960s, these estimates have been central to important policy debates and decisions about environmental programs and regulations and their costs. The loss of such data may make it more difficult to defend cost-effective environmental regulations and complicate intelligent reform of rules that are especially burdensome (Landefeld 1996).
- If budget cuts continue, there are plans to cut collection of data on the operations of multinational corporations. These data explain how manufacturing is

distributed within U.S. companies between their domestic and foreign facilities (Wallich 1996, 32).

- BEA has turned calculation of the Leading Economic Indicators and Business Cycle Indicators over to the private-sector Conference Board.[9] The decision was forced on BEA by shrinking resources. It could have either cut back on data collection or on data analysis. Choosing the latter was not without a price. The charge for these series has risen above the reach of most individuals.[10]

The Census Bureau has also faced the prospect of cutbacks due to reduced funding (Riche 1996). Some of the types of cuts proposed have been:

- Data which do not feed directly into other important statistics, such as stand-alone surveys that do not support larger statistical functions mandated by statute. Other series at Census, such as those used by BEA to calculate national income accounts, cannot be cut without substantial harm to our most prominent national indicators.
- Marginal data series will face elimination. Examples could include special populations which are difficult and therefore expensive to count, such as migrants or the homeless, or peripheral regions like Guam and the Virgin Islands.
- Functions that the private sector can fulfill will be privatized.

The Bureau of Labor Statistics, the statistical agency with the largest annual budget, has also been forced to trim its sails lately (Wallich 1996, 32).

- BLS has stopped keeping track of the turnover of workers in manufacturing, which is important in looking behind unemployment figures to see what is happening to labor markets and productivity.
- The employment status of older women and long-term economic projections of employment within specific occupational groups are two series that are scheduled for elimination if budget tightening continues.

The Small Business Administration was never a major statistical agency. Nevertheless, during the 1980s, it provided very useful information about the

---

9 . Census, BLS, state universities, and other public agencies still collect the component statistical series, which remain widely available, while the Conference Board analyzes the data and their implications and charges for the analytical product. It might seem there has been no real loss of information to the public. Yet, the combining of components into time series, which is now the function of the Conference Board, is not merely a mechanical procedure. Judgments are made at many junctures. The issue is whether private interests might influence some of those judgments.

10 . Information the Commerce Department once made available to citizens at no charge now costs $125 from the Conference Board.

condition of small business and entrepreneurial activity. Much of that has been lost (Useem 1996, 26).

- From 1971 to 1991, the SBA supported a database containing basic information on the start-ups, finances, and dissolutions of small businesses. Since 1991, when support for that database ended, tracking small business has been close to impossible.
- One year after SBA's statistical office was phased out, the database was physically eliminated (erased) from the computer where it was stored. A potential baseline for future analyses was thereby destroyed. The loss of data archives is one of the hidden dangers of cutbacks in federal statistical offices.
- Other federal agencies do not provide adequate reporting on small businesses. Census and BLS datasets on the number of self-employed people differ by 20%. That category of worker is peripheral to the primary focus of their surveys. That suggests paying the extra money for data gathering by specialized agencies may be worthwhile.

Loss By Attrition - Declining Quality Of Data

Eliminating data series entirely is the most visible form of data loss, but it may not be the greatest threat. An even greater menace to the federal statistical system would be the gradual deterioration of the quality of data. Only a few experts would know the difference. The rest of the country, including Congress, would imagine that the system remained intact, when in fact it was providing increasingly misleading information.

If agency statistical budgets are cut and no data series are dropped, the following problems, invisible to most observers, will emerge:

- The frequency and sample sizes for critical statistics will decline, possibly to the point of compromising their integrity, a process that has already begun at BLS (Wallich 1996, 32). The analysis of trends will become less reliable. Variations within and between cities or counties will not be detectable. As a result, local governments will not be able to rely on federal statistics for planning purposes.
- Essential and long-overdue improvements will not occur. Investment in better statistical methods and information infrastructure (e.g., computers, databases and classification systems) and the development of new data sets and surveys will be postponed indefinitely. Statistical agencies will stagnate. They will be unable to attract talented young people. Lack of technical improvement is very difficult to recognize until deterioration reaches a crisis stage. It is much more costly to repair than to prevent.

- New data to reflect emerging priorities such as high-tech and service industries or critical environmental risks such as climate change will not be developed. Already, the lack of accurate information about the service sector of the economy is harming the ability of analysts to understand the true condition of the economy.
- Congress will not be able to evaluate the performance and accountability of social programs devolved to the states (e.g., recent welfare reform legislation).

Ultimately, the interplay of internal agency politics and the influence of interest groups will determine which government data survive and in what form. Several people interviewed for this report expressed hope that agencies will undertake a rational review of their programs and make strategic programmatic decisions. Yet, they worry that narrow political interests will instead drive decisions about what to cut.

Ironically, a lack of agency willingness to make tough choices presents the greatest risk of all. Deciding when certain data have outlived their utility is never easy. Worthy uses can be found for even the most obscure and tangential data series, and each one enjoys its own lobby. However, trying to satisfy all interests with a smaller budget pie could have disastrous consequences.

Outmoded Technologies Due To Insufficient Funds

Another effect of inadequate funding is inadequate equipment. Statistical agencies are hamstrung by antiquated technology and software of the sort long since mothballed in the private sector. For example, the BEA, responsible for producing the national income accounts, still uses 1970s-era mainframe computers. Despite the BEA's influential constituency in the business and academic community, and ample evidence of cost-cutting and internal reform accomplished at the behest of Congress, lawmakers have still failed to fund an upgrade of its computer system.

In addition to technological obsolescence, agencies are also constrained by outmoded and conflicting databases that prevent data sharing across programs. Even within an agency, different divisions have their own databases. These databases frequently cannot be cross-referenced because each uses different software programs and lacks common identifiers. This is less a function of money than of the failure of Congress to pass legislation that would mandate greater coordination of databases across agencies. Individual agency departments and programs have written their own computer programs, which often only a handful of people know how to use. One official told the story of an employee who went on maternity leave. She was the only one who knew the program for a particular survey, so it could not be revised in her absence. (For an example of the harm caused by outmoded equipment and software, see Appendix E regarding EPA information systems.)

Organizing and integrating databases and bringing them on-line is an example of government reform that would raise costs before it lowers them. New computer hardware and software represent additional up-front costs, which are difficult to fund in the federal budgeting and appropriations process.[11] Moreover, managing the conversion from old information systems to new ones is exceedingly labor intensive, and staff resources are often fully dedicated to maintaining existing systems for data collection, processing and dissemination. Yet, once completed, updated information systems would perform better and with fewer staff, thus saving agency resources for other activities. Cutting statistical agency budgets without investing in enhanced information systems is penny-wise and pound-foolish. While potential productivity gains are difficult to calculate, a Presidential task force in 1983 estimated a cumulative savings (for all Federal agencies) of over $4 billion from 1987-1990 through improvements in technology alone (Duncan 1987, 400). The savings today would surely be greater.

PRIVATIZATON AND ACCESS: DATA TO THE HIGHEST BIDDER?

The legal context

Even if the government produces valuable information, that is no guarantee that it will be made available to the public in a useful form and at reasonable price. As a matter of law, the public's right to know is official policy of the federal government, enforceable in court. Failure to make data available is a violation of the law. However, the law is not always observed. In a recent law review article, Gellman (1995) discusses agency circumvention of three areas of law that promote low-cost access to federal data: the Freedom of Information Act (FOIA), the Copyright Act of 1976, and OMB Circular A-130.

---

11. The process for procurement of new technology by federal statistical agencies can take up to five years, a lifetime for computer and telecommunications technologies. In a world of dynamic and rapidly changing technology, five years to obtain new programs and equipment perpetually consigns agencies to working with archaic tools that cannot meet the needs and expectations of data users outside government.

There are several reasons for the snailÍs pace of the procurement process. First, statistical agencies generally make budget decisions two years before the appropriation year, to allow time for Congressional approval and changes. Hence, it is impossible to have an accurate budget or plan for equipment procurement, since many technologies are nearly obsolete in two to three years. Second, the two-year lag increases as appropriations usually do not occur at the beginning of the fiscal year, but often after the year has already begun. Third, vendors cannot be solicited until funding has been allocated. If the appropriation of funds happens in the middle of the year, choosing a vendor must normally wait until the next year and next budget plan. Finally, even selecting a vendor and purchasing equipment is a convoluted process. The solicitation of vendors is done by publishing an announcement in the *Federal Register*. The choice of vendor is then announced to the public. Once decided upon, the process of actually procuring the product takes additional time, as does training staff for the new equipment.

The most basic law protecting access is FOIA. It requires that any records produced by the government, whether published or not, be made available to researchers and the media at the cost of duplication (5 U.S.C. § 552(a)(4)(A)(ii)(II)). Those who seek information for commercial use are supposed to pay the cost of searching for and reviewing requested records as well.

The Copyright Act of 1976 is also intended to increase public access to data produced by the federal government. It permits citizens to use information produced by the government without restriction. Unlike most European countries, the U.S. government is expressly prohibited from copyrighting any of its works (17 U.S.C. § 105 (1988)).

OMB Circular A-130 (need full cite, revised February 8, 1996) is even more insistent on the importance of making information available. It states:

> The free flow of information between the government and the public is essential to a democratic society. . . . Because the public disclosure of government information is essential to the operation of a democracy, the management of Federal information resources should protect the public's right of access to government information. . . . Agencies shall: (a) Avoid establishing, or permitting others to establish on their behalf, exclusive, restricted, or other distribution arrangements that interfere with the availability of information dissemination products on a timely and equitable basis; . . . . (c) Set user charges for information dissemination products at a level sufficient to recover the cost of dissemination but no higher. They shall exclude from calculation of the charges costs associated with original collection and processing of the information. (Section 7, paragraphs c and f and Section 8, subsection a.)

The final sentence is extremely important. It expressly forbids agencies from pricing information high enough to recoup the cost of producing it.

The language and spirit of all of these laws have been violated by agencies. With respect to FOIA requests, some agencies routinely drag their feet, provide data in an inconvenient form, or impose excessive charges. Nevertheless, violation of FOIA is not a primary concern of this report since the key issue is no longer a question of disclosure versus nondisclosure. Instead, the crucial conflicts pertain to relative denial of access by making data inconvenient or expensive. With respect to copyright law, Gellman (1995, 1004-1005) describes how federal agencies often behave as if they had copyrighted their products, particularly electronic data, through "license agreements, royalties for use of data, restrictions on redisclosure of information products, and denial of access to digital versions of publicly available data."[12] These and similar practices also violate the clear intent of OMB Circular A-130.

---

12 . There is some ambiguity regarding copyrights of databases produced under contract with the U.S. government if the court regards the process as a subterfuge whereby the government indirectly obtains what amounts to a copyright (Gellman 1995,

Access is thus never guaranteed by statute. It is assured only by the vigilance of groups like the Taxpayer Assets Project, OMB Watch, and the American Library Association. Although many of the cases of undue restrictions on data discussed below were ultimately resolved in favor of free and open access, future access will never be entirely secure. Federal agencies are likely to treat data in proprietary ways, and private interests (the Information Industry Association) will continue to press for laws that enable them to reap private gains from public expenditures. In order to counter these practices, it is important to understand in some detail why the federal government has a positive duty to make data accessible to the public.

The peculiar nature of information

There are three features involved in collecting and disseminating information, including statistical data, that require that it be treated in a manner unlike other services performed by government.

First, reliable and impartial information, produced with confidentiality, regularity and continuity is essential in a democracy. Even if information could be more efficiently produced and disseminated by private companies, the importance of having a relatively impartial and trustworthy source of economic and social information might necessitate a federal role.[13] This is true even for specialized data that only a handful of people use, if the availability of that information serves the public indirectly. An open society depends on reporters, researchers, and public advocacy groups acting as watchdogs for the rest of us. The most important information for the preservation of democracy is the material they depend on to hold institutions accountable.

A second reason for a strong government role in producing and disseminating data is that, unlike most commodities, the benefits of information are generally not

---

1023, fn. 101). However, under legislation passed in 1998, data produced by universities and other nonprofit organizations (but not contractors) with grants from the federal government are subject to FOIA requirements. (See Federal Register, Vol. 64, No. 23, pg. 5684 ff.)

13 . According to the National Research Council (1979, 65), 31% of the population believe that government is most likely to get accurate information through surveys, 12% think universities will, and only 5% believe private companies will.

Even though business and industry may complain about the reporting burdens of federal surveys and other data collection, the private sector has faith in government's ability to protect the confidentiality of highly-sensitive information provided by individual firms. The National Research Council study (1979, 68) found that among citizens who think there is a difference, nearly four times as many trust government agencies over private companies to keep survey responses confidential. If respondents (particularly businesses) feared that the confidentiality of their responses might be violated, they would be reluctant to provide the information that makes economic surveys useful. Federal penalties for agency employee misuse of privileged data are severe, but private information companies are not similarly bound. Consequently, response rates for federal surveys are generally 85 to 95 percent, compared to far below 75 percent for surveys from information companies (National Research Council, 1979, 157). The response rates for a sample of federal surveys, drawn from Appendix III of the *Statistical Abstract* are as follows: for the Annual Survey of Manufactures and the Annual Surveys of State and Local Government, 85%, for Department of Education surveys, 87% to 92%, for the National Crime Survey, 93%, for the Current Population Survey, 95%, for the National Health Interview Study, 96%, and for the American Housing Survey, 97%.)

diminished by being available to many people simultaneously.[14] In other words, it is a quasi-public good.[15] Like listening to music, one person's benefit from access to information does not diminish the benefits to other people. In fact, each person benefits when other people are well-informed. If information is treated as a private commodity, the reduction in access harms not only those who are directly affected but the entire society.

Third, the cost structure of information production makes it a natural monopoly. Economic theory suggests that goods and services should be priced according to marginal cost--the cost of providing one more unit of the item. A private company cannot sell a commodity at marginal cost if that price would be lower than its average cost. It would lose money. Since information generally has high average costs (associated with collection) and very low marginal costs (associated with dissemination), companies sell information only if they can maintain a monopoly and sell it at or above average cost.[16] That means information is undersupplied by private markets. Economic theory suggests that information should be therefore be provided by (or subsidized by) government. On this basis, government has an affirmative duty to expand, not reduce, its role in the collection and dissemination of data, even to the extent of buying (or subsidizing) privately created electronic

---

14 . This is not always true. Some information is valuable precisely because it is exclusive, such as a trade secret or information that will give prior knowledge of a commodity price. Hayek (1945) and Machlup (1962, 18-19) distinguished between two types of knowledge: 1) information that is sensitive either to time or place and 2) information about enduring facts, ideas, and principles. This distinction is approximately the same as the difference between information as a private good and as a public good.

Some information passes from being type 1 to type 2 in a short time. For example, information about the geology of a site is worth a great deal of money before companies bid on the mineral rights, but it has no commercial value after the site has been mined. (This is why geological information is exempt from the FOIA.) Timely agricultural reports have value in the short run because of their influence on commodity futures, but soon they have value only as general knowledge of trends.

In these and other cases, the government might reasonably sell the first type of data for short periods as a commodity, then release it at no charge to those who want the data for research purposes when it becomes the second type. Or the government might release the time-sensitive or place-sensitive information freely to everyone at the same time in order to promote competitive bidding for actual commodities.

The access issues in this report are not about the procedures for selling or releasing time-sensitive or location-sensitive data during the period of its greatest commercial value. Instead, the focus is on data of the second type, which provide general information about social and economic conditions. In the case of geological and agricultural information, for example, the public has a right to know whether lease sales have been yielding adequate revenues and whether commodity markets operate efficiently and equitably.

15 . It is not a pure public good in the standard definition used by economists because it is possible to exclude people from it. National defense is the typical example used of a pure public good because it is not possible to exclude any member of a society from the benefits of collective security. Information is, however, much more clearly public (beneficiaries not being excludable) in this sense than either medicine or education in that the attention paid by a doctor or teacher to one person is likely to diminish the attention paid to someone else.

16 . Private monopolists and government agencies functioning as monopolists tend to sell information according to its value to users, not the cost of dissemination. Thus, they might charge $10,000 for access to a database, even though it would cost only a few dollars or less to provide it to an additional user on the Internet.

databases and making the information available at marginal cost.[17] There is no clearer case of a natural monopoly than information.[18] By permitting private monopolies to charge high (average cost or value-based) prices for many types of information, the government needlessly impedes its flow.

Monopoly privatization

In this report, we refer to agency practices that create a private, monopolistic, financial interest in data produced at public expense as "monopoly privatization" or more simply as "privatization."[19] This can occur either when a federal agency overprices data (particularly in electronic form) or when it contracts with a single private entity to disseminate data and allows that entity to charge monopolistic rates.

Privatization does not include adding value to widely accessible public data and selling it in a new form. There are several forms this value-added service can take:

- Customizing data, software or products produced by federal agencies in order to make them applicable to specific users;
- Combining data from private industries with federal data;
- Providing on-line access to public data, combined with indexing and processing facilities or superior software; and
- Offering consulting services to help clients understand and use the data. (Starr and Corson, 1987)

In principle, it should be possible to separate the form in which data is presented (a patentable service) from the raw data provided by the government, but in practice there are ambiguities that have been used to justify monopolistic behavior. If a company gains an exclusive right to issue documents in a particular form, it may extend its claim of ownership from its format to the public data as well, as in the case of West Publishing Company's monopoly control of legal citations.[20] This issue

---

17 . The merits of buying or subsidizing privately created databases apply only to data not already collected in some form by the government. There is no legitimate basis for the government to buy information that it has produced, which has happened several times in the past. Or as Harders (1995) put the matter: "Why should citizens pay a private vendor for public documents, which in every sense they already own?"

18 . For a readable, yet sophisticated, treatment of the correct pricing of information (dealing with joint production, economies of scale, economies of scope, and many other factors), see Love (1995a, chapter 10).

19 . The ideological debate over privatization is vacuous unless a crucial distinction is made between a) creating a competitive market to break a government monopoly and b) transferring a government monopoly into private hands. The former is generally desirable, but the latter is not. The critique in this report is solely about the privatization that occurs with natural monopolies: services like information collection and dissemination that have high fixed costs and extremely low variable costs.

20 . For example, West Publishing has long claimed a copyright over legal decisions in the United States. West considers the pagination used in published legal decisions and the editorial comments it adds to judges' decisions to be "value added." In effect, because the West page numbers had become the standard citation used in most legal documents, West asserted ownership of the

should be no more difficult than many others that courts have handled. It should be possible to distinguish proprietary features from public ones and to continue publishing public data without infringement of copyrights. Still, government agencies should ensure that future contracts for value-added services (such as indexing and online search capabilities) protect the rights of the public so the government information does not become monopolized, forcing citizens to buy from a private company what they already paid for in taxes. The aim of policy should be to encourage innovation and competition in private systems of delivering and indexing information, without infringing on free access to the information in its raw form.

The history of monopoly privatization

The long tradition of free information | For many decades, government information was available to any citizen who had access to one of the 1,382 depository libraries in the United States that house government documents. Federal agencies were required to provide their publications to the Government Printing Office (GPO), which in turn sent them to the depository libraries. For documents on paper or microfilm, this system provided truly open access, and depository libraries remain an important public asset.[21] Any individual, regardless of place of residence or income, could read documents produced by the federal government. Publicly created information was treated as a public good.

Starting in the 1970s, two trends began that modify the old system. The first was the shift to electronic databases, which created a large difference in value between paper documents and electronic data. When government agencies discovered they had a marketable commmodity, they began to think of the revenue-generating potential of acting like monopolists. Second, an ideological shift popularized the idea that government operations should be privatized, an attitude that runs directly counter to the law and tradition of public access.

Avoiding "competition" with the private sector | Although the Carter Administration began work on deregulation and privatization, the more serious effort to turn the

---

entire case law of the United States. Based on West's copyright claim, courts prevented the use of the West page breaks in computer databases. In 1992, West defeated proposals before Congress and the Administrative Office of the U.S. Courts to establish a public domain legal database and new citation system. The fact that people affiliated with West gave $738,000 to the Democratic National Committee over a five-year period may have influenced those outcomes. (Love 1995b; Hansen 1997) In 1995, the Minneapolis Star Tribune broke a story revealing that West had paid for the junkets of a number of judges, including U.S. Supreme Court justices, who later decided cases in favor of West (Schmickle and Hamburger 1995a, 1995b, and 1995c).

21 . "The general public uses these [government] documents; in fact, they really use them. There are 8.6 million visits per year (167,000 people each week using depository libraries). These figures are impressive and constitute an enormous, on-going system of access. The American Library Association states that libraries save business leaders, scientists and engineers an estimated $10 billion a year in locating necessary information" (Thompson 1994, 84).

functions of the federal government over to the private sector took place in the Reagan years. Under Reagan, an Office of Information and Regulatory Affairs (OIRA) was created within OMB. Its mission was to cut back on regulations that burdened business, including surveys and information-disclosure requirements.

In 1985, OMB issued Circular A-130, "Management of Federal Information Resources." It called on agencies to leave the field of data dissemination as much as possible to private companies. They were supposed "to ensure that major information systems do not 'unnecessarily duplicate' systems available from the private sector, and to place 'maximum feasible reliance' upon the private sector for the dissemination of products and services" (Love 1992, 399, citing OMB Circular A-130, 50 Federal Register 52730, December 24, 1985). The idea was for the federal government to turn all of the profitable activities over to private vendors and to keep the unprofitable work for the government. This suited the large companies (such as DRI, McGraw Hill, West Publishing, LEXIS, and Dun and Bradstreet) that comprised the Information Industry Association (IIA). They have espoused the philosophy that government should keep its hands off data dissemination, even the data that government produces. This logic once led Paul Zurkowski, former executive director of the IIA, to call free information provided by the government "an iron curtain ... descending across the competitive marketplace of ideas" (Berry 1975, 795, cited in Starr and Carson 1987, 434-35).[22]

Circular A-130 was revised by both the Bush and the Clinton administrations. It now calls on federal agencies to comply with the Paperwork Reduction Act (PRA) of 1995 (first enacted in 1980) by charging no more for information products and services than the "cost of dissemination."[23] This fundamental reversal of federal policy (from promoting private monopolies to low-cost public access) represented a great triumph for the public's right to know.

WIPO: a treaty that would have created monopolies | The greatest potential threat to the public interest that has so far emerged is the possibility of privatizing entire data sets that have been produced by public bodies. It is even conceivable that the meteorological data provided by government weather satellites could be privatized, forcing all weather reporting to pay the company that gained that control.

---

22 . The IIA members continue to defend their data monopolies in the name of freedom and competition. "The large commercial data vendors, through the Information Industry Association (IIA), have sought language that would prohibit federal agencies from creating new information products when the private sector already had similar products on the market" (Love 1995e).

23 . Among other things, the Paperwork Reduction Act (PRA) of 1995 seeks to limit the pricing of government information to the costs of dissemination (whether distributed by federal agencies or by private businesses). In 1996, West Publishing (along with other private information packagers) attempted to have a clause (subsection 3518(f)) inserted into H.R. 830, which would have eliminated the rights of the federal government to any information or database it produces itself if anyone added value to it (Love 1995e; Obey and Eisele 1996). Fortunately, public interest groups prevented this egregious effort to lay claim to public domain.

An international treaty proposed in December 1996 by the World Intellectual Property Organization (WIPO) would have opened the door to the wholesale transfer of public data to private control. Although the treaty was not adopted, the challenge to the public domain embodied in it remains a threat. This treaty would, among others things, have restricted the public's right to use information stored in databases.[24] It would also have transfered control of some public databases to private hands and restricted access to government data that have been used in the construction of private databases.

The ostensible aim of the treaty was to protect the rights of private database owners, but it went far beyond legitimate protections.[25] It raised the possibility of massive privatization of public information by allowing anyone with proprietary elements in a database to claim ownership of the nonproprietary elements as well. According to Shari Steele (1996), Staff Attorney for the Electronic Frontier Foundation, "Under the treaty, private monopolies for the maintenance of public data will be sanctioned, and individuals will have to pay for facts in the public domain."

Domestic legislation could achieve the same effect as the treaty. A bill introduced in Congress in 1996, the Database Investment and Intellectual Property Antipiracy Act (H.R. 3531), would have given copyright protection to any database that includes privately gathered information. That is legitimate. The bill would also have allowed copyrighting of public data that has been combined with privately gathered data. Publication of federal data that have already been published as part of an earlier private dataset could land the second publisher in court. In other words, the first publisher would be granted an effective monopoly over data in the public domain.[26]

---

24 . If passed, the database language in the treaty would expand monopolistic protections on government data internationally, making exchange of data across countries more expensive and difficult. The increased costs of information in fields such as science and medicine, where international exchange of information is vital to research and development, would be damaging. Universities and research centers would be similarly hindered in their ability to perform research. Not surprisingly, much of the library, scientific, and research community is up in arms about WIPO.

25 . Part of the impetus for this treaty is the inadequate protection of private databases under U.S. copyright laws. Current laws do not cover databases containing compilations of different public datasets. There does exist a lead time which allows companies to recoup their value-added investments in data. However, once data have been disseminated to the public in any form, no legal deterrent exists for others who then take the database contents, modify the format, and resell it, without paying the original compiler any fee. New technologies have greatly simplified the copying of databases. This leaves little incentive to invest in data improvements because there is little to discourage the copying of the data by competitors and customers alike. Thus, there is a legitimate basis for action to protect private interests, but the proposed treaty is excessive in giving rights to private database owners.

26 . According to Reichman and Samuelson (1996, 18), the bill would have prohibited extraction, use and reuse of whole or substantial parts of a database in which a vendor has made significant investments in the collection, assembly, verification, organization, or presentation of the database contents. The bill would have explicitly excluded "a database made by a government entity" but it later asserted that "any database otherwise subject to this Act. . .is not excluded because its contents have been obtained from a governmental entity."

Although this legislation was not enacted, it represents the sort of efforts toward privatization that will have to be guarded against.

The situation today | Privatization will continue to be a threat as long as private companies and government agencies find in their interests to create information monopolies. Thanks to the concerted action of groups such as the Taxpayer Assets Project, the American Library Association, the American Newspaper Publishes Association, and OMB Watch, some of the more egregious cases of monopolization have been thwarted. The issue has not been laid to rest, however. It is likely to arise repeatedly in the future.

Three forms of monopoly privatization of government data

In order to understand precisely how monopolistic privatization functions, it is necessary to examine the various forms it takes. One form occurs when the government acts like a private vendor and charges an inappropriately high price for data. The second and most obvious form of monopolistic privatization involves granting a private company the exclusive right to sell government data. Third, privatization can take the form of inaction, as when a government agency favors a private monopoly by failing to collect data that would compete with the private company.

Excessive prices of government retail data | Federal agencies have been under continual pressure to cut their budgets in recent years. As a result of that pressure, combined with a growing ideology that government should operate like a private firm (i.e., a monopolist), some agencies have sold data at commercial rates that are supposed, by law, to be available for a nominal fee.[27] In effect, agencies have functioned like businesses that charge monopoly rates.[28]

James Love, director of the Taxpayer Assets Project, which actively promotes public access to information, describes a case of which he had personal knowledge (Love 1992, 398-99). A student at Princeton needed the "call reports" of the Federal

---

27 . In fact, according to federal law, unless authorized by statute to retain user fees for its programs, an agency is required to deposit any receipts from the sale of services in a general Treasury account. Thus, selling information at a high price should not help an agency, unless the agency fails to follow these rules. Gellman (1995, 1042, fn. 205) cites as authority for this statement 31 U.S.C. §3302(b) (1988) and General Accounting Office, II *Principles of Federal Appropriations Law 6-105 (1992) (OGC-92-13)*.

28 . Several of the following cases involve the sale of data by the National Technical Information Service (NTIS), which is a branch of the Department of Commerce. Unlike the Government Printing Office (GPO), NTIS is funded entirely through user fees, which means that NTIS usually charges more for information than the GPO, but less than private sources. Because it fails to cover its costs on some items, NTIS charges exhorbitant fees for others. Thus, NTIS pricing often conflicts with the general federal policy of making information freely available. The mere fact that NTIS is required to cover its costs with fees does not justify current policies.

Reserve in order to conduct a study of the liquidity of the banking system in the aftermath of the savings and loan scandals. Until at least 1986, those reports were available free to researchers. In 1991, however, the student was told he would have had to pay $20,000 to obtain the ten years of quarterly data required for his project. (The cost of the magnetic tapes for the entire series would have been something like $600.)

Several EPA databases were at one time treated as profitable commodities even though the information in them was of importance to every citizen. Two databases dealing with hazardous wastes were on sale to the public on CD-ROM from the National Technical Information Service (NTIS) at a rate of $480 for a single issue, or on an annual subscription rate of $1,440 a year for three issues a year. These databases were and still are available on-line commercially through the Chemical Information System (CIS). For on-line access to this database, CIS charges a rate of $130 per hour, prorated down to the second, plus a $300 annual subscription fee, which enables access to all CIS databases. Few nonprofit organizations and community groups could afford the rates charged by either source for access to environmental information affecting their regions and communities. That is why the Right to Know Network has now made them available on-line at no charge.[29]

NTIS also charges a large amount for databases dealing with public health issues. The Vaccine Adverse Event Reporting (VAERS) database, produced by the Food and Drug Administration costs $1,080 for two diskettes (Love 1995a, 132). This would suggest that NTIS charges public health agencies large amounts for essential information because they are willing to pay. If information about vaccines were simply a private concern, that might be reasonable, but given their importance in public health, it appears to be a rather perverse pricing policy.

In another case involving highly significant public documents, the Government Printing Office, was mandated in 1993 to make the Federal Register and

---

29. The two databases which contain information on hazardous waste inventories are the Resource Conservation and Recovery Information System (RCRIS) and the Biennial Reporting System (BRS). The RCRIS database compiles information regarding operation, classification, permitting and violations for handlers of hazardous waste materials. The BRS database, on the other hand, holds information on actual wastes generated. The Right to Know Network now makes these two databases available to the public at no charge through their Web-site [http://rtk.net]. RTK receives the databases for free from EPA on magnetic tape and directly from EPA's mainframe computer and makes them available to the public in a user-friendly format. However, given the fact the information is collected at taxpayers' expense, the availability of such data should not be dependent upon the willingness of the non-profit sector to provide such information to the public.

Another EPA hazardous material database is also available on-line from CIS at $80 an hour, prorated per second (a figure that comes directly from Chemical Information System, 810 Gleneagles Court, Suite 300, Towson, MD 21286). This database is called the Oil and Hazardous Material/Technical Assistance Data System--OHM/TADS. It contains information on 1,402 different hazardous materials, including physical, chemical, biological, toxicological and commercial data. As with the RCRIS/BRS database, there is also the $300 annual subscription fee to CIS. NTIS also provides this data on seven diskettes, for $322.

Congressional Record available on the Internet.[30] The GPO did so in 1994, but at a high price that would have deterred the average citizen from making use of either set of documents. The online versions of the Federal Register and the Congressional Record were then sold only a subscription basis. The cost was $30 per month, $200 for six months, or $375 for one year.[31] In 1995, as a result of pressure from citizens and advocacy groups, this policy was reversed. Now GPO offers both the Federal Register and the Congressional Record online at no charge.

A final case that indicates the connection between budget issues and fees for public data is the plan by the U.S. Department of Education to charge for access to ERIC, a database of education literature. For a number of years ERIC was provided to the public by a contractor at the cost of reproduction. Both commercial and non-profit providers purchased the tapes from the contractor and offered the service online. Congress annually appropriated $7 million to maintain the system. Starting in 1991, DOE proposed to modify the contract, allowing the contractor to copyright the database and charge commercial providers a fee beyond the cost of reproduction. The $200,000 to $300,000 in additional revenue was to be used by DOE to improve the database beyond what was possible with the existing appropriation. DOE claimed that external funding would be needed to compensate for a decline in expected funding in fiscal year 1993. Congress blocked this plan by statute in 1992. If the DOE plan had become a precedent, other agencies might also have begun charging royalties in the name of economic necessity.

Low wholesale prices to private monopoly retailers | The second form of privatization has involved selling or giving data in bulk to private companies that earn monopoly profits by retailing the information at high prices.

An early case that demonstrates how this process works involved the sale of BLS data by Date Resources, Inc. (DRI), a commercial vendor. Until recently, the public had access to the the BLS labor statistics database (LABSTAT) in one of two ways. On the one hand, a researcher could pay DRI $75/hour connect time, and 45 cents per number downloaded, with a minimum annual charge of $1,200 (1989 prices). On the other hand, he or she could buy computer tapes from BLS, write a computer program, and buy time on someone's mainframe. DRI liked this arrangement. Although technically anyone could get the data from BLS, the transaction costs of doing so were so high as to give DRI a virtual monopoly. Joseph Kasputys of DRI said in a 1980

---

30. This was mandated under the GPO Enhanced Electronic Access Act of 1993.

31. As James Love (1995a, 129-30) summarizes the situation: "The GPO's failure to provide a method of pricing that allows episodic users to access the database represents an alarming misunderstanding of the purposes for which the GPO Access program was enacted. Congress clearly intended the GPO Access program to enhance public access to important collections of public documents which are vital to a democracy. The GPO personnel deliberately designed a pricing system which will only serve a handful of large organizations. . . ."

interview that he supported "the role of NTIS as a 'one stop' wholesaler of such government tapes," but he "caution[ed] against NTIS acting as a retailer of these data" (Duncan 1991, 319, cited in Love 1992, 405). In other words, Kasputys did not want any competition from NTIS that would have broken DRI's near-monopoly. Kasputys made these remarks after BLS had backed away, under pressure from DRI, from a plan to put LABSTAT on-line in an accessible form through NTIS.

In 1987, the Federal Maritime Commission (FMC) planned to save itself and shippers money by creating an electronic filing system for ocean shipping rates. The FMC planned to develop an on-line filing and retrieval system for ocean tariffs, the Automated Tariff Filing and Information system (ATFI). This move enjoyed support from shippers, who supplied the freight rate information to begin with, as they found the rates charged by TRANSAX/RATES (the commercial vendor, owned by Knight-Ridder, who published the information on-line) to be extremely high. When the Commission announced its plan in the *Federal Register*, TRANSAX/RATES lobbied Congress and secured a rider on FMC's appropriation that forced FMC to wholesale its database and establish "reasonable controls on the system to limit remote access usage by any one person." As in the case of LABSTAT, the commercial vendor was able to sustain a virtual monopoly by ensuring that the government sold its information only in bulk (on magnetic tape)so that individuals could not use it (Love 1992, 403, citing P.L. 101-92 (1989)). In a strange twist of fate, Congress reversed course in 1992 and, as a budget balancing maneuver, required all indirect users of the FMC data (i.e., users of commercial on-line data) to pay 46 cents per minute for their access time. In effect, Congress decided to collect the value of the information rather than allow TRANSAX/RATES to be the sole beneficiary. That still violated the spirit of free (or marginal cost) access to information, but at least it reduced the burden on taxpayers.

The National Agricultural Library (part of the U.S. Department of Agriculture) produces a database entitled AGRICOLA. Silver Platter, a commercial vendor, puts that database on a CD-ROM, then charges $825 per year for a single workstation subscription. Backfiles cost an additional $950 (on a one-time basis). Citizens are forced to pay three times: once to cover the cost of creating the database, second to pay the fees so that USDA employees can access the database, and third to access the database as individual researchers. The initial cost of mastering a CD-ROM is only around $1,500. Each additional one can be produced for a few dollars. By making this database available for the cost of dissemination (a few dollars per CD-ROM or free on the Internet), USDA could save both itself and researchers millions of dollars (Love 1995a, 125).

Until 1993, the only source of electronic data from the Securities and Exchange Commission (10K reports and other company financial records) was the EDGAR database, which was operated by Mead Data Central (operators of LEXIS/NEXIS)

under contract to SEC. Access to the database was limited and very expensive. As a result of a public access project, undertaken by the New York University Stern School for Business (funded by the National Science Foundation), the data were made available to everyone on the Internet without charge for two years as an experiment. In 1995, a bill was introduced that would have privatized EDGAR once again. As a result of public pressure, however, the free Internet dissemination program was continued, at least for the time being (Love 1995c; Love 1995d). Achieving this goal was a breakthrough that could serve as a model for many other federal databases that remain available only to those who can afford to pay high fees to private vendors.

Protection of private monopolies | In addition to raising the rates on data itself, agencies have avoided collecting or disseminating data in order to protect an existing private supplier of data. In effect, the government has sanctioned private monopolies and the high prices that result from them. In protecting the interests of a few data suppliers, the government has ignored the interests of the general public.

One example of a protected monopoly is building permit data. The Census Bureau provides data on the number and value of building permits issued in cities and counties. But it does not provide data on the area (square feet) of building permits that have been issued, which is a more relevant factor for predicting future building patterns in an area. As of 1988, those data were available only from the F.W. Dodge Company for $18,826 for the current year and $9,413 for each previous year. The same data could be collected by the Census Bureau at virtually no cost since it would require adding only one question to an existing survey. Ironically, Census has to pay F.W. Dodge for these data when it needs them. Census does not collect these data because it wishes to protect the Dodge monopoly (Love 1992, 406)

Presumably there are cases similar to the building permit data, but they are the hardest cases to discover. Like the dog that did not bark in the Sherlock Holmes story, the process is invisible when the government covers for a monopolist by not collecting data. Moreover, since there is no positive duty for the government to collect all data that might be of value, there is no legal recourse even if this sort of relationship is exposed.

Conclusion | Fiscal pressures, coupled with an ideological desire to downsize government, are likely to encourage efforts by federal agencies to treat their databases as commodities to be sold on the basis of value rather than the cost of distribution. If that happens, the gain to the overall federal budget will be trivial, and the loss to the ideal of self-governance will be immense.

An informed citizenry is essential to a democratic society, and information collected at the taxpayer's expense ought to be available to the general public at little or no cost. Thus far, there have been both successes and setbacks in preserving the

public's right to know. Only the addition of new voices to the debate on federal statistics can ensure that this democratic vision prevails.

DEVOLUTION: THE ROLE OF FEDERAL STATISTICS?

Devolution, or shifting authority for federal programs to the states, also presents a challenge to the integrity of federal statistics. As programs are transferred to states and local authorities, there will pressure to disband the federal statistical offices that have tracked and monitored the related federal programs. If that happens, uniform statistical series that can be used to make comparisons across states may simply be lost.

This process has already begun. Devolution has become a central theme of American politics. The proponents of devolution are ideologically committed to cutting the size and scope of federal programs in dramatic gestures. As a practical matter, though, devolution has been occurring on a piecemeal basis for years within specific programs, and states have been given greater authority to make decisions.[32] The next stage involves the wholesale transfer of authority for an entire program to states.

The ideal of shifting authority is full of irony and internal contradiction, however. Advocates of devolution must find a way to reconcile decentralized management with centralized data collection. Most proponents of devolution are pragmatic and not merely ideological: they favor devolution as a way of improving the performance of government. They want to turn power over to the states but they also want to be sure that programs are managed rationally. They want to monitor the performance of states. Those two goals are at odds with each other as long as the federal government relies on program records to evaluate performance.[33]

---

32. A recent survey by the Council of Governors' Policy Advisers (1996) reveals that, in light of expected reductions in federal funding, states are taking advantage of more flexible waiver provisions to assume greater responsibility for managing Medicare, Medicaid, education and environmental programs.

33. Using program records as the database to evaluate performance is a source of conflict. Uniform national standards in data collection might seem inherently contrary to the spirit of allowing states and local governments to decide how best to manage programs as long as program records are used. Yet, maintaining a national identity requires a minimal degree of information sharing. The fact that Arizona, Alaska, and Alabama might choose to spend federal money on housing programs in different ways does not mean the rest of the country does not want to know about the outcomes. Even on an international basis, there is a need for comparable statistics. If unemployment, saving, or inflation rates are to be compared among different countries, we want to be sure the same methods of measurement are being used. A desire for uniform statistics is not the same as a desire to control policy, but if statistical work is closely tied to programs, a tension will always remain.

In principle, the conflicting goals of independent state programs and uniform national reporting might be resolved by spending more money at the federal level to develop detailed social indicators that would measure the health and well-being of the population on an ongoing basis. If such measures were in place as devolution took place, there would be a feedback mechanism outside the specific program records that would test the efficacy of state policies. If malnutrition or illiteracy or homelessness started rising in a state after a program were introduced or modified, this would provide independent evidence that the program was not working. However, since that sort of information would be politically embarrassing to governors and legislators in poorly

If a program is turned over to the states, each will develop its own forms and procedures. Record-keeping will no longer be uniform across the states. Comparisons between states will be difficult, at best. On the other hand, if the federal government insists on each state continuing to collect data in a uniform manner, that will severely limit the capacity of states to experiment and develop their own approaches to social and environmental programs. The information-gathering tail will wag the programmatic dog. The conflicting desires for programmatic variation and uniform national data will create confusion and frustration.

Welfare Reform As A Test Case

The first major test case of that tension is the Personal Responsibility and Work Opportunity Act of 1996, which transferred authority for welfare programs to the states. Even as it gave freedom to states with one hand, Congress took away state autonomy with the other hand by imposing massive new reporting requirements that states probably cannot fulfill.[34] The law requires states to track the number of years each person spends on welfare, his or her work history, marital status, timeliness of child support payments, and other factors as he or she moves from state to state.[35] Aside from the invasion of privacy involved, the task may be technically impossible.

Tracking program participants over time and across different programs and political jurisdictions will necessitate the creation of new longitudinal and relational databases of a kind and complexity which do not currently exist. The tough child support enforcement provisions call for several new, easily updatable databases, including federal and state registries of new hires to provide social service agencies with needed information. Finally, the legislation mandates several new studies. For example, a new national survey of children at risk must be longtidudinal, provide data for many states, describe out-of-home child placements, and establish the frequency with which children come into contact with state and local agencies (Brady and Snow 1996 and 1997). All of this is to be done without additional federal funding. (Appendix F provides a more complete discussion of the welfare reform issue.)

---

performing states, the federal government will not collect it unless there is strong public pressure to do so. Thus, the original conflict is not easily resolved.

34. The ultimate irony is that Congress mandated all of this just one year after supposedly limiting Congressional power to impose "unfunded mandates" (programs and regulations without federal funds) on state and local governments. In the statistical arena, welfare reform is full of unfunded mandates.

35. Robert Pear (1996) offers an example of the sort of case history that statistical agencies will have to track. A couple and their children receive public assistance in California for one year until the husband finds employment. They later divorce, the mother returns to welfare, and the father moves to Illinois. Authorities there must receive notification about the father's child support obligations and compel him to make payments. California authorities must have on record the mother's previous time on welfare to determine her current eligibility. If the father then moves to Arizona and the mother to Texas, Arizona and Texas officials must have access to their respective case histories from Illinois and California to ensure that the father continues child support and that the mother has not exceeded her five-year lifetime welfare limit. Such a scenario is not uncommon in our society of extreme mobility and a high proportion of single-parent families.

Lessons From Past Block Grants

The latest devolution craze is not the first time information management and data needs have been treated as secondary in a move toward federalism. In 1981, Congress consolidated 50 categorical programs into block grants.[36] (Block grants give wider authority to states than categorical grants in deciding how to spend money.) These block grants mandated data collection to ensure money was being spent for relevant purposes, but there were no uniform national standards for form and content.

Realizing that the lack of comparability of data among the states was a problem of block-grant programs, Congress soon responded by imposing uniform data collection requirements. At the same time, federal agencies developed cooperative arrangements with states to facilitate the collection of nationally comparable data. Such federal-state cooperation on data issues has become increasingly urgent as devolution proceeds. Nevertheless, according to one report by the General Accounting Office (GAO 1988, 3-4, 8-9, 15), national comparability of data suffered under block-grant arrangements.[37] That GAO report found block-grant data programs to provide the best, nationally-comparable data under the following circumstances:

- national agency leadership in supervising the methodology of data collection;
- recognition by states of the need for accurate data;
- federal statutes to encourage cooperation in data collection;
- federal funding for data collection;
- allocation of national staff to work with states; and
- involvement by state officials in program design.

These conditions can only be assured with sufficient federal funding for statistics. Yet, block grants weaken the programmatic justification of much federal data collection, thereby rendering statistical agency budgets even more vulnerable (McMillen and Spar 1986, 10). To make matters worse, the dispersal of authority among a dozen major federal statistical agencies (and sixty minor ones) makes a

---

36. Additional block grants have since been created. Prior to welfare reform legislation, 15 block grants distributed $32 billion in federal funding (PRC and COPAFS 1996, 6).

37. This study evaluated the performance of data programs associated with four block grants: Alcohol, Drug Abuse and Mental Health Services; Low-Income Home Energy Assistance Program; Community Services Block Grant; and the Education Block Grant. It confirmed federal-state cooperation under block grants as a viable strategy for obtaining national data on funding, services, and characteristics of program clients, and for reducing collection and reporting burdens on state governments. However, the study also pointed to weaknesses and inconsistencies in the accuracy and timeliness of data collected. Finally, it concluded that block grants do not represent a viable option for providing data for allocating federal funds to the states or for ensuring compliance with federal laws and regulations--both profound accountability concerns raised by the devolution of federal programs.

concerted effort to promote greater federal-state cooperation more difficult.[38] Finally, the experience of block grants has failed to clarify how to balance the desire of states for fewer statistical mandates with the broader public interest in uniform national data (PRC and COPAFS 1996, 7).

Successful Devolution: Lessons From Federal-State Cooperation On Statistics

Unless the federal government begins gathering data that can be used to construct social indicators for cities and even neighborhoods (see section below on local uses of statistics), administrative statistics will remain the only basis for judging devolution. Although the tension between programmatic autonomy and uniform data will persist under devolution, federal-state cooperative arrangements can promote comparability (if not uniformity) of statistics.

The Bureau of Labor Statistics (BLS) has the longest standing federal-state cooperative program in the federal statistical system (Ruddick 1996). Through written agreements, BLS relies on states to collect standardized employment data. This avoids duplication of effort and reduces the burden on the people who provide the data.[39] The long track record of this cooperative program makes it an important model for evaluating how to encourage further cooperation on statistics between federal, state and local governments. It is important to note, however, that this cooperation comes with a price tag: federal funding was $61 million in FY 1996. If that funding were to decline, states would seek greater autonomy, and differences in state definitions and standards would quickly emerge.

PRIVACY: A BREWING BACKLASH AGAINST STATISTICS

Another threat faced by federal statistics is the growing hostility to anyone who wants to track the activities of businesses or households. This concern for privacy is heightened because of the growing capacity of private agencies to invade people's private lives with various forms of electronic devices. Even if the federal government continues to protect citizens from misuse of its own surveys, it will likely experience fallout from the response to private surveys and infiltration methods.

---

38. Appendix A describes the multitude of federal statistical agencies, each of which has its own priorities, methods, and databases. Appendix B discusses some of the history of failed efforts to coordinate or consolidate the work of statistical agencies. For the time being, the diversity is likely to continue, which will retard efforts to cooperate more closely with state and local governments.

39. Agreements between BLS and states extend beyond data collection. BLS has, for example, worked with states to standardize software for shared use, and it has consulted states in the updating of programs and activities. In addition to employment statistics, BLS also cooperates with states in the collection of data about occupational incidents and illnesses by industry.

The safeguards put in place by the federal government to protect citizens already reduces the efficiency of statistics collection. The Justice Department, the Internal Revenue Service, the Census Bureau, the Immigration and Naturalization Service, the Veterans Administration, and other agencies must separately maintain raw data gleaned from surveys and other sources. Current law precludes development of master files that would allow agencies to share raw data that can identify individual persons or households, even for statistical purposes. This legislation has achieved its purpose of preventing disclosure of personal and proprietary information in the production of public statistics at the federal level. However, in protecting individual privacy, federal confidentiality statutes have had the unintended effect of balkanizing the data collection activities of federal statistical agencies and increasing their costs.

The same safeguards do not apply to the private sector. State-of-the-art information technologies have transformed mundane acts of daily life--placing an 800 number call, filling a medical prescription, submitting a product rebate, making a credit card purchase--into the feedstock for personal information databases that sustain industries ranging from direct marketing to insurance. Credit history, medical history, political affiliations, financial transactions with credit cards, and many other aspects of an individual's life can be obtained for a price. Employers use information obtained through credit checks in condsidering potential employees. If the information in these private databases is incorrect, correcting it can be a daunting task. Even when accurate, such information gives those with access to it enormous power over individuals (Rule and Hunter 1996, 18-19).

Americans are understandably fearful that large institutions will invade their privacy.[40] Although the Privacy Act of 1974 restricts public access to government data about individuals, there is no equivalent protection against intrusions by private companies that compile information about citizens. Recently, a Los Angeles television reporter purchased from direct marketer Metromail Corporation a list of 5,500 children, complete with family names and addresses. Worse, he successfully placed the order in the name of convicted child rapist and murderer Richard Allen Davis (Rule and Hunter 1996). Another example is the recent furor over the announcment that Lexis-Nexis was selling a commercial database called Ptrax which featured very private information on approximately three million Americans. Lexis-Nexis was inundated with thousands of e-mails and phone calls from irate citizens. As a result

---

40. In 1993, a Harris poll showed 83 percent of Americans concerned about intrusions into their privacy, more than double the 34 percent who expressed that concern in 1970. More than half of respondents were "very concerned" (Bryant and Dunn 1995). Bryant and Dunn also cite Alan F. Westin (no date): "Growing privacy worries are due to a general distrust of institutions and the government process, and to public fears about the misuse of computer technology."

Another report (NRC & SSRC 1993) recently concluded that "many citizens believe increasingly and with some justification that their privacy is being eroded by organizations that develop and control the use of large databases that contain detailed information about them.

of this history of privacy intrusions by private companies, citizens have begun to question even legitimate government surveys and become less inclined to take part (NRC 1979, 149).

Anti-government ideologues are using public anxiety about loss of privacy to feed suspicions about government. The proposed Family Privacy Act introduced in the last Congressional sesssion was aimed entirely at curbing government data collection. For example, it would have forbidden government statistical agencies from interviewing minors for surveys without parental consent. A GOP-led initiative, the Act was rich with irony. If passed, it would have interfered with critical surveys such as the National Household Survey on Drug Abuse at a time when Sen. Robert Dole and the Republican party had launched a broad attack on the record of the Clinton Administrations in curbing teenage drug use. The statistical ammunition used by Dole and his allies came from the very survey data the Act would have eliminated.

Meanwhile, the ideologues ignore obvious and growing threats to privacy posed by private-sector databases and argue that risks to privacy come from the gathering of data by federal statistical agencies, including the Census Bureau.[41] While some federal agencies may have illegally "snooped" on citizens, none of them belongs to the federal statistical system.[42]

Stirring up fears about the Census only serves to raise its cost and decrease its accuracy. The more households are reluctant to provide information when they first receive the forms, the more must be spent following up in personal, at-the-door interviews.[43] The number responding to the initial Census questionnaire fell from 75% in 1980 to 63% in 1990 (GAO 1995b, 5). It appears this may have been caused in part by anxieties about government use of data. A Gallup poll in March 1990 indicated that just 67% of Americans felt confident that their privacy would be protected by the Census Bureau (Bryant and Dunn 1995).[44]

---

41. Most people make no distinction among survey organizations, whether public or private. They either trust all of them or none of them. Among those who do make a distinction, three times as many people mistrust private companies as the federal government in keeping data confidential--with state and local governments considered even more trustworthy (NRC 1979, 69).

42. If those critics were focusing their attention on the intelligence services (particularly the National Security Agency, which has a massive capacity for electronic surveillance of the population, and which is shrouded in secrecy), their concerns about Big Brother might be well-founded.

43. According to former Census Bureau director John G. Keane: "When the public attitude is not good, the census suffers" (Bryant and Dunn 1995).

44. Concerns about confidentiality of data are not new. In 1976 (at a point following Watergate, when public confidence in government was at a low ebb), NRC (1979, 70) researchers found that 18% of the public believed that an individual's responses to the census were open to the public, and an additional 22% thought they were open to other federal agencies. Only 9% thought the records were completely confidential. Most respondents said they did not know. Because of these doubts about confidentiality, a significant number of people have refused to answer the questions on family income on the Current Population Survey. In 1948, the non-response / inadequate response percentage was 7.5. In 1969, this had jumped to 19.0, and in 1971, it was 14.6. (NRC 1979, 145)

There was a time when concerns about census records were legitimate. During WWI and WWII, the Census Bureau cooperated with other federal agencies and violated the rights of individuals. (The worst abuse was providing the FBI with information about the neighborhoods occupied by Japanese-Americans so they could be rounded up and put in camps.) Since 1954, when confidentiality rules were tightened, no such disclosures have occurred.[45]

In Pursuit Of Ignorance: Politicization Of Statistics

A final threat to the integrity of the federal statistical system is the prospect of various interests politicizing the determination of what data are collected and distributed.

In a broad sense, data are always politicized. The choice of which data to gather is already a political matter because it involves a determination of which facts are socially significant. It is impossible to have a purely apolitical statistical system. Disputes among statistical experts about the best methods of sampling and analysis are essential to the continual improvement of the science of statistics. Partisan debates over the proper interpretation of statistics are a healthy part of the democratic process.

Maintaining a long-term, professional approach to statistics probably requires the sort of balance of factions that James Madison regarded as the hallmark of democracy.[46] If a single ideology becomes strong enough to control the apparatus of

---

Another reason for declining response rates to government surveys may be the growth of private surveys by marketing agencies. An increase in the overall response burden on individuals has reduced their willingness to cooperate with government statistical surveys.

45. As Bryant and Dunn (1995) explain (Bryant is former director of the Census Bureau): "Title 13 of the U.S. Code requires census information to be used only for statistical purposes, and it prohibits publication of the data in which any individual business or person can be identified. Thus, it limits release of census information to the aggregate level. Title 13 also states that no one other than sworn officers and employees of the Census Bureau and its parent Department of Commerce is allowed to examine individual survey and census reports. All sworn officers and employees, including temporary employees, take a nondisclosure oath of confidentiality. The penalties for breaching confidentiality are up to a $5,000 fine and / or up to five years' imprisonment. These penalties apply to former as well as present employees."

46. The relative stability of political factions in the United States in this century may be one reason why the federal statistical system has been allowed to become professionalized and reasonably immune from political pressure. No president could call up the head of the Bureau of Labor Statistics and demand that a new method of counting unemployment be instituted to hide a sharp rise in the official rate. The methodology of the crime rate has not been manipulated to create pleasing outcomes.

Politicians who tried tampering with the system in the past got their fingers burned. When President Nixon appointed people without statistical backgrounds to head the Census Bureau and the Commerce Department's statistical division, he was criticized for it (Innes 1990, 282). When he suspended briefings by BLS staff on the meaning of unemployment figures, the Joint Economic Committee in Congress held monthly hearings to give the BLS officials a platform to speak (Innes 1990, 133, 286-87). After a consensus formed around the definition of unemployment by the 1940s, political leaders have had to treat it as a technical issue and leave it to professionals. Other measures, such as the official poverty line, are also free of arbitrary modification by presidents, but until they attain the same degree of consensus as the unemployment rate, they are subject to ongoing conflicts between ideological groups.

data gathering and reporting, there is a danger that statistics will be modified to fit preconceived ideas. If one interest group gains too much leverage, data may be reported to conform to its aims.[47]

The narrow forms of politicization that are the subject of this section of the report involve subterfuge and intrigue. Thus, as used here, politicization is limited to efforts to intervene in the process of data gathering or reporting in order to achieve a goal that would not have political support if done openly.

The case studies in this section show how ideology and interest have been allowed in two ways to politicize federal statistics.

First, some data have been eliminated because they would reveal inconvenient facts. Opponents of active environmental policies, for example, have undermined the credibility of those policies by eliminating the information needed to justify them.

Second, the methods of collecting and processing information have been modified for ideological reasons or to serve particular interests but rationalized publicly in technical terms.

The Elimination of Statistics

First, we look at examples of the way in which Congress is trying to hide inconvenient facts (and sometimes succeeding) by discontinuing data gathering. In the section following this one, we will examine cases in which s tatistical methods are being manipulated for partisan purposes. In the case of the decennial census, both types of pressure are being applied.

Data On Biological Diversity | The frontal assault on the Department of Interior's National Biological Service (NBS) in 1995 typifies the elimination of data that might reveal ideologically inconvenient facts. Wil_dlife surveys conducted by NBS might have jeopardized some property developments by revealing new species to be endangered. Rather than directly attacking the Endangered Species Act, which enjoys widespread public support, opponents chose to attack the database on which listings could be based.

Interior Secretary Bruce Babbitt had envisioned NBS as an effort to consolidate disparate scientific, research and data functions into one agency and separate them from the regulatory responsibilities of agencies such as the U.S. Fish and Wildlife Service. The most controversial component of the NBS (and a tiny portion of its budget) was the proposed National Biological Survey, an effort to map the nation's

---

47. Ideology distorts the gathering of statistics whenever a government wants only positive news. In 1958 in China, for example, during the 'great leap forward,' economic planners at every level were implicitly encouraged to exaggerate their actual output. The result of this organizational self-deception was a catastrophe for the nation. (Li 1962, cited in Innes, Chap. 1) A similar process occurred during the war in Vietnam. Glowing field reports, intended to win favor with those at headquarters, exaggerated success and ignored failure.

biological resources and to identify more accurately the ranges and populations of threatened and endangered species. Babbitt saw the Survey as a planning tool to rationalize enforcement of the Endangered Species Act and prevent the economically costly train-wrecks that occur when listed species are discovered on land after development or other economic activity has already begun.

However, Congressional supporters of the property rights movement, funded by developers and oil, timber, and mining interests, opposed the Survey. If threats to wildlife were not recorded, support for the protection of species would diminish. Rep. Don Young (R-AK), Chairman of the House Resources Committee, summed up their view of government scientists: by describing them as "federal goons." The conservative Heritage Foundation reinforced that view, claiming that the NBS is "nothing more than an attempt to convert private land to public use without compensation" (Kenworthy 1995). When asked about the western GOP attack on our nation's ability to collect and analyze biological data, Rep. Ralph Regula (R-OH), chair of the House Appropriations subcommittee for Interior exclaimed, "Doggone it, this is democracy" (Morgan 1995).

While attacks on the NBS were not wholly successful, they did inflict considerable damage. In order to marginalize NBS, Congress ordered the Secretary of the Interior to merge it (under a new name--the Biological Resources Division) with the U.S. Geological Survey (USGS) and slashed its budget by over 20 percent in two separate cuts. USGS is widely viewed as unresponsive to broader Department of Interior needs. Moreover, an agency specializing in geology has no unique expertise in managing a broader program focused on biological sciences. Worse, the original conception of an agency that integrated Interior's scientific, data and research functions under one roof has been lost.

CDC Statistics On Firearms Injuries | At the behest of the National Rifle Association (NRA), Congress passed an ammendment in July 1996 to cut $2.6 million from the budget of National Center for Injury Prevention and Control (part of the Centers for Disease Control and Prevention or CDC) and to transfer this money to a program supporting rural health care centers. The amount of the cut equalled the cost of the Center's study which produced statistics showing rates of gun violence. The study in previous years had shown that 37,000 out of the 145,000 deaths from injuries are a result of firearms, and that guns are the second leading killer of black teens (Herbert 1996). (But see also Kates et al. (1995) and other articles cited in Bijlefeld (1997, part VIII) arguing that CDC has intentionally ignored contrary evidence.) Knowing that a forthright repeal of the Brady bill or the assault weapons ban would win little public support, the NRA successfully gutted collection of the data that has contributed to public awareness and concern over gun violence.

Representative Nita Lowey (D-NY) introduced an amendment to restore the $2.6 million to the Center, but the amendment was defeated. As justification for their advocacy of public ignorance, the NRA and its supporters stated that the CDC showed a bias in their study toward gun control. The NRA also stated that the CDC study duplicated studies of other agencies. However, Dr. Mark Rosenberg, director of the Center, stated that "in terms of firearm injuries. . . the CDC is the only agency that is trying to [collect this information]" (Herbert 1996). The Government Accounting Office also dismissed the NRA duplication charge as false.

OSHA Data On Repetitive Motion Strain | Repetitive motion injuries, commonly called carpal tunnel syndrome among office workers, represent the fastest growing category of workplace injury in the U.S., affecting up to 1,000 workers every day in data entry, assembly-line production, meat processing, grocery check-outs and other jobs where repeated motion is the norm. The Department of Labor's Occupational Safety and Health Administration (OSHA) reports on repetitive motion injury, which are drawn from data reported to BLS. In 1994, 332,00 cases of repetitive motion injuries were recorded, costing industry an estimated $20 to $100 billion a year.

Industry groups feared both increased regulation from OSHA and definitive causal evidence of a link between specific workplace activities and injury--the latter opening the door to lawsuits by affected workers. In order to head off both possibilities, major business interests such as the National Association of Manufacturers (NAM), the American Trucking Association, and the United Parcel Service lobbied to prevent the collection of relevant data. Congressman Henry Bonilla (R-TX) attached a 59-word "ergonomics rider" to a 1997 appropriations bill that prohibited OSHA from "recording and reporting occupational injuries and illnesses directly related [to ergonomic protection]." In 1995 Congress had already prohibited OSHA from issuing ergonomic regulations to protect against repetitive motion injury. Yet, powerful special interests remained irritated by what one high-level NAM official described as the "agency's arrogance" in continuing to study the issue (Suplee 1996). For the time being, the idea that society should acquire data in order to understand an emerging problem has prevailed over a campaign of willful ignorance. Moderate GOP members of Congress joined Democracts to pass an amendment by Rep. Nancy Pelosi (D-CA) which eliminated the rider (Lohr 1996).

No Air Quality Data - No Standards | The dangers of imposed ignorance are further exemplified by the growing debate on federal air particulate standards. In April 1997, television viewers were treated to the spectacle of C. Boyden Gray, Chairman of Citizens for a Sound Economy and former White House Counsel in the Bush Administration, chastising EPA for making scientific claims about the health effects of particulate emissions based on the agency's poor monitoring coverage. Gray's

disingenuous critique conveniently neglected to mention the fact that the Reagan Administration cut EPA's air quality monitoring capabilities in the 1980s. This is the politics of stealth. The press fails to remember such details and hold special-interest prophets of ignorance accountable.

The Manipulation of Statistical Methods

Some statistical series cannot be eliminated completely. In those cases, the ideologues have fought to modify the way in which the data are collected and analyzed to yield new results more to their liking.

Adjusting the Consumer Price Index | In 1995 House Speaker Newt Gingrich (R-GA) made a high-profile threat to cut BLS's budget if it did not adjust how it calculates the Consumer Price Index (CPI). Since the CPI is used to determine annual cost-of-living adjustments in entitlement programs, the Speaker sought to force changes in the methodology to slow entitlement growth. Something had to give on entitlements, if GOP commitments to cut taxes and balance the budget were to be reconciled. While a methodological case can be made for some revisions to the CPI, the rhetorical attacks on the Bureau of Labor Statistics foreshadowed the growing vulnerability of the statistical system to political interests in an era of diminishing fiscal resources.

Over a year later, the Advisory Commission to Study the Consumer Price Index (1996), chaired by Michael Boskin, issued a report that continued the politicizing of the CPI. Members of the commission had been chosen by the Senate Finance Committee to include only those who would argue that the CPI overstates the true rate of inflation. (The argument is based, in part, on the idea that quality changes in products are undervalued by BLS, which calculates the official CPI.) Although the Senate Finance Committee did not attempt to change the basis of calculation directly with legislation, it sent a message that BLS should reconsider its method of calculating the CPI or produce a supplementary "cost of living" index more in line with the Boskin methodology.

The Politics Of Census 2000 | Since apportionment of seats in the House of Representatives depends on the results of the decennial census, the political stakes in preparing for it have always been high. Census 2000 was no exception to the rule, but a few new elements were interjected.

Long Form Cuts First, Rep. Hal Rogers (R-KY), chair of the appropriations subcommittee that deals with the Department of Commerce (and thus the Census Bureau), proposed limiting Census 2000 to a pure head count, as required by the

Constitution.[48] If that concept were carried to the extreme, all of the questions on the "long form" would be eliminated.

Already congressional pressure caused the Census Bureau to cut questions on five subjects that were on the 1990 census form: children ever born (fertility), year last worked, source of water, sewage disposal, and condominium status. In the Survey of Census Needs of Non-Federal Data Users, around 2/3 of the respondents said they had used the data based on those five questions in local planning, and most knew of no other source of similar information (Census Bureau 1997b). Since the long-form questions provide information about housing, income, and other features of households that are needed to fulfill the requirements of certain federal laws, eliminating them would have caused legal turmoil. That information is also used by local governments in carrying out social policy and in planning the construction of schools and roads. (See section below on local uses of statistics).

Conflict Over Sampling To avoid the problem of undercounting, which was worse in 1990 than in 1980,[49] the Census Bureau planned to use sampling to estimate the characteristics of the "hard-to-count" and "don't-want-to-cooperate." Since the most costly aspect of the census is repeated efforts to contact these nonrespondents, sampling could save hundreds of millions of dollars and increase overall accuracy. The Census Bureau proposed to count directly at least 90% of the population in each census tract and to sample one tenth of the remainder. Despite some drawbacks associated with sampling, most experts believe it would have be the best course to follow.[50]

The issue of sampling has distilled partisanship in its purest form. The Democrats encourage sampling because it increases the accuracy of the count. As a partisan matter, the traditionally undercounted groups in society (which will be picked up by

---

48. Rep. Rogers has presumably made proposals such as this for rhetorical purposes, to send a message to the Census Bureau that he was serious about cutbacks, not because he literally wanted to eliminate the long form. The long form is a questionnaire that has been sent to around one in six households as part of the census. It asks almost fifty questions about housing quality and cost, citizenship, migration, health, employment, income, and other household characteristics. Critics such as Rogers claim that it is not essential to collect this kind of information.

49. The Census Bureau estimates that it missed 1.8 percent of the population in 1990. Minorities, in particular, were undercounted. Whereas 1.3 percent of whites were missed, 5.7 percent of blacks were. The Government Accounting Office (1995, 11) estimates that closer to 3.9 percent of the population was left out. The Cenus Bureau has been faced with a declining response rate by citizens, and the cost of the census has grown in inflation adjusted dollars from $10 per person in 1970 to $20 in 1980 to $25 in 1990. See also Appendix D of this report for more discussion of census-related issues.

50. Sampling poses some risks in the form of biased samples, distortions in small-area estimates, and operational problems resulting from the complex mathematical procedures that must be followed. An error in a computer program at the Census Bureau would have mistakenly shifted a seat in the House from Pennsylvania to Arizona after the 1990 census if the adjusted count had been used (Wachter and Freedman 1996, 96). (Wachter and Freedman are professors of statistics and demography who argue that the statistical methods proposed by the Census Bureau to adjust the raw data are too complex.) Still, without sampling, the next census is likely to be either very expensive or very inaccurate. The American Statistical Association and three expert panels commissioned by the National Academy of Sciences have endorsed sampling because it will add to the accuracy of the census if full funding is not available.

sampling) tend to be aligned with the Democratic Party. Most Republicans oppose sampling and say they are willing to spend "whatever it takes" to do a full head count. They claim that sampling will lead to inaccuracies at the block or census tract level, and they ignore the difficulties of locating nonrespondents. An undercount of urban and rural areas will increase the relative strength of the suburbs, which traditionally favor Republicans. Thus, debates couched in terms of the technical merits and flaws of sampling are thoroughly partisan.

The issue has been at least partially resolved in the courts. On January 25, 1999, the Supreme Court ruled that a full enumeration must be used (with no sampling) to allocate seats in the House of Representatives among the states. However, the Census Bureau still plans to use sampling as the basis for distributing Federal aid to state and local governments and as the basis for legislative redistricting (Holmes 1999).

Racial and ethnic definitions The use of whether to use a category called "multi-racial" also created considerable controversy in the preparations for Census 2000. At present, the census asks household members to identify their race. The 1990 options included White, Black, Hispanic, Asian or Pacific Islander, and American Indian or Alaska Native. Those not fitting those categories could choose "Other Race." Groups calling for a multi-racial category have filed suit to require the government to use this category. If they succeed, some organizations based on existing racial and ethnic categories are prepared to fight in Congress to restore the status quo. They see the multiracial category as watering down hard-won political and economic gains (such as affirmative action and the Voting Rights Act). This is, however, primarily a symbolic issue. category. According to a survey conducted by an advisory group on the census, just over one percent of people are likely to choose the multi-racial category (Skerry 1996, 36-39). The politics of identity could descend into prolonged trench warfare between new groups that want the multi-racial category and old-line groups that do not. (For a more extended discussion of this topic see Appendix D.)

SUMMARY OF THREATS TO STATISTICS

The federal statistical system is being threatened in a variety of ways at present. As a result, both the quantity and quality of policy-relevant information is likely to decline.

Fiscal austerity is slowly bleeding the statistical system and reducing the effectiveness of federal agencies. The current budget surplus (in 1999) has offered a reprieve, but the growth of entitlements almost guarantees that statistical agencies will be faced with severe fiscal constraints in the future.

Devolution threatens the integrity of federal statistics by cutting off the source of a great deal of information that currently comes from reporting within public

assistance programs. Less and less will be known about the effectiveness of devolution because the data to verify progress or failure will be lost.

As a consequence of privatization, entire data sets have been turned over to profit-making entities, requiring the public to pay twice: once as taxpayers and a second time as users of the data. Although the law protects public access to the data that citizens have paid for, maintaining access at a reasonable price requires constant vigilance.

Privacy concerns threaten to undermine the gathering statistics by reducing the willingness of businesses and individual citizens to participate in surveys. The Census has already begun to suffer from declining response rates as have other surveys. While this withdrawal from surveys might only seem to affect the quality of statistics, it should also be taken as a sign of the general deterioration of civic life.

The politicizing of data collection represents the subtlest danger to the statistical system. If a policy is followed of preventing the collection of data that challenges ideological assumptions, the basis of democratic conversation will be destroyed.

III. THE SIGNFICANCE FOR LOCAL COMMUNITIES

The threat to federal statistics will affect every citizen in some manner, although most people will not be able to observe the difference directly. The chain of causation from quality of information to quality of life is far too complex to specify.

In an effort to assess the effects of potential reductions in the quantity and quality of statistical data on users, the editors of *American Demographics* asked various analysts to imagine how marketing, health care, and other aspects of life would be different in 2003 if the long form were eliminated from the census. Dowell Myers (1995, 38), a professor of planning at the University of Southern California, was asked for a scenario at the local level:

> Local agencies had to look elsewhere for the data they used to get from the census to tell them about households and citizens requiring services, such as transportation for the disabled or meals on wheels. Private vendors rushed to fill the void, but they competed on price, rather than quality. After all, without census data, how could anyone judge good data from bad? . . . For lack of a common standard, neighboring districts made opposing assumptions, reached inconsistent decisions, and threw a great cloud of confusion over localities. Mixing and matching data from different sources led to the discovery of many false trends, alternately alarming and exciting the public. State and federal agencies became mired in data disputes. . . . The court system was soon overburdened with cases hinging on disputed data, giving rise to a booming industry of data consultants and a new specialization known as legal demography.

This may be an exaggeration of the problems that would arise without small-area data, but its emphasis on the value of standardization of data across all jurisidictions is important. Locally generated data cannot replace federally collected data because they lack this crucial component: a consistent methodology that permits comparison from place to place.

Richard Thomas (1995, 40), with a medical research firm, proposes that the movement toward community-based health care would be sidetracked by the loss of small-area census data. Blocking that cost-saving change would cost consumers tens of billions of dollars.

Other analysts (Hodges 1995, 39; Wardell 1995, 39-40; Kintner 1995, 41) said that business would also suffer from the loss of small-area data. Yet, they note that private

credit-reporting companies and marketing firms already possess demographic and financial information at the household level, which is more precise (and more invasive of privacy) than the census data that would be lost. Although business would lose some capacity for niche marketing, it would not be as severely harmed as public sector data users. The major exception to that is the financial services industry, which currently relies on national data on housing values that would be extremely costly to collect if the census did not include it (Gelman 1995, 42).

This section of our report briefly examines some of the effects of declining data on local governments and citizen groups. There are two reasons for focusing on end-users in the public and nonprofit sectors rather than businesses. First, the former are the groups that most clearly represent the public interest. This is not to deny the vital importance of federal statistics to the health of the private sector, but some people might question the rationale for public financing of business-oriented statistics. Second, editorials in *American Demographics* have been alerting those with a marketing interest in statistics to threats to the system for years. There has not been a similar effort to address the needs of public sector users of statistics.

FEDERAL DATA FOR LOCAL PLANNING

One recently developed local planning tool relies heavily on data from a variety of sources, including federal statistics and satellite images. Geographic information systems (GIS) technology is computer software that allows users to map a variety of social and physical characteristics in order to identify spatial relationships that would not otherwise be apparent. For example, by tracing the spatial patterns of a disease outbreak in relation to social and physical features of a region, public health authorities might more readily determine the cause.

Local governments around the country have been using GIS for well over a decade. It has been applied extensively to zoning, public works engineering, property records management, and other site-specific activities. The data for those purely physical features can be supplied locally.

Recently, however, cities and counties are moving toward GIS programs that deal with the full range of services offered by local governments. As the uses of GIS expand, so also will data needs about the characteristics of local populations. That will mean that cities, counties, school districts, and other special districts will either have to conduct their own very expensive surveys or rely on the federal government. Although much of the data that is needed at the local level is collected as part of the decennial census, the ten-year interval between data points is too long for many purposes. In order to track annual demographic changes, the Census Bureau plans to develop the American Community Survey (see below on page XX).

Local Indicators

Local governments are not the only institutions that have been making use of small area data provided by the federal government. Nonprofit research and advocacy organizations in over 100 cities and counties throughout the United States have established community indicator projects or CIPs to evaluate progress within their communities (self-defined geographic regions).[51]

The development of community indicators creates a meaningful framework for approaching economic, environmental, and social problems in a way that (1) informs policy discussions (2) educates the private and public sectors about positive and negatives trends that characterize their local economy, and (3) enables goverments to start measuring the effects of what they do rather than what they spend money on doing.

As another example, organizations in Atlanta, Boston, Chicago, Cleveland, Denver, Oakland, and Providence that have been involved with the Urban Institute's National Neighborhood Indicators Project (NNIP) are relying in part on federal data in constructing their databases on neighborhoods within each city.[52] "The indicators projects in the seven pilot cities aim to provide reliable and consistent information on the social and economic conditions . . ., with a focus on informing discussions and policy deliberations relating to persistent poverty and neighborhood decline" (Sawicki and Flynn 1996, 173).

This is precisely the kind of local initiative that the proponents of devolution envision. Yet, without the basic data provided by the federal government, these kinds of projects cannot be carried out. It would be cost-effective for the federal government to gather more small area data, so more groups might take on projects of this type.

The flow of information need not be in a single direction. CIPs rely on federally collected data to measure, for example, the percent of households paying a high percentage of income for housing and utilities in the Truckee Meadows Region (NV) or the number of affordable housing units in King County (WA). Still many other CIPs, such as the cities of Fremont and San Jose (CA), rely on local level data to measure things like the number of business licenses outstanding for operations with a single employee.

---

51. Redefining Progress has been identifying and supporting CIPs in an attempt to better link them together, and to facilitate the development of these efforts. Community indicators are being initiated, designed and researched by regional planning agencies, Chambers of Commerce, city governments, nonprofit environmental organizations, to name of few.

52. Patrice Flynn, who formerly worked on that project at the Urban Institute, points to this as an example of how nongovernmental organizations have played an important role in transforming federal data into a form that is useful to local citizen groups. The mere existence of federal data does not matter until some individual or group finds a way to make it useful in the lives of ordinary people.

City, county, and school district administrative records are valuable sources of local data. Other data pertaining to accessibility of public services might be included in an indicators project. In Atlanta, for example, research confirmed complaints of a shortage of supermarkets in low-income areas of the city (Sawicki and Craig 1996, 518). That sort of information (address-coded for GIS) could be fed back into a comprehensive database at the federal level. On that basis, it would be possible to assess conditions across cities at a level of detail that is not currently possible.

One kind of national database that could only be constructed from local data is one that used records of property ownership to estimate the concentration of real estate holdings on a national basis. Very little is known about the distribution of wealth in the United States, so measures of income distribution are consistently used by default. Estimates of property ownership within individual jurisdictions have shown concentrations far higher than for income. Since the largest property owners have assets in a number of jurisdictions, a multi-jurisdiction database would make it possible for the first time to gain an accurate picture of the distribution of tangible wealth in the United States. This information could also be used to correct the national income accounts, which currently must rely on income tax records to estimate annual rents from property. Eventually, a continuing property ownership database might be used to answer a perennial question about American society: to what extent is inherited wealth, as opposed to accumulated wealth, the basis of differences in economic opportunity?

Preserving Data in Accessible Form

Assessment of local conditions by community groups will become increasingly difficult if data are available only in electronic form. The use of the Internet to provide data to the public is cost-effective for many sophisticated users today, and eventually will be accessible to most households. But for the next decade or so, there are millions of people who do not have access to computers or who lack the skills to use them, particularly more sophisticated applications involving large databases. This will be especially true in low-income areas of cities and rural areas where educational levels are also low.

Thus, it will be important for the federal government to continue (or resume) publishing in paper form data that are of use to local citizen groups. If present trends continue, many local research and advocacy groups concerned about conditions in inner cities could be effectively denied the information they need to participate in decision-making.

BETTER FEDERAL STATISTICS FOR LOCAL AREAS

Just at the point when cities, counties, and citizen groups are preparing to expand the work they are doing in policy-relevant GIS and indicators projects, the federal government is in the process of cutting back on available data. If this important work at the local is to proceed, however, simply maintaining existing relationships with federal statistical agencies will not be sufficient. A closer partnership between federal agencies and local area users will need to develop if the full potential for this kind of work is to be realized.

With adequate data, devolution could become a meaningful transfer of responsibilities rather than the thoughtless dumping of federal programs on states and local governments, as now seems planned. The GIS technology that could make possible new ways of analyzing and responding to local issues is available, but it is currently being used almost exclusively for physical planning. As William Drummond, an expert in GIS, explains:

Over the last 20 years, geographic information systems (GIS) transformed the way in which urban planners conduct environmental analyses, register parcel boundaries, map infrastructure location, and model transportation systems. Surprisingly, though, GIS technology has had much less effect on the analysis of the human activity patterns that underlie the countryÕs most pressing urban problems, including poverty, crime, education, teenage pregnancy, public health, and unemployment. One of the major reasons for this neglect is the difficulty of generating accurate, timely, and inexpensive locational information for human activities (Drummond 1995, 240).

Much of this work can be done already by applying local databases to reference databases that are available from the federal government. In order to achieve comparablity across different jurisdictions, which will aid local governments in learning from experience elsewhere, federal agencies will need to take a more active role in coordinating what is now a completely decentralized operation.

Coordinating Agencies to Provide Local Area Data

At present, federal agencies primarily have experience collecting national-level data through surveys. While BLS, NCHS and Census have developed cooperative programs with states and local governments, overall agency capabilities for local-area data are weak. For example, the Census Bureau began producing county and city population estimates in 1970, but it does not provide small-area measures of poverty or data by age, sex, and race. BLS collects local area data for employment and unemployment measures, but these are not linked to the variables that Census uses (age, race, income, education level). The IRS Statistics of Income Division cannot easily convert its tax return data into broader measures of the population and households (Spar 1986, 10).

At present no plan exists that would aid local agencies and citizen groups in bringing together local area data in an inexpensive way. Local entities have to start

almost from scratch in each case. What is needed at this point is a federal initiative to assist local agencies in linking city and county administrative records--public assistance program data, drivers licenses, etc.--to existing national surveys, and to extend the scope of national surveys to make them more compatible with the needs of those working at the local level. Making this link will be costly in the short run, but an investment today in common national definitions and standards could pay off.

Developing Outcome Measures, Not Just Traditional Statistics.

Federal statistical agencies could be of greater assistance to local citizen groups if the agencies gathered data useful for performance measures. This would enable citizen groups to compare cities and counties around the country on the basis of relevant characteristics and hold local officials accountable for failure to make improvements.

In developing performance measures, however, statistical agencies enter the realm of politics. State officials complain that federal statistics often lack policy relevance, yet governors and mayors often have an active aversion to data that can be used for national ranking and public comparison of state and local performance. For their part, federal statistical agencies fear losing their credibility and objectivity if the traditional boundary between statistics and policymaking is blurred. Yet, it will be hard for agencies to remain relevant in a devolved system without more attention to performance and outcomes.

Need For A Coordinated Federal Statistical System

In order to provide the kind of information needed by local users, federal agencies need to develop a more unified approach that will permit the combination of multiple databases. Currently, the entire onus for coordinating the work of federal agencies lies on local users who must deal with a wide array of technical requirements to access the variety of databases at the federal level. This is a daunting task even for experts in the field of information management. The problem is that federal statistics are oriented toward specific programs, not for end-users who cut across programmatic boundaries.

Strong leadership will be needed at the federal level to overcome this impasse. Yet, in the absence of pressure from citizen groups, states, and local governments, the synthesis at the national level that would make federal agencies user-friendly is not likely to occur. The Office of Statistical Policy in the Office of Management and Budget is theoretically responsible for encouraging cooperation among agencies and with public users, but with only a five-person staff, its effectiveness is limited. Moreover, statutory confidentiality requirements governing individual agencies will have to change to allow greater data sharing among statistical agencies. [see Statistical Confidentiality Act, p. ?].

Need For Federal Leadership

Because of economies of scale and coordination in the collection and analysis of statistics, much of the expertise on data issues within the public sector is currently at the federal level. There are, of course, specialists within state data centers around the country who are higly qualified. Nevertheless, it is naive to imagine that data collection can be transferred readily to the state level as funding is cut at the federal level. Programs at the federal level have taken decades to develop. Moreover, federal budgetary pressures are mirrored at the state level by a growing number of governors committed to reducing the size of state government and staffs, even in the absence of budget deficits. In the case of data, devolution is likely to mean demise.

The value of statistical work is simply too obscure to develop a political constituency in every state and local jurisdiction. Unless the federal government maintains its role as a leader in this field, the deterioration of quality statistical agencies at the national level will be magnified locally. If Congress wants to take devolution seriously, federal agencies need to be funded to develop training programs for policymakers and analysts at state and local levels.

IV. RECOMMENDATIONS: RESPONDING TO THE THREAT

FULL FUNDING FOR STATISTICAL NEEDS

When national security is at stake, few would argue that skimping on centralized information makes sense. When potential threats to the command and control system of America's nuclear arsenal became apparent in the 1980s, the problem was taken seriously. There was no point in having massive firepower if the institutions that controlled it could be easily destroyed. Weak links in the chain were discovered and repaired, to the extent possible.

Preserving information about social, environmental, and economic conditions within the U.S. should be seen as equally vital to the national interest as the information system used to control military force. Since domestic chaos and conflict will weaken the nation internally, the case for civil statistics that enable leaders to track social changes may be even stronger than the case of military intelligence.

Thus, it is ironic that fiscal conservatives should be leading the assault on funding for the federal statistical system. In the name of saving a tiny fraction of 1% of federal spending, they seem intent on allowing the social command and control system to deteriorate.

The budget that is needed is more than a maintenance diet for statistical agencies. Although it runs counter to the prevailing ideology of uniform, across-the-board cuts, there should be *added* spending on statistics. The increased funding should be tied to a general reform of the statistical system by decreasing the number of separate surveys but making the remaining ones more frequent and more comprehensive in scope. That will require coordination among statistical agencies or the creation of a single statistical super-agency that would house most existing agencies.

THE AMERICAN COMMUNITY SURVEY: BETTER QUALITY FOR LESS MONEY

In a mobile and rapidly changing society, a decennial census is an inadequate basis for planning and evaluation. Almost half of all households change residence within a five-year period, and about 20% of the population moves to a different county (based on 1990 census data). From now until 2002, when the data from a new

census are made public, information about the population is going to be nine to twelve years out of date. For counties with significant growth or decline in the intervening years, the census provides a relatively poor picture of actual conditions.

The existing decennial census is like a snapshot of the characteristics of Americans. To observe a society in constant motion, what is now needed is a movie.

The Census Bureau has begun working on the "script" for such a movie. The proposed American Community Survey (ACS) would take a snapshot of 250,000 households each month, or three million per year, with an initial sample size, from 1999 to 2001, of 400,000 households per month. (Alexander 1996, 1). It would ask all of the questions now included on the "long form" of the decennial census: questions dealing with income, poverty, education, commuting, housing characteristics, and other details about households. By the year 2010, if all goes according to plan, the long form would no longer be needed in the decennial census.

For Congressional districts and states, plus cities, counties or other areas with a population greater than 65,000, one year of data from ACS should be sufficient to make statistically reliable judgments (http://www.census.gov/CMS/www/acs.htm). For less populated areas (including census tracts within large cities), the data would need to be averaged over several years to give statistically reliable results. Rural areas (with less than 2,500 people per jurisdiction) would be sampled at about three times the standard rate in order to increase the validity of small area statistics.[53]

One of the main goals of the ACS, according to the Census Bureau, is to "aid state and local officials in meeting their new responsibilities under devolution" <http://www.census.gov/CMS/ www/acs.htm#back>. As we saw in an earlier section of this report, devolution will depend heavily on the availability of information, much of which is not currently available at the state or substate level.

There are at present only a few surveys that provide continuous data about social or economic conditions at the state or local level. The Current Population Survey of 60,000 households, conducted by the Census Bureau each month for the Bureau of Labor Statistics is the source of raw data about unemployment. The March supplement (of 60,000 households) provides annual data on poverty and income. Only the monthly survey is large enough to provide statistically reliable data below the national level. By providing calibration data from three million households per year, ACS would enable federal surveys dealing with health, crime, poverty, and many other issues to be disaggregated below the national level.

---

53. If standard samples were used in a rural area with 200 households, only 25 to 30 (12.5% to 15%) would be sampled in a five year period. That is far below the 50% rate of long-form sampling in rural areas for the 1990 census.

Advantages of the ACS

The ACS would permit states or local governments, for a fee, to include supplemental questions or to expand the sample size in particular areas in any given month. This would be cheaper than conducting a stand-alone survey. The flexibility to piggy-back on a continuing survey is one of the most important features of ACS.

Providing data about annual changes would be a major boon to city planners in cities and urban counties. Even in rural areas, a five-year rolling average from the survey would provide better information about changing trends than the current decennial census does.

For local governments and citizen groups seeking better indicators of economic and social progress, the ACS would provide a new method of tracking change. As the Center for the Study of Social Policy (1995, 10, 11-12.) notes, the ACS:

> would also provide the data to facilitate outcomes-based accountability [by states and local governments]. . . . [A]ll of these jurisdictions are limited by the current lack of detailed, up-to-date data at the state and local levels. For example, Minnesota wants to measure its percentage of children living below the poverty line, but this data is available only every ten years from the decennial census. Oregon wants to measure unemployment by race; however, statistically significant small-area unemployment rates for racial and ethnic groups are only available through the decennial census.

Problems with the ACS

Despite the numerous advantages of the ACS, there are some drawbacks that need to be taken seriously.

First, transportation analysts, rural planners, and others who model activity of small population units will have to become accustomed to dealing with averages rather than single year estimates (U.S. DOT 1996, 24, 54, 95-96). This is likely to cause some confusion, since city councils and county commissions are used to dealing with point estimates. Some software that is currently used in planning will have to be redesigned. In cases where accurate baseline estimates are needed, local governments may have to pay for an increased sample size for one month.

Second, there is some concern that the promise of ACS would not be fully realized due to funding constraints. The ACS would cost about $75 to $100 million per year (Edmondson 1997). As a new program, this would require additional funding. In principle, it could lead to some savings in 2010 by eliminating the need to use the "long form" that year. The net cost over the decade 2000 to 2010 could actually be lower than continuing on a business-as-usual course. But if Congress decides to trim some of the funding for the ACS, which would lead to smaller sample sizes, that could

wreak havoc with the program. It would reduce the statistical significance of the small-area samples and reduce public confidence in the entire endeavor.

Third, because of the relatively large on-going budget for the ACS, some other agencies that conduct their own surveys are worried that the ACS will eventually take over their work (presumably by adding questions to the existing census questions). This would require a degree of coordination between agencies that neither Congress nor OMB has yet been able to attain. Thus, any fears (or hopes) that ACS would become a kind of unified federal survey are not likely to be realized.

COORDINATED STATISTICAL ACTIVITY AT THE FEDERAL LEVEL

The federal statistical system in the United States is unique among industrialized countries in its degree of dispersion among program agencies. Most statistical systems are centralized in one agency which collects, analyzes, and disseminates data. Statistics Canada is such an example. Although there are countries with dispersed authority over statistics, they usually have a strong coordinating agency to oversee activities of different agencies and to promote overall system coherence and efficiency. Individual U.S. agencies have far greater autonomy than in other countries, and OMB's Statistical Policy Branch staff of five lacks the personnel and authority of its foreign counterparts.

Consolidating collection and analysis of statistics in a single agency would give that type of work greater visibility and political leverage, but it would reduce the flexibility that the current system offers.

What is needed is greater coordination among existing statistical agencies. That will require congressional action to remove some of the legal barriers that currently prevent agencies from sharing information. It will also require granting a larger budget and greater authority to the Chief Statistician of the U.S., who is housed in the Office of Management and Budget. That office should have the power to oversee the work of all statistical agencies 1) to reduce survey burden on individuals and businesses, and 2) to create more uniform methods of data management to allow public users to cross-reference agency data. Janet Norwood (1995, 69-87), former Commissioner of BLS, has outlined a proposal such as this to promote cooperation and information exchange between agencies within the federal statistical system.

ENSURING CONFIDENTIALITY OF PERSONAL AND PROPRIETARY INFORMATION

One of the barriers to interagency data-sharing and cooperation is legislation governing privacy and confidentiality of information (NRC & SSRC 1993). As each statistical agency was set up, the rules dealing with confidentiality were different. These conflicting requirements have made interagency sharing of data difficult or

impossible. The Statistical Confidentiality Act was introduced in the 104th Congress and has been reintroduced subsequently as an effort to clarify and simplify this situation.

The proposed Act represents an effort to preserve existing rules that prohibit disclosure of data identifying specific individuals or businesses, while accommodating the growing need for federal agencies to share data for strictly statistical purposes. The proposed Act would enable agencies that now collect the same or similar data through separate surveys to share information, thereby decreasing both federal costs and reporting burdens for citizens and private institutions. It retains safeguards guaranteeing public privacy and confidentiality.

The Act would improve coordination and assure privacy in the following ways:

- Reduce the public reporting burden and duplication expenses by allowing interagency exchange of information;
- Lower the cost and improve the accuracy of statistical programs by facilitating cooperative projects between agencies;
- Ensure that information given directly or indirectly to a designated statistical agency for statistical purposes will not be disclosed in individually identifiable form without individual consent; and
- Reduce the risk of inappropriate disclosure by designating specific agencies as authorized to receive information from other agencies.

The Act would enable designated Statistical Data Centers (BEA, Census, BLS, NASS, NCES, NCHS, and designated units in DOE and NSF) to manage data resources as if they were a single entity. The Centers would have the responsibility of identifying opportunities for eliminating duplication and reducing costs by sharing information and entering into joint projects. Policies established by the Act will be overseen by OMB.

The Statistical Confidentiality Act is a necessary step towards greater cooperation among statistical agencies and the integration and streamlining of their activities. At the same time, the Act attends to the growing privacy concerns of Americans, while still allowing for a minimum of system-wide reform. It has the potential to increase agency productivity and reduce reporting burdens, without jeopardizing the quality of statistics.

INTEGRATED ENVIRONMENTAL AND SOCIAL INDICATORS

Indicators are an important element in advancing public understanding and appreciation of the connections between the environment and human well-being. At

present, environmental and social interests often seem outside the realm of traditional economic concerns. That is partially due to the fact that statistical agencies collect data for narrow purposes. No agency has yet brought together all of the data collected by the government that bears on the health of the nation as a whole.

As a first step in illustrating relationships between various factors in the nation's health, the Sustainable Development Indicators Group (SDIG), an interagency working group established in 1996 by President Clinton, has begun a process of developing an integrated set of indicators.[54] SDIG staff and agency advisers released a candidate set of national indicators in early 1997 for review and comment. The purpose of the indicators is to highlight critical trends and show how economic, social and environmental concerns interact.

Over time these integrated indicators have the potential to become a much more comprehensive and meaningful approach to assessing national well-being than our current fixation on narrow economic indicators such as the GDP.

COMMISSION TO STUDY THE FEDERAL STATISTICAL SYSTEM

Citing nearly a century of Congressional neglect of federal statistics, Senators Daniel Patrick Moynihan (D-NY) and Robert Kerrey (D-NE) introduced a bill in 1996 to establish a commission to study the federal statistical system and to make recommendations to Congress for reform. Moynihan and Kerrey drafted the legislation out of concern over poor coordination within the system, inadequate adaptation to a changing U.S. economy, and failure to adopt new technologies for collecting and disseminating data (Moynihan 1996, S11296).

The proposed commission would have the authority to review, evaluate and make recommendations on a range of matters, including (U.S. Congress 1996, 6-8):

I. the role of multipurpose agencies that collect and analyze data of broad interest, such as BEA, BLS and Census;
II. the mission and organization of various statistical agencies with an emphasis on expansion or elimination of activities, and prioritization of those activities;
III. the advantages and disadvantages of a centralized statistical agency or partial consolidation;
IV. statistical methodologies and the accuracy and appropriateness of key statistical indicators;

---

54. The SDIG was initiated in response to a recommendation by the President's Council on Sustainable Development, a group composed of government officials, corporate executives, and environmental leaders. The SDIG is an interagency effort to produce economic, social and environmental indicators of national sustainable development. SDIG convenes the Departments of Agriculture, Commerce, Education, Energy, Interior, Justice, Health and Human Services, Housing and Urban Development, State, and Transportation, as well as EPA and NASA.

V. interagency coordination and methods for standardizing collection procedures, surveys and presentation of data throughout the system;

VI. foreign statistical systems in comparison to the U.S. and statistical coordination with other nations and international institutions.

VII. Ultimately, the commission would be charged with recommending a strategy for maintaining a modern and efficient federal statistical infrastructure capable of accomodating changing national needs.[55]

While the idea of a commission is appealing if Congress continues to devalue the importance of stastistical work, the Kerry-Moynihan proposal should only be viewed as a second-best solution. Sen. Moyhihan himself acknowledges that 16 different committees, commissions, and study groups have addressed the statistical system from 1903-1990 and little or no action followed their recommendations. Another commission is no guarantee of meaningful action. Indeed, some leading statistical officials oppose it on that basis. They argue that a commission would accomplish nothing and postpone, and possibly derail, the Statistical Confidentiality Act, something which the statistical community has labored for years to bring to the point of passage.

There is a further risk--which the current legislation does little to guard against--that the proposed commission will be dominated by a small circle of insider statisticians and economists. Some old hands are surely needed to provide the wisdom of past experience, but the statistical system desperately needs the creativity and fresh ideas that new faces would bring. As the CPI debate has shown, the economics profession largely sets the terms of statistical debate, and economic indicators unsurprisingly receive undue attention over other statistics. Voices for social and environmental data deserve their due, as do those with experience working with state and local statistics. The devolution of welfare and other programs warrants substantial representation from data users at the state and local level. A commission composed of members with a broad range of experience will be necessary to craft an effective blueprint for the fundamental reform the system needs.

Awareness in professional statistical circles of the need for comprehensive and fundamental reform of the federal statistical system is growing, and an adequate road

---

55. With the recent controversy over the CPI, and growing worries in the academic and business communities about the quality of economic statistics, the bill stands a reasonable chance of passage. It has also been blessed by nine former chairmen of the Council of Economic Advisers, who signed a letter in support of the proposal (Letter to Moynihan and Kerrey 1996). In the letter, the economists point to many problems already discussed in this report:

All of us . . .remember the problems . . . under the current system of widely scattered responsibilities. [S]tatistical priorities are set in a system within which individual Cabinet Secretaries recommend budgetary tradeoffs between their own substantive programs and statistical operations which their departments, sometimes by historical accident, are responsible for collecting. Moreover, long range planning of improvements in the federal statistical system . . . [are] hard to organize in the present framework. The Office of Management and Budget and the Council of Economic Advisers put a lot of effort into trying to coordinate the system, often with success, but often swimming upstream. . .

map is evolving for that reform. In the past, general public and official disinterest in statistics, combined with the decentralization and mind-numbing complexity of the current system, have bedeviled previous reform efforts. To some extent, obscurity can be a political blessing; the American political system has not been friendly to serious public policy and government reform in recent decades. Small budgets and the low profile of federal statistics mean that much can be accomplished with a pittance of money (by the standards of federal spending) and with strategic leadership from key elected officials. However, getting political leaders to step to the plate will require new constituencies for federal statistics at the community level to inform and encourage their elected officials to take action.

> All of us . . .remember the problems . . . under the current system of widely scattered responsibilities. [S]tatistical priorities are set in a system within which individual Cabinet Secretaries recommend budgetary tradeoffs between their own substantive programs and statistical operations which their departments, sometimes by historical accident, are responsible for collecting. Moreover, long range planning of improvements in the federal statistical system . . . [are] hard to organize in the present framework. The Office of Management and Budget and the Council of Economic Advisers put a lot of effort into trying to coordinate the system, often with success, but often swimming upstream. . .

APPENDIX A: OVERVIEW OF THE FEDERAL STATISTICAL SYSTEM

The federal statistical system often seems a complex maze of institutions and programs, even to those who work within the system. Thus, a brief overview will be helpful for those readers who are unfamiliar with it. (See Norwood (1995) for a more complete discussion of the statistical agencies.)

The official federal statistical system consists of the eleven agencies listed in the table below:

TABLE 2: STATISTICAL AGENCIES OF THE U.S. GOVERNMENT

Statistical Agency	Governmental Department
Bureau of Labor Statistics (BLS)	Dept. of Labor
Bureau of the Census (Census)	Dept. of Commerce
National Center for Educational Statistics (NCES)	Dept. of Education
Energy Information Administration (EIA	Dept. of Energy
National Center for Health Statistics and Human Services (NCHS)	Dept. of Health
National Agricultural Statistical Service (NASS)	Dept. of Agriculture
Economic Research Service (ERS)	Dept. of Agriculture
Bureau of Economic Analysis (BDA)	Dept. of Commerce
Statistics of Income Division (SOI), Internal Revenue Service	Dept. of Treasury
Bureau of Justice Statistics (BJS)	Dept. of Justice
Bureau of Transportation Statistics (BTS)	Dept. of Transportation

The dispersal of these agencies across nine different executive departments adds to the complexity. Each agency is also organized and managed differently, depending upon its position and status within its department. Finally, over seventy other federal agencies perform statistical activities, contributing yet another layer of complication to the system.

The first statistical bureau, SOI, was established in 1860, in the Treasury Department. ERS, NASS, and NCES emerged soon after in 1867, followed by BLS in

1884. The twentieth century gave rise to several new agencies, largely due to government expansion and changes in the economy and society, both of which increased the need for new information and statistics. Census was created as a permanent agency in 1902, NCHS in 1912, EIA in 1977, BJS in 1979, and BTS in 1991.

The Office of Management and Budget (OMB) coordinates and oversees the different activities in all of the different agencies. OMBOs Office of Information and Regulatory Affairs (OIRA) determines specific classifications and definitions, controls the quality and accuracy of data, ensures objectivity and confidentiality, and coordinates with statistical agencies from other countries. A staff of five under the Chief Statistician of the U.S. make up OIRA's Statistical Policy Branch. This small group performs the daunting task of overseeing and coordinating the whole federal statistical system.

ORGANIZATION OF STATISTICAL AGENCIES

There are large disparities in status among the various agencies. The heads of BLS, EIA, and BTS report to cabinet secretaries, Census and BEA to undersecretaries, NCES and ERS to assistant secretaries, and BJS, NCHS, and SOI through several layers of officials (Norwood 1995, 27).

That some agencies enjoy considerable independence in decisionmaking and budgets, while other agencies report to second or third-tier department officials, makes coordination and collaboration among agencies difficult. BLS and NCES provide an excellent example of differences in agency management and organization. Appropriations for BLS are completely independent of its parent, the Department of Labor. The BLS commissioner commands respect throughout the Administration and Congress. Although the commissioner is appointed by the President and approved by the Senate, s/he has a four-year term, and reappointment to more than one term has become a tradition. The BLS Commissioner answers directly to the Secretary of the Department of Labor, and has the same rank as the Assistant Secretary, assuring her a highly independent and respected role in the Department.

The NCES commissioner, on the other hand, does not command similar independence or clout. The budget of NCES is not independent of the Department of Education, nor does the Commissioner report to the Secretary, but rather the Assistant Secretary of the Office of Educational Research and Improvement. Obviously, despite the same title, the NCES Commissioner does not enjoy the same stature as that of the BLS Commissioner.

AGENCY ROLES AND ACTIVITIES

Congress mandates that each statistical agency perform specific activities and appropriates funding to do so. An overview of the individual agencies follows.

Bureau of the Census

Census has a broad mandate to collect and publish general-purpose statistics based on surveys. Most people know of the decennial census of population and housing sent to every American household, and many are aware of the quinquennial census of agriculture and business. Census activities are not limited, however, to large, infrequent surveys. The agency also conducts other surveys on a more frequent basis, often under contract to other agencies such as BLS and NCHS. Research and data from the Census Bureau provide an important foundation of information for use by other agencies and policy makers. Only BLS tops Census in terms of its budget allocation, which was $146.1 million for FY 1999 (not including the separate budget for the decennial census).

Bureau of Labor Statistics

Created in 1884, BLS was given a mandate to collect and disseminate information on subjects relating to employment and the workplace. BLS collects, compiles, analyzes and disseminates data relating to employment and unemployment statistics, industrial relations, wage and cost of living, technology, productivity, current and projected economic growth, and worker safety. Given its range of activities, the BLS budget ($398.9 million for FY 1999) constitutes one third of the overall budget for statistical agencies. BLS is not only the statistical arm of the Department of Labor, but it also interacts and coordinates independently with Congress, executive branch agencies, and other statistical agencies. The agency has maintained a reputation for independence and objectivity that has allowed for stability and longevity in personnel and planning and made it one of the most highly respected agencies in the system.

National Center for Educational Statistics

NCES collects, develops, and disseminate information on K-12, secondary and post-secondary education in the areas of achievement, institutions, and finance, as well as student access to education, the educational environment, and education professions. NCES also works with states to develop uniform standards, definitions, and collaborative data collection activities. The Center has a small staff, and it contracts out much of its work (data collection in particular) to the private sector and research groups. Its FY 1999 budget was $104 million.

The Department of Agriculture has two statistical agencies, NASS and ERS. Although they collaborate on data collection, they are independent agencies, with different heads and budgets.

National Agricultural Statistical Service

NASS has a broad mandate to collect and disseminate general information about agriculture. The agency collects information on crop productivity, crop and livestock prices, agricultural employment and income, and general farming conditions. It often performs activities for or works closely with other agencies on issues relating to agriculture. For example, NASS works with Census in the preparation and implementation of the agricultural census. The budget for NASS in FY 1999 is $80.4 million.

Economic Research Service

Although it generates some data and information, in addition to producing agricultural indicators, the main activities of ERS include analyzing NASS data and forecasting economic trends related to agriculture and agricultural commodities. The budget for the ERS is less than that of the NASS ($65.8 million for FY 1999).

Energy Information Administration

EIA collects and disseminates data on energy reserves, production, demand, and new technology. As with BLS, EIA data is often politically sensitive. Therefore, with its creation, Congress put in place legislation which would ensure independence of the agency and of its administrator to plan its activities. The administrator does not have a fixed term of office, providing for greater stability and long-term planning within the agency. The EIA budget for FY 1999 was $65.8 million.

National Center for Health Statistics

NCHS collects and disseminates statistics on broad issues relating to health, such as illness, disability, and births and deaths. The agency itself is located in the Centers for Disease Control and Prevention (CDC) in the Department of Health and Human Service (HHS), and the director of the agency reports to the secretary of HHS through several administrative layers. The NCHS works closely with other agencies, in particular Census and NCES, as well as contracting out to the private sector. The NCHS budget for FY 1999 is $94.6 million.

Bureau of Economic Analysis

Despite having one of the smaller budgets of the formal statistical agencies ($43.1 million for FY 1999), the BEA develops, analyzes, interprets and publishes some of the most widely used and politically influential economic statistics. Its main activity is to develop the national income and product accounts (the GDP among them) and trade accounts. Simply put, the BEA is the nation's accountant, charged with developing measures for every economic activity that takes place in the country. BEA does not conduct many of its own surveys. Rather, most of the data come from other agencies, primarily Census and BLS.

Bureau of Transportation Statistics

BTS is the newest statistical agency, established in 1991 by the Intermodal Surface Transportation Efficiency Act. This Act gave the Agency the broad mission of compiling, analyzing and publishing statistics on highway, rail, and intermodal transportation systems. Its architects took into account problems of older statistical agencies, and it is often referred to as a model when considering agency reform. BTS remains a small agency in its formative stages, with a budget of $31 million for FY 1999.

Statistics of Income Division

The oldest statistical agency, SOI recently moved to the Internal Revenue Service (IRS). The agency provides information on income, financial and tax issues to the Treasury Department and Congress, most of which is derived from tax returns. The budget for the SOI ($28.8 million in FY 1999) became part of the IRS budget in 1994.

Bureau of Justice Statistics

BJS was established in 1979 to collect and analyze statistical information on crime, juvenile delinquency, and the criminal justice system. BJS works closely with state and local statistical centers to support the collection of different criminal justice data. Most data come from BJS-sponsored surveys, such as the National Crime and Victimization Survey, conducted by Census. BJS falls under the Office of Justice Programs within the Department of Justice, and some of its funding comes from the Department. Its budget for FY 1999 was $25 million.

Little 60

Called the Little 60 (although there are over seventy such agencies), there are many federal agencies not classified as part of the formal statistical system, but which perform statistical functions. These agencies devote a half million dollars or more to

statistical activities, and include such important and well-known agencies as the Environmental Protection Agency (EPA), National Institutes of Health (NIH), and Immigration and Naturalization Service (INS). These agencies' combined statistical budgets amount to approximately $1.5 billion.

FEDERAL AND STATE COOPERATION

Many cooperative statistical programs exist between federal agencies and states, and they differ widely in nature and scope. Some concentrate on data collection, such as the one between BLS and State Employment Security Agencies (SESAs). In this relationship, each SESA collects data, with technical help and funding from BLS. The Federal-State Cooperative Program for Population Estimates in the Census Bureau also collects data at the state level, but states fund the program. Other programs have close cooperation and integration between state and federal agencies, such as the Cooperative Agreement between NASS and State Statistical Offices. Others have very little integration, and cooperation is merely for gathering and exchanging information. Finally, some programs are voluntary, whereas others have a legislative requirement for state participation.

Appendix B: HISTORY OF STATISTICAL REFORM INITIATIVES

The tendency of Congress and the President to appoint blueribbon commissions, panels and working groups, and then ignore their recommendations, is legendary. As with many important policy issues, the history of federal statistics is littered with august bodies whose impact was incremental in the best of cases. This history is pertinent, given that Sens. Daniel Patrick Moynihan (D-NY) and Robert Kerrey (D-NE) introduced legislation (S. 205) in 1999 to establish yet another commission to study the federal statistical system. Congress has generally failed to follow-up on advisory recommendations with appropriate oversight and funding for special initiatives, often leaving leadership to the agencies themselves.

The following discussion of the history of federal statistics, derived primarily from Norwood (1995, 7-24, 53-59) is intended to provide an overview of reform efforts in the past, which might serve as an intellectual base from which to pursue future reforms.

THE BUREAU OF EFFICIENCY

An early attempt at overall reform of the statistical system occurred in 1919, soon after World War I. Congress mandated a system-wide review in order to obtain needed information for the war. The Bureau of Efficiency (BEF) conducted the analysis and published a final report in 1922. BEF made several recommendations, including the establishment of a centralized bureau in order 1) to reduce costs, 2) to make statistics more accessible to users, and 3) to reduce the burden on citizens who are asked to respond to surveys.

By 1922, the statistical agencies housed in the Departments of Commerce, Labor, Agriculture, and the Treasury, among others, had developed stakeholders who opposed consolidation. Thus, BEF's proposal to integrate and rationalize the statistical system was the first of a long line of failed efforts.

THE ESTABLISHMENT OF AN ADVISORY COMMITTEE

The New Deal in the 1930s saw a significant increase in the number of statistical agencies, as well as in the amount of data collected. This subsequently required more

formal coordination and organization of the overall system. The Committee on Government Statistics and Information Services (COGSIS) was established in June 1933 as an advisory committee to the American Statistical Association and the Social Science Research Council. It was charged with the task of evaluating many of the system's functions, including methodology and staffing. COGSIS conducted studies at agency request and operated much like a law firm managing a number of cases (Duncan and Shelton 1978, 27-28).

COGSIS recommended that the dispersed character of the federal statistical system be continued. At the same time, it urged that the powers of the Central Statistical Board be strengthened. That Board (renamed the Division of Statistical Standards in the Bureau of the Budget, which later became the Office of Management and Budget) was instrumental in providing leadership to and coordinating the activities of statistical agencies for several decades. In more recent years, the Statistical Policy Branch of OMB, the latest successor to the Board, has lost much of its clout. The branch was cut from 26 to 15 by Reagan, and now has only 5 professional staff (Norwood, 57).

THE MILLS-LONG REPORT

In 1948, the Hoover Commission commissioned the National Bureau of Economic Research to study the federal statistical system as part of a government-wide reorganization plan. The report, by F.C. Mills and Clarence Long (1949, 12), proposed the creation of a presidentially appointed position in the Executive Office of the President. (A similar recommendation was made in 1966 by the Kaysen Committee. Presumably, the aim was to give statistical issues more visibility and to enable the new office to impose greater uniformity of procedures.) The Mills-Long Report also suggested that the Bureau of the Census be the centralizing agency for certain repetitive activities. The Hoover Commission, after considering the report, did not propose any significant changes in the statistical system.

ECONOMIC STATISTICS

During the 1950s and 1960s, several committees were established to look at the development and methodologies of economic statistics programs (Duncan and Shelton 1978, 169-74). The National Accounts Review Committee, set up in November 1956, made many important changes in federal economic statistics. The Stigler Committee of 1959-60 looked specifically at price statistics, and the CPI in particular. The Gordon Committee of 1961 sought to improve the collection of employment and unemployment statistics through the use of new techniques and better definitions. The Bernstein Committee of 1963 reviewed statistics relating to balance of payments.

The 1970s saw the Advisory Committee on the Presentation of Balance of Payments Statistics in 1975-76. It advised OMB on how to make balance of payments statistics more useful for policy analysis. Several recommendations made by these committees were implemented and have improved the accuracy and efficiency of economic statistics.

THREE INITIATIVES IN THE 1970S

The 1970s saw three initiatives to study and evaluate the overall federal statistical system, and each produced a similar set of recommendations.

The President's Commission on Federal Statistics

In August 1970, President Nixon appointed a commission, headed by Alan Wallis, to study federal statistics, and directed it to examine the following issues: 1) statistical needs, 2) respondent burden and personal privacy, and 3) the effectiveness of statistical operations (Duncan and Shelton 1978, 172). The Commission suggested increased coordination of statistical activities and elimination of obsolete programs. It also recommended the establishment of an advisory board to review confidentiality issues (U.S. President's Commission on Federal Statistics 1971, Vol. One, 2). As with past efforts, most of its recommendations were not implemented.

The most important outcome of the Wallis Commission was the creation of the Committee on National Statistics (CNSTAT). Established under the auspices of the National Academy of Sciences, CNSTAT was originally envisioned as an advisor to OMB on the quality of work by various agencies, but it has not functioned that way. Instead, CNSTAT has offered evaluations of statistical agencies upon the request of the agencies themselves.

The Joint Ad Hoc Committee on Government Statistics

The Joint Ad Hoc Committee on Government Statistics, representing members of professional associations, met in 1976. Members were drawn from the American Sociological Association, the American Statistical Association, the Federal Statistics Users' Conference, the National Association of Business Economists, and the Population Association of America. Like previous commissions and committees, it focused on the problems of coordination and on the need to clarify the role of professional organizations and their review functions (Duncan and Shelton 1978, 173). It also requested further study of these issues by CNSTAT. A study was conducted by A. Ross Eckler and Thomas J. Mills, which again emphasized the need for increased coordination (Joint Ad Hoc Committee on Government Statistics 1976). It

also laid out several options for reorganizing the statistical system, ranging from a coordinating group to a completely centralized statistical system.

The President's Reorganization Project for the Federal Statistical System

In 1977, President Carter asked James T. Bonnen to oversee a study focusing on the policy relevance, quality, confidentiality, and integrity of federal data. The Bonnen Report was one of the most comprehensive studies ever done on the federal statistical system. Like the Mills-Long report and the Kaysen Committee, it recommended that a statistical office be housed in the Executive Office of the President. It also made recommendations regarding protection of personal privacy and uniform confidentiality provisions. The report was released at the end of the Carter Administration. Despite high-level support, no Congressional action was taken. Instead, the Paperwork Reduction Act was passed, moving the Statistical Policy Division (SPD) back to OMB (after three years in the Commerce Department).

DETERIORATION AND DEREGULATION: THE REAGAN YEARS

Ronald Reagan came into office intent on reducing the burdens of government on citizens. But rather than surgical removal of agencies or activities that were especially troublesome, the Reagan Administration made across-the-board reductions in agency budget proposals. As Janet Norwood, then Commissioner of BLS, explains: "A number of data programs were reduced or eliminated. The budget in each of the major statistical agencies was cut, with little organized review of the effects of those cuts on the data produced by the other statistical agencies" (Norwood 1995, 18).

Since OMB was pushing for this uniform reduction in the size of government agencies, it had to take its own medicine. The staff of the Statistical Policy Branch, which had provided some degree of coordination to the statistical system, was severely cut, and it has declined further in size since then.

The Paperwork Reduction Act, enacted in the closing days of the Carter Administration, placed the Statistical Policy Branch in the Office of Information and Regulatory Affairs at OMB. That assignment assured that statistical work would become merely an adjunct to deregulation. During the Reagan years, the Director of Statistical Policy shifted attention away from the coordination of statistical agencies toward the reduction of the burden on citizens caused by statistical surveys.

The focus of the Reagan Administration on deregulation thus directly undermined prior efforts to improve the statistical capacities of the government by minimizing the role of OMB in those efforts. In addition, it seems likely that deregulation led to a loss of available statistics by reducing federal oversight of certain economic activities. A major source of knowledge about the economy comes from the

reports generated by agencies that oversee power generation, transportation, finance, and other regulated activities. Whether deregulation has significantly reduced data availability is an unexplored question.

RENEWED INTEREST IN STATISTICS: THE BUSH ADMINISTRATION

The Bush Administration reversed course from the Reagan years and devoted more energy to statistical coordination and organization, especially in economic statistics. The Chairman of the Council of Economic Advisors under Bush, Michael Boskin, headed up a Working Group on Statistics in 1990 and 1991. It was also known as the Economic Statistics Initiative or ESI (GAO 1995a). Members included representatives from OMB, the President's Office of Policy Development, the Treasury Department, the U.S. Trade Representative, HUD, BLS, USDA, the Federal Reserve, the Commerce Department (both BEA and Census), and CEA.

The proposals from ESI about how to improve economic statistics included more funding for BEA, BLS, and Census, the statistical agencies charged with producing economic statistics. The ESI set forth specific goals: better economic accounts, particularly those dealing with the service sector and with price changes, improved surveys of households and businesses, plans for future personnel for statistical agencies, and sharing of data. The Bush Administration and Congress failed to support the funding increases that would have been necessary to achieve the goals set by the ESI.[56]

ESI revealed a great deal about the challenges of statistical reform and improvement. Congress failed to fund modest investments in economic statistics, even though these statistics have the most powerful political constituencies. The interest of the private sector in an accurate GDP has not ensured sufficient funding for work on the national income accounts. Building support for improvements in social and environmental statistics will be even harder.

Despite inadequate funding, some progress has been made toward the goals of the ESI by the agencies themselves. The Bureau of Economic Analysis (BEA) has revised the System of National Accounts to conform to United Nations standards. It has also sought to fill in gaps in national income data formerly provided by other agencies, even as it was forced to eliminate a number of long-standing data series in those accounts. The Census Bureau has conducted an annual survey of communications services, just one branch of the growing service sector. That sector is growing in

---

56. After release of the second set of recommendations in February 1991, the Bush Administration announced that it would request a $230 million increase in funding for economic statistics for fiscal years 1992 through 1996 to support implementation of the recommendations. Later, that budgetary commitment evaporated amidst fiscal constraints. For example, from 1990 to 1994, statistical agencies requested over $95 million to implement the ESI recommendations, yet only $50 million was appropriated by Congress.

economic importance, yet it is difficult to measure and requires special study.[57] The BLS completed the automation of payroll data collection, which is important in providing timely estimates of economic activity. In that case, BLS received its requested budget.

An important feature of managing a changing economy is the updating of the way in which commercial activities are classified. The Standard Industrial Classification (SIC) system that has been used for decades to categorize economic activity is badly out of date. The ESI recommended funding for BLS, BEA, and Census to update the system, but Congress provided a tiny fraction of what was needed ($100,000 to Census, nothing to BEA). Still, the agencies have devised a new system, the North American Industry Classification System (NAICS), in conjunction with Mexico and Canada, but lack of funding has slowed implementation in the U.S.

CONCLUSION

Despite occasional recommendations over the years that existing statistical agencies be combined into a centralized statistical bureau, that approach seems unlikely to gain political support. A dispersed statistical system is probably too deeply entrenched to change.

In the absence of consolidation, however, it is important that a single agency have the power to ensure that the 70 major and minor statistical agencies have the capacity to share data with each other and jointly manage projects of common interest. It is too much to ask citizens who want data from federal agencies to master 70 different sets of rules and categories for accessing statistics. This is becoming even more true as agencies place their data online. Some degree of standardization of protocols would be very helpful in making data compatible between agencies.

A number of commissions have proposed the basic idea of a coordinating agency throughout this century. An office in the predecessor to OMB was assigned that task in the 1930s. For awhile, the coordinating task was taken seriously. In 1947, there were 69 staff trying to make the statistical agencies into a unified system. By the time of the Carter Administration, the staff had been cut to 29. The Reagan Administration cut back further, to only 15 staff, and the Statistical Policy Branch now has only 5 professional staff, which cannot begin to do the integrative work that is required to bring coherence out of the diversity that remains within the federal statistical system.

---

57. Census requested approximately $7 million to expand service sector data, but received little more than $1 million total for fiscal years 1990 to 1994.

APPENDIX C: FIGHTING POLLUTION WITH DATA

The Toxics Release Inventory (TRI) stands as a model for how federal data can achieve policy ends without additional regulations or government programs. It has dramatically cut pollution just by making public certain data on emissions. (Up-to-date information about TRI is available at: <http://www.epa.gov/opptintr/tri/whatsnew.htm> and <http://www.rtk.org>).

In 1986, the Emergency Planning and Community Right to Know Act (EPCRA), signed into law by President Reagan, included a provision establishing the TRI, the first publicly accessible, on-line computer database ever created by federal statute. By requiring major facilities of some 25,000 companies and, more recently, the federal government to monitor and report their emissions of toxic chemicals, it gave citizens and community groups a new weapon in the war on pollution: information. It also gave the CEOs of corporations feedback they had not had before about their own companies.[58]

Monsanto was first with a 1987 pledge to reduce its worldwide air toxics emissions 90 percent by 1992 (MacLean and Orum 1992, 14 and passim). Since 1988, reporting industries have cut toxic emissions of listed chemicals by 43 percent and prevented the release of 2.1 billion pounds of toxic chemicals into the nation's environment (Browner 1995).[59] The mere public existence of the data has motivated plant managers to improve environmental performance and to seek regular input from members of the community on environmental, health and safety issues affecting them.

Similar state laws in Massachusetts, New Jersey and Oregon have produced equally remarkable outcomes. In Massachusetts, legislation to reduce use (not just

---

58. Prior to TRI, industry representatives had disputed Congressman Henry Waxman's (D-CA) estimate that chemical companies released 85 million pounds of toxics into the air annually. TRI revealed actual emissions to be far greater and created an enormous incentive to reduce emissions (MacLean and Orum 1992).

59. Despite its success, TRI improvements are needed. The current legislation only mandates the reporting of emissions, not the use of listed chemicals in manufacturing processes and products (where they can end up as hazardous waste or be emitted into the environment at the end of a product's life-cycle), workplace exposure to them, or their transportation through local communities. Moreover, only product manufacturers that produce, process or use toxic chemicals must report. This exempts some of the country's largest polluters, such as pesticide users, oil and gas companies, utilities, mining companies, incinerators, and hazardous waste recyclers. Finally, of the 72,000 man-made chemicals presently used in commerce, the Interagency Testing Committee, which advises EPA, has issued testing decisions on 40,000 chemicals (USEPA 1996). EPA has evaluated over 30,000 chemicals for their threats to health. Yet, under Section 313 of EPCRA, only 600 fall under TRI requirements.

emissions) of toxics has achieved the following results among 29 firms that competed for state awards in 1995 (with the results from the 16 firms competing in 1996 in parentheses) (Geiser 1995, 9; Massachusetts Executive Office of Environmental Affairs 1996):[60]

- a 5.7 (2.3) million pound reduction in toxic chemical use;
- a 1.5 (5.9) million pound reduction in hazardous waste generated;
- 431 (2) million gallons of water conserved;
- $2.3 ($4.0) million saved from reduced chemical purchases;
- $1.5 ($1.1) million saved from avoided regulatory compliance costs; and
- $2.1 million saved from water conservation (no data for 1996).

In New Jersey, corporate pollution prevention plans conducted at 16 facilities in response to state materials accounting requirements (to provide use data) saved $6.34 million, for a private-sector savings of $5-8 for every $1 invested by the Pollution Prevention Program (Page 1996, A1, A10; Johnson 1996; both cite Warren Muir and Thomas Natan, Jr., unnamed study by Hampshire Research Associates in Virginia). In Oregon, every facility required to report under the state's toxics use reduction legislation responded in a survey that they found the data so useful that they would continue to collect it whether required to or not.

60. Markoe (1997, 9) provides a more recent update on the program.

Appendix D: TWO RACIAL ISSUES IN THE CENSUS: MULTI-RACIALISM AND SAMPLING

Two factors brought racial issues to the fore in debates over the year 2000 census. The first issue was overtly about race: whether to create a separate category for those who identify themselves as multi-racial. The second issue is only indirectly related to race: whether to use sampling to ensure full enumeration of groups that are traditionally undercounted. Since racial minorities have tended to be undercounted far more than whites, it is reasonable to think of this as a racial issue as well.

MULTI-RACIALISM AND THE HISTORY OF CHANGING CATEGORIES

Perhaps no other issue stokes the fires of America's painful neurosis over identity more than the terminology by which we classify ourselves. In our transition from slave society to free, this country's racial and cultural categories and terminology have evolved considerably.

Following OMB Statistical Policy Directive No. 15, "Race and Ethnic Standards for Federal Statistics and Administrative Reporting," which was issued in 1978, survey respondents in the United States were asked to identify themselves racially as white, black, Asian or Pacific Islander, and American Indian or Alaska Native. (They could also identify themselves as Hispanic, but that was not a racial category.)

OMB carefully rejected any scientific and anthropological basis for these designations, reflecting the ambiguity and political compromises such categories necessarily represent. That classification scheme came under attack from Hispanics who wanted to be called Latino, blacks who prefer African American, and so forth. Also, Arab Americans sought their own category, as have native Hawaiians. Finally, the children of inter-racial marriages became increasingly vocal in their demands for a multiracial category, even proposing a federal lawsuit, if necessary.

OMB was under pressure from traditional minority interest groups not to change the racial and ethnic categories used in surveys. The statistical results affect the reported size of different minority populations, their representation in American politics, and determinations of compliance with federal and state affirmative action policies and the Voting Rights Act. Traditional civil rights groups opposed the

multiracial category because they feared an eventual loss of clout and resources for themselves and their respective constituencies.

In an attempt to resolve this quagmire, OMB created the Interagency Committee for the Review of Racial and Ethnic Standards. In 1995, this Committee sponsored a 60,000 household supplement to BLS monthly Current Population Survey in order to gauge racial and ethnic self-identification in the U.S. (Skerry 1996, 36). Findings from this survey were supplemented by the 1996 National Content Survey, which probed more deeply into how Americans define themselves (U.S. Bureau of the Census 1996). The results are complex, even tedious, but worth reporting given their political import:

- Adding a multiracial category to the race question first (and the Hispanic origin question second) did not significantly affect the overall percentages of people reporting in the White, Black, American Indian, or Asian or Pacific Islander categories. There were, however, write-ins to the multiracial category from the Asian and Pacific Islander categories, suggesting statistically significant changes in a larger survey that included a multiracial category.
- The response "Other Race" dropped with the inclusion of a multiracial category, indicating that inclusion of a multiracial category would eliminate some "Other Race" respondents.
- With the Hispanic origin question preceding race, and with a multiracial category, the inconsistency of reporting from Whites declined regarding race. It also reduced inconsistent reporting for the "Other Race" category. Inclusion of the multiracial category had no significant effect on consistency of response to the "Hispanic Origin" question.
- The overall percentage of individuals reporting multiracial was slightly over one percent.
- The federal lawsuit demanding a multiracial category and the results of the Census Bureau's survey rekindled a number of thorny issues:
- Race is an imprecise concept, and individuals do not use it consistently over time. Among those who identify themselves as black or African-American, 94% identify themselves by the same race the following year. Among Hispanics, that number falls to 88% (Skerry 1992, 25)
- Almost any classification regime is bound to conflate geographic and ethnic origin with race. For example, Hispanic (or Latino) is not considered a racial category, but Asian ancestry is. In what sense are Japanese and Bengalis part of the same racial group?
- Race questions ask individuals to define themselves--a thoroughly subjective act. Responses then feed into a comparatively objective statistical framework, inevitably leading to analytical problems.

The advisory committees did seem to agree one point. There was little controversy about the inclusion of a Hispanic origin question before the race question since it reduced the number of "Other Race" responses and related inconsistencies.

On October 30, 1997, OMB resolved the controversy over the use of a multi-racial category by issuing "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity" (Federal Register, 62 FR 58781 - 58790). It recently issued guidelines for how agencies are supposed to comply with the new classification system U.S. OMB 1999). All agencies are now required to allow respondents to choose more than one category when they identify their race.[61]

THE CONTROVERSY OVER SAMPLING AND THE UNDERCOUNT

Resolving the knotty issue of how we define ourselves for statistical purposes leaves unanswered the question of how best to count ourselves. The failure to enumerate large numbers of citizens in the decennial census has enormous significance, both symbolically and politically. The disproportionate undercount of minorities in the 1990 decennial census and the subsequent controversy over whether to adjust statistically for that undercount contributed to the image of the federal government as racially insensitive. (Secretary of Commerce Robert Mosbacher ultimately decided against an adjustment that would have partially corrected the undercount.)

Background: the 1990 undercount in perspective

The 1990 census was the first whose accuracy did not improve over its predecessors since Census began measuring coverage accuracy in 1940 (Skerrry 1992, 19; citing Robinson et al. 1990 and Robinson et al. 1991). In 1990 the net undercount of 4.7 million persons was 1.8 percent of the population, compared to 1.2 percent in 1980. (This 1.8 percent undercount is a 0.6 percent increase from 1980, but still an improvement of 3.6 percent over 1940. Given conditions similar to 1990, the Census Bureau predicts an undercount of 5.5 million people in 2000.) Those undercounted represented primarily hard-to-enumerate minorities, immigrant and disadvantaged populations. The differential between the white (1.3 percent) and black (5.7 percent) undercounts reached a record 4.4 percentage points.

Inaccuracies were actually worse than these figures suggest. The GAO has estimated 14.1 million gross errors in the 1990 census (US GAO 1995b, 1-5). According to its calculations, 9.7 million persons, or 3.9 percent of the population,

---

61. The categories are now American Indian or Alasaka Native, African-American or black, white, Asian, Native Hawaiian or Other Pacific Islander, or any combination of them.

were not counted at all, but overcounts (from duplicate forms and the like) and other errors offset the gross undercount. The decision not to adjust the 1990 census for the undercount provoked a flurry of unsuccessful lawsuits, some 50 in total, and assured that the decision of whether to adjust the 2000 decennial would occupy center-stage.

Sampling in order to count everyone

In Census 2000, the Census Bureau is proposing to count 90% of the population in each census tract (around 4,000 households in each one) and then to determine the characteristics of the remaining households by sampling. To the purists who think that every household should be counted directly, this seems like fraud. Yet, no matter how hard the Census Bureau tries, direct enumeration cannot work. It inevitably leaves out millions of people.

Much of the undercounting occurs because some people do not want to be counted. Undocumented aliens are understandably wary of people coming to the door asking questions about them. If you faced deportation, would you trust someone who said that Census officials would not give the immigration authorities information about you? It is remarkable that 70% of illegal aliens are actually counted, not that 30% evade the census takers.[62]

In minority neighborhoods, according to a National Research Council report, "the largest number of individuals are missed through incomplete reporting of household members rather than through failure to enumerate the households themselves" (Skerry 1992, 23; citing Citro and Cohen 1985). In 1990, the census missed 5.7% of all blacks, but 10.1% of black men from 20 to 29 years old and 12.4% of those ages 30 to 44.[63] The census missed far fewer black women. Evidently, many African-American men do not show up in the census because the respondents in the households choose not to reveal their existence.

It is significant (and not surprising) that the census process is directly affected by conditions within society. A portion of those who feel most alienated or shut out of genuine opportunities choose not to be counted. That simple fact may account for why the undercount of African Americans from 1940 to 1980 remained consistently high (around 6 percent), even as the overall undercount fell. The "race gap" (the black/non-black enumeration difference) is not a technical problem. It is a mirror of

---

62. Using that Census Bureau estimate from 1980, Peter Skerry (1992, 22) estimates that the undercount of Hispanics, not counting those who are undocumented, was only about 2.1% in 1990, only slightly above the national average. Around 800,000 undocumented aliens were not counted in the census.

63. These estimates, cited in Skerry (1992, 22) are made using sex ratios from demographic studies. The total missing from these two categories was about 720,000, accounting for 7% of the missing, but only 2.6% of the total population. Combined with uncounted illegal aliens, about 4% of the population accounted for 15% of those not counted.

the continuing failure to create opportunities for those who historically have owned little property. (Renters are also disproportionately among the undercounted.)

Better coordination with the U.S. Postal Service to confirm addresses and other technical changes may improve Census 2000 in some ways. However, the Census Bureau is not going to find those who want to escape notice by sending more enumerators into the field. In the long run, accurate counts depend on reducing the number of people who are so alienated from society that they run from the census taker.

In the near term, the only way to get an accurate count is to use estimation techniques based on sampling. In other words, rather than trying to count everyone, the Census Bureau can estimate the characteristics of the missing persons by assuming the entire group of the uncounted is the same as those who are randomly selected.

Counting everyone directly is impossible. The only way to get an accurate count is by inferring the existence of people who do not want to be observed. That can be accomplished by sampling. There are some problems with it, however, which will be discussed below. First, we turn to the fiscal reasons for sampling.

Fiscal reasons for sampling

The decennial census has experienced spiraling costs in recent decades, and practical efforts to whittle down the undercount using direct enumeration have exceeded the budgetary limits of what Congress appears willing to spend. Decennial census expenditures doubled in constant dollars from 1970 to 1980 and increased 25 percent in real terms in 1990 (U.S. GAO 1995b, 4-5). Adjusting for inflation, the census cost $10 per household in 1970, $20 per household in 1980, and $25 per household in 1990. At the same time, mail returns of census forms have declined from 78 percent in 1970 to 75 percent in 1980 to 65 percent in 1990. There is a general consensus that census forms with too many questions and too much complexity contributed to this decline. Declining mail-in rates increase expenditures on personal interviews and other means of securing responses.[64] For the 2000 decennial, the Census Bureau estimates an additional $25 million in costs for each additional percentage point increase in nonresponse, an incremental increase of $8 million over the 1990 level.

Not all of the rising costs come from falling mail-in response. The twofold increase in costs in 1980 corresponded with only a 3 percent fall in the response rate. By contrast, the 1990 decennial suffered a 10 percent decline in the mail-in response, but only a 25 percent real increase in expenditures. An NAS panel concluded that the

64. The U.S. GAO (1992) attributes declining mail-in returns to illiteracy, non-English-speaking immigrants, privacy concerns, the increasing pace of daily life, undocumented aliens, growing commercial mail and telephone solicitations, growing mistrust of public institutions, and growing numbers of nontraditional households and family arrangements.

disproportionate cost increases may be due to more labor-intensive efforts to enumerate everybody, deteriorating quality of temporary Census personnel and demands for better small-area counts (Edmonston and Schultze 1995, 53; U.S. House Committeee on Government Reform 1996, 29).

Congress will understandably be reluctant to pay for further real increases in costs.[65] Between 1984 and 1993 the Census Bureau spent $2.6 billion on all phases of the 1990 decennial. Were the census to be conducted similarly in 2000, the Bureau estimates that it would cost almost $5 billion (U.S. GAO 1995b, 4-5). The Bureau pledged to hold the line at $3.9 billion, or the 1990 cost of $25 per household after inflation, assuming it would be permitted to use sampling. However, with the Supreme Court decision requiring a full enumeration (Holmes 1999) for purposes of reapportionment of House seats, the cost is likely to escalate. The Census Bureau now plans to carry out a full enumeration *plus* a post-enumeration sampling procedure to be used for all other purposes. Nevertheless, Census has developed a variety of cost-saving techniques that will locate occupied households more accurately and make the forms easier to file.[66]

Of the approximately $1 billion in projected savings from these changes, $500 million was intended to come from sampling. GAO had previously argued that sampling at 70 percent would save the most money and increase overall accuracy (see Table D-1), but focus group research by Census suggested that such extensive sampling would have compromised public faith in the integrity of the count. However, most members of the public do not understand that sampling could provide a more accurate count than direct enumeration.

---

65. Because of budget constraints, Census has had to postpone some staffing and preparatory activities, despite highly-critical outside evaluations that point to delays in preparations for 2000 and inadequate management and decisionmaking within the agency (US DOC, Office of Inspector General 1994 and 1995).

66. The methods of improving the census include (U.S. Census Bureau 1997a):

close cooperation with the U.S. Postal Service on the creation of a continuously updated address system;

mproved, more user-friendly forms with fewer questions;

multiple mailings of reminder letters and replacement forms for those who do not respond;

multiple sites to pick up documents;

expanded state and local partnerships to get out the count;

first-ever, $100 million paid advertising campaign targeting hard-to-count populations;

telephone reporting with commercial telephone lists and use of administrative records to supplement traditional door-to-door interviews of nonrespondents; and

incorporation of sampling and adjustment for nonresponse.

TABLE D-1: ESTIMATED POTENTIAL BUDGET SAVINGS FROM SAMPLING (billions)

Assumed Mail Response Rate	Full Enumeration	Sampling Costs at Various Percentage Cut-offs			
	100%	95%	90%	80%	70%
66.9%	$4.3	$4.2	$3.9	$3.4	$3.2
56.9%	$4.7	$4.5	$4.3	$4.0	$3.4

Source: U.S. GAO 1995b

METHODOLOGICAL BACKGROUND AND DISAGREEMENTS

The official Census Bureau proposal was to count 90 percent of the population in each census tract and to sample the remaining 10 percent at a rate of 1 in 10. Results from the sample were to be used to revise the final count, a process known as adjustment. Then, in an effort called Integrated Coverage Measurement (ICM), Census would undertake a survey of 750,000 households in order to evaluate any overcounting or undercounting in the initial enumeration and sampling. Finally, these statistical adjustments would be integrated and a single tally announced by December 31, 2000.

Support for this approach to sampling in the scientific and statistical communities has run deep. A National Academy of Sciences panel, the American Statistical Association, other professional associations, and even Department of Commerce Inspector General Francis DeGeorge, a frequent critic of Census in other respects, have endorsed sampling (U.S. House Committeee on Government Reform 1996, 31). The NAS panel perhaps sums up the pragmatic rationale for sampling:

An effort to conduct a census in 2000 using 1990 methods--this is, attempting to the fullest extent to physically enumerate every household, with the funding levels that now seem probable-will likely result in a census of substantially lower quality than previous censuses. . . . A combination of sampling for nonresponse follow-up and for integrated coverage measurement is key to conducting a decennial census at an acceptable cost, with increased accuracy and overall quality, and reduced differential coverage (White and Rust 1996).

Although sampling is more accurate and reliable than a direct count, Congressional opponents have responded with a variety of criticisms (U.S. House Committee on Government Reform 1995, 1-18; 23-27; US GAO 1995b):

- The choice of one sample and its assumptions over another can yield very different outcomes over large geographic areas (as can the choice of methodology for ICM), thus rendering the final count somewhat arbitrary and

potentially subject to partisan manipulation. Since legislative redistricting is based on the census, the collection of data needs to appear impartial.

- Sampling would undermine the perceived integrity of the census at a time when survey response rates and overall confidence in government are falling.
- Sampling and adjustment would open the door to more lawsuits, not fewer. (Indeed, Census has specifically allocated funds in its budget to cover legal expenses for anticipated lawsuits.)
- Census relies on state, local and nonprofit cooperation to promote public participation, yet the expectation of certain sampling and adjustment would reduce the incentive to help "get out the count."
- Federal small-area data are already poor, and once 90 percent of a population has been enumerated, the remaining 10 percent are the most difficult to count--especially in urban and rural areas. Thus, statistical uncertainty under sampling would be magnified at lower geographic levels, possibly distorting federal funding and state-level legislative redistricting.
- Sampling and adjustment would significantly increase the operational complexity of the 2000 decennial census, when it already had difficulties carrying out the 1990 decennial. Moreover, complex statistical procedures aimed at incremental gains in accuracy can also lead to large errors.[67] This will be especially true if Census is forced to reduce quality control measures due to budget constraints.

Criticisms of sampling and adjustment come with weighty references to Constitutional obligation, and the critical need to maintain the integrity of the census in the eyes of the public. Critics' concerns would carry more weight if their proponents acknowledged the obvious: Sampling is a pragmatic response to the declining accuracy and rising costs of the decennial census.

Ultimately, the sampling issue comes down to a simple question. Should the census count everyone using statistical techniques with known flaws, or should it be carried out in a way that will leave out 10 to 15 million people? Both choices have serious drawbacks. Both choices are political. Both choices mean that legislative districts will be drawn less fairly than they would if perfect knowledge of the population were available. But since perfect knowledge is beyond human reach, a decision between two imperfect methods of counting must be made.

---

67. In 1991, a computer program error in the 1990 census was discovered that affected a million people and would have shifted one Congressional seat from Pennsylvania to Arizona if Commerce Secretary Mosbacher had decided to adjust the count (Wachter and Freedman 1996, 96).

Appendix E: AN EXAMPLE OF OUTMODED TECHNICAL SYSTEMS: THE ENVIRONMENTAL PROTECTION AGENCY

CURRENT PROBLEMS

Historically, environmental legislation has emerged haphazardly to address environmental problems in isolation from each other. Statutes require separate collection and maintenance of data on air emissions, water emissions, pesticide use or toxic waste sites, despite obvious ecological interactions. This reduces the value of the information available and imposes multiple reporting obligations on individual facilities, leading to increased costs and duplication. Many EPA officials recognize the need to streamline reporting and integrate data across environmental media (e.g., air, land and water), but they are hindered by statutory requirements for data collection.

The problem cannot be attributed fully to statutory restrictions. Agency officials have failed to develop a comprehensive data management program that would overcome the tendency toward fragmentation of monitoring and regulatory reporting. Separate databases, different computer programs, incompatible data formats and a lack of common data identifiers undermine the efficiency of the agency.

Enforcement officials, for example, cannot retrieve a complete list of a company's violations in all of its U.S. locations through a single search, even though this information is collected throughout the Agency. Agency officials cannot search a particular city or geographical location and access a broad array of environmental information for that location. Poor integration of EPA environmental data is spurring states to develop their own Geographic Information Systems (GIS)-integrated systems, raising concerns about unnecessary duplication and data incompatibility.

Finally, EPA faces the same technological hurdles that bedevil statistical programs at other agencies.[68] Out-of-date equipment and software in many offices (such as 286 processors and DOS-based applications as recently as 1996):

- Cripple communication with state and local governments and private institutions
- Prevent staff access to Envirofacts, one of EPA's few integrated information resources.
- Deprive the EPA of the capacity to make use of GIS systems, which have begun to revolutionize environmental management. (GIS systems allow planners and enforcement officials to integrate environmental and related data with physical maps and to assess and project the impacts of different economic activities, emissions and policies rapidly and at low cost.)

RECENT INITIATIVES

EPA has begun to move in the right direction by implementing new information systems:

- Envirofacts is a database that takes important steps toward providing integrated environmental data across media.
- The Toxic Release Inventory - TRI (see Appendix C for more about TRI) provides EPA officials and the public with integrated emissions data from private sector and federal facilities.
- The Key Identifiers Project develops common identifiers for chemical and other data to enable the creation of a relational database system. While the architecture of that system has yet to take shape, the goal is to provide GISintegrated, point-and-click access to all EPA data in an integrated format.

The Community-Based Environmental Protection (CBEP) program is an EPA strategy to provide local communities with the information they need to develop integrated policies and public involvement. This is a step toward increased local control over environmental management. As in the case of the TRI, the idea of CBEP is to use information rather than regulation to promote changes in harmful behavior. This is especially relevant in areas such as nonpoint source pollution where EPA lacks the statutory authority or ability to regulate effectively (US EPA 1996b, 10, 17, 23). However, nearly half of EPA regional staff interviewed see poor data collection and

68. Some of the observations on computer technology and applications are drawn from meeting minutes of the Information Impacts Committee of EPA's National Advisory Council for Environmental Policy and Technology, July 11-12, 1996.

management capabilities as a major impediment to effective community assistance. Barriers cited include:

- a lack of expertise, training and equipment for GIS systems;
- an inability to provide program and environmental media- specific data on a geographic and community basis;
- poor quality of locational data (e.g., the exact source of an ambient pollutant);
- lack of monitoring and data comparability across communities; and insufficient funding and expertise for local-area monitoring and indicators (US EPA 1996b, 10, 17, 23).

CBEP has the potential to become a positive example of devolution of environmental policy that grants communities more say in local indicators development, goal-setting and decision-making, but only if the information base exists to support it.

LOOKING AHEAD

Uniform regulatory control may have made sense in the industrial era when standardized problems outweighed local variations. Faster information processing systems have made a uniform approach to management unnecessarily wasteful.

With state-of-the-art monitoring systems, it should become increasingly possible for local governments and states to set performance targets and determine how to reach them, thereby reducing the cumbersome reporting burdens that now fall on facilities. This will only be possible if the potential capacities of information systems are put to use.

There are many possible directions to move in. Leading companies have proposed continuous emissions reporting on the Internet. New Jersey is experimenting with five-year, facility-wide permitting. This suggests the potential for a single national environmental reporting system that combines state and federal permits, TRI data, and other reporting requirements (Jeff Johnson 1996, 72). Eventually, satellite remote sensing may also provide additional data to verify a streamlined national environmental reporting system, thereby increasing accountability without generating further burdens on the private sector (Mathews 1996, A-15).

APPENDIX F: WELFARE REFORM: Statistician's FULL EMPLOYMENT ACT?

One of the best kept secrets of recent welfare reform legislation--the Personal Responsibility and Work Opportunity Reconciliation Act of 1996--is its radical implications for federal and state statistical systems. Statistics officials agree that the Act establishes unprecedented and breathtaking mandates for data collection, as well as for monitoring and reporting on state program participation, expenditures and results. One former high-ranking BLS official has dubbed it the Statistician's Full Employment Act. In reality, such a characterization misleads: The Act envisions a vast new array of statistical activities and functions, but it provides few new resources to pay statisticians to get them done.

While public debate about welfare legislation has centered around termination of the decades-long entitlement to public assistance, the Act foreshadows equally historic changes in the use of government data.[69] States must conform to strict performance standards for welfare recipients in order to avoid significant reductions in federal funding. This includes highly detailed quarterly reports that states must submit to the Secretary of Health and Human Services.

Statisticians Henry Brady and Barbara Snow (1996, 1997) were commissioned by the National Academy of Sciences to review the Act from a statistical perspective. Their observations are revealing. They were "charmed" that Congress had such "faith in the ability of public administrators, survey researchers, database managers, and statisticians to track people over time, update databases regularly and accurately, to measure work effort in enough detail to develop weekly logs of the number of hours and kinds of work undertaken by someone on assistance, and to keep track of the complicated living arrangements of modern American households in this era of single parent families" (Brady and Snow 1996, 3). They conclude that the principal means of obtaining such information--sample surveys and administrative databases-will be tested to their limits and possibly beyond. Indeed, the experience of state compliance with the Family Support Act of 1988 does not augur well for the far more complex data requirements of welfare reform. The 1988 Act established standards for the

69. The legislation requires that states impose a five-year lifetime limit on aid, that recipients reenter the workforce after receiving two years of assistance, and that parents comply with child support orders.

automation of state child support payment and collection systems, with the aim of cracking down on delinquent fathers. Only one state met the October 1995 deadline for compliance (Pear 1996, 1).[70]

Brady and Snow illustrate their concerns with their discussion of Temporary Assistance for Needy Families (TANF). This program replaced Aid for Families with Dependent Children, or AFDC, and imposed mandatory work requirements. The authors have compiled the types of data that states must now collect for quarterly reports. As Table F-1 shows, the data requirements are daunting. If reports are late, states must forfeit 4 percent of their block grant; failure to satisfy work requirements will incur a penalty of 5 percent in the first year, followed by incremental increases of 2 percent in succeeding years. Brady and Snow suggest that to generate this data, each state will have to develop mini versions of the Census Bureau's Survey of Income and Program Participation, or SIPP (Brady and Snow 1996, 8). Ironically, at the time TANF was enacted, budget cuts had delayed data collection and reduced household coverage in SIPP, the best source of information on welfare programs (US OMB 1997).

Compelling states to develop new data-gathering capacity, without federal guidance, runs the risk that they will collect data that is not nationally comparable. Indeed, California and some other states do not have a centralized system themselves, and counties have different data formats, definitions and computer languages. Brady and Snow surmise that survey design could become a bone of contention, with conservatives prefering questions emphasizing work and program participation, and liberals supporting questions about the quality of life of participants (Brady and Snow 1996, 10-11).

TABLE F-1 : DATA TO BE PROVIDED IN QUARTERLY REPORTS

FROM A SAMPLE OF FAMILIES RECEIVING "TEMPORARY ASSISTANCE TO NEEDY FAMILIES" (TANF):

Demographics:

- County of residence of family
- Number of individuals in the family and relationship of each family member to youngest child in the family
- Ages of family members
- Marital status of the adults in the family
- Race of each adult and child in the family
- Educational status of each adult and child in the family
- Citizenship

70. Others counter that the experience of states with the Family Support Act and significant improvements in computer technology will make for more successful implementation of welfare reform.

TABLE F-1 : DATA TO BE PROVIDED IN QUARTERLY REPORTS (continued)

Employment Status and Data:

- Rate of workforce participation among TANF recipient families
- Average weekly hours in work and work-related activities (education, training, provision of child care to others in program, etc.)
- Employment status, hours and earnings of the employed adult in the family
- Participation of adults in education, subsidized private sector employment, job search, job skills training or onthe-job training and vocational education
- Unearned income received by any member of the family

Social Program Use and Status:

- Type and amount of assistance received under TANF
- Whether family received subsidized housing, Medicaid, food stamps, subsidized child care, and amount received for latter two, and number of months receiving assistance under each program
- Disability of child or adult

FROM A SAMPLE OF CLOSED CASES:

- Whether the family left the program
- If the family left the program, whether the family left due to employment, marriage, five-year time limit, sanction, or state policy

Source: Brady and Snow 1996.

In its final form, the Personal Accountability and Work Opportunity Act represents a compromise between GOP conservatives, who sought a virtually unrestricted block grant to the states, and moderates, who demanded that accountability provisions be retained through extensive reporting and performance requirements. Thus, while the secretary of HHS loses the ability to regulate states, (s)he can nonetheless apply 12 specific budgetary penalties on states for nonperformance, as well as rank their performance among each other (Primus 1996).

A big risk underlies this uneasy compromise. States may find themselves simply unable to cope with the extraordinary statutory requirements for data collection and reporting, or the data a state collects might demonstrate noncompliance and threaten embarassing political exposure. In the former case, state governments might be compelled to hazard estimates; in the latter case, they might be tempted to modify inconvenient numbers. Thus, the technical, fiscal and political pressures inherent in the welfare reform bill risk undermining the integrity of state statistics.

The lack of information on current state welfare reform initiatives highlights the danger in failing to give adequate consideration to the technical and financial needs of states which must now produce outcome and performance related data under federal welfare legislation. President Clinton recently boasted declining welfare caseloads of 2.1 million people, the largest in history. Some state declines are in fact dramatic: 28 percent in Massachusetts, 40 percent in Wisconsin, 37 percent in Indiana, and 34 percent in Oklahoma (Stolberg 1996, A-1; Miller 1997, 4).

However, data on people leaving welfare are so scant that states do not know whether individuals gained employment and independence, became dependent upon a relative or friend, moved out of state, or simply fell into destitution and obscurity. Commenting on the ambitious welfare-to-work effort in Massachusetts begun in 1995, no less a figure than national welfare reform proponent David Ellwood of Harvard University observes that "[t]hey have reduced the caseload. But have they got people working? The truth is, we just don't know" (Vobejda and Havemann 1997, A-1). This concern was shared by Gov. William Weld (R), who ordered a survey to track adults who have left the state's program. At the federal level, welfare reform marks the biggest change in social policy in some 60 years, and we risk not understanding its ultimate consequences-good or bad.

In addition to the sheer volume and diversity of data to be collected, the need to share data across programs and institutions will conflict with the need to safeguard individual privacy--an area where the welfare legislation is critically weak (Brady and Snow 1996, 51-52). In order to determine whether applicants meet the Act's rigorous eligibility requirements, caseworkers at social services field offices across the country must have access to the most intimate details of people's private lives. The improper disclosure in Florida of the names of HIV-positive individuals underscores the risk to the confidentialy of TANF participants.

Heightening the risks to privacy, state and local governments are experimenting with privatization of social service delivery. Holli Ploog, senior vice president of Lockheed Martin IMS, an information management subsidiary of Lockheed Martin Corporation, observes that "[s]tates will have a lot of difficulties keeping track of time limits and myriad other things the [Act] requires." In September of last year, her company already had 33 contracts to run aspects of state welfare, food stamp and other social service programs (Pear 1996, 1). How will confidentiality be ensured in privately-administered programs, let alone state services? At the federal level, statistical agencies work under strict confidentiality rules; indeed, the frequency and exent of information sharing among federal, state and local agencies envisioned under federal welfare reform would currently be illegal, if implemented at the national level. The jury is still out as to whether devolution and private contracting of services will yield improved efficiency and performance. It is harder to imagine how these trends will guarantee continued protection of individual privacy without the blanket

extension of federal confidentiality rules to state and local governments and private contractors, and stiff sanctions for their violation.

Finally, the Act already mandates establishment of a national register of all child support orders and a national directory of newly-hired employees. States will report this information from similarly-mandated databases at the state level. The federal government will then alert states of matches in order to enforce child support payments across state boundaries. While lawmakers did not have a centralized and comprehensive national database in mind to implement welfare reform, it is an open question whether the goals of the welfare legislation can be attained without such a system--this in a country in which national identity cards are bitterly opposed by large segments of Congress and the public alike.

In its current form, the welfare reform bill is a fiscal, technical and legal quagmire for federal and state statistical agencies and programs. Recently, state officials have begun to criticize Congressional architects of the bill for heaping on them a blizzard of unfunded statistical mandates. Worse, federal officials are unsure how to respond to state requests for guidance and technical assistance because the legislation states that "[n]o officer or employee of the Federal Government may regulate the conduct of states," except where expressly permitted (Pear 1996, 1). Clearly, those activities and requirements actually supported by additional funding will enjoy the greatest likelihood of successful implementation. This is sobering. Congress has appropriated $500 million over 3 years to help fund development of new information systems as Medicaid is delinked from welfare. Otherwise, no new money exists for 50 states, some of which have scarcely begun to gear up for these statistical challenges. Congress has granted states the right to exceed the 15 percent limit on administrative overhead, if additional resources go to the development of data systems. However, this would require that funding for statistics cannabalize money for the poor--a politically and ethically hard choice to make, even though the disadvantaged would clearly benefit from accurate and comprehensive analysis of their situation.

Given the paltry additional resources to accompany these mandates, there is little confidence about the ability of agencies to comply with the spirit of the law, let alone the letter. And it is not clear that lawmakers on Capitol Hill or the White House even cared; BLS was specifically discouraged from voicing its concerns about the bill to OMB (Plewes 1996).[71] The statistical community's absence from the table while legislation was crafted, debated and revised is glaringly revealed in the final product. It is an example of what former Bush Administration OMB Director Richard Darman sees as a trend toward government as riverboat gambling: bet the farm, and let the chips fall where they may (Darman 1996, 116-17). The treatment of statistics

71. For example, BLS was prevented from voicing objections about using unemployment as a measure of need since the official figure counts only those actively seeking work and not those who have given up seeking employment.

under welfare reform risks setting a dangerous precedent for the devolution of future programs. We need only look at the human costs of AFDC's failure to appreciate the repercussions that will come from botching its reform as well.

SOURCES

ALEXANDER, CHARLES H. 1996.

Some basic technical information about the American Community Survey (draft). Demographic Statistical Methods Division, Bureau of the Census. June 12.

BERRY, JOHN. 1975.

Free Information and the Economy. *Library Journal.* Vol. 100 (May 1): 795ff.

BIJLEFELD, MARJOLIJN. 1997.

The Gun Control Debate: A Documentary History. Westport, Conn.: Greenwood Press.

BRADY, HENRY E. AND BARBARA WEST SNOW. 1996.

Data systems and statistical requirements for the Personal Responsibility and Work Opportunity Act of 1996. A paper presented to the National Academy of Sciences (NAS) Committee on National Statistics (CNSTAT). October 14.

BRADY, HENRY AND BARBARA WEST SNOW. 1997.

Recommendations for research priorities and data sources for monitoring and evaluating welfare reform. Berkeley, CA: California Policy Seminar, University of California.

BRYANT, BARBARA EVERITT AND WILLIAN DUNN. 1995.

The Census And Privacy. *American Demographics*. May. 48-54.

CENTER FOR THE STUDY OF SOCIAL POLICY (CSSP). 1995.

Making Decisions Count: How the Census Bureau's New Survey Could Transform Government. Washington, D.C.: CSSP.

CITRO, CONSTANCE F. AND MICHAEL L. COHEN, EDS.. 1985.

The Bicentennial Census. Panel on Decennial Census Methodology, Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council (NRC). Washington, D.C.: National Academy Press.

COUNCIL OF ECONOMIC ADVISERS. 1997.

Economic Report of the President. Washington, D.C.: Government Printing Office.

COUNCIL OF GOVERNORS' POLICY ADVISERS (CGPA). 1996.

The States Forge Ahead Despite the Federal Impasse: Survey of States on the Devolution Revolution. Washington, D.C.: CGPA. February.

DARMAN, RICHARD. 1996.

Riverboat gambling with government. *New York Times Magazine*. December 1: 116-17.

DRUMMOND, WILLIAM. 1995.

Address matching: GIS technology for mapping human activity patterns. *Journal of the American Planning* Association. Vol. 61, No. 2 (Spring): 240-251.

DUNCAN, JOSEPH W. 1987.

Technology, Costs, and the New Economics of Statistics. In William Alonso and Paul Starr, eds. *The Politics of Numbers*. New York: Russell Sage Foundation. 395-414.

**DUNCAN, JOSEPH W. ET AL. 1981.

Remarks of Joseph Kasputys in Private Versus Public Sector Responsibility for the Collection, Distribution, and Analysis of Statistical Data. *Review of Public Data Use*, Vol. 8 (1981).

DUNCAN JOSEPH W.AND WILLIAM C. SHELTON. 1978.

Revolution in United States Government Statistics, 1926-1976. Washington, D.C. U.S. Department of Commerce.

EDMONDSON, BRAD. 1997.

Will Super Survey Fly? The proposed American Community Survey is a cost-effective way to bring federal data up to date. So far, Congress doesn't see the point. *American Demographics*. April. 8-13.

EDMONSTON, BARRY AND CHARLES SCHULTZE, EDS. 1995.

Modernizing the U.S. Census, Panel on Census Requirements in the Year 2000 and Beyond. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council (NRC). Washington, D.C.: National Academy Press.

FLEENOR, PATRICK ED. 1995.

Facts and Figures on Government Finance. 30th edition. Washington, D.C.: Tax Foundation.

GEISER, KENNETH. 1995.

Tracking Pollution Prevention Progress in Massachusetts. Lowell, Mass.: Massachusetts Toxics Use Reduction Institute. December.

GELMAN, KENNETH. 1995.

Life without the census; financial services. *American Demographics.* October: 42.

GELLMAN, ROBERT. 1995.

"Twin Evils: Government Copyright and Copyright-like Controls over Government Information," *Syracuse Law Review*, Vol. 45, No. 3: 999-1072.

HANSEN, MARK. 1997.

A Question of Influence: Critics see dividends in political contributions by former publisher. American Bar Association Journal. Vol. 83. June. 6-7.

HARDERS, JULIE. 1995.

Pay and pay again; access to public records. The Quill. Vol. 83, No. 8 (October): 17-18.

HAYEK, FRIEDRICH. 1945.

The Use of Knowledge in Society. American Economic Review 35: 519-30.

HERBERT, BOB. 1996.

More N.R.A. Mischief. *New York Times*. July 5. A-23.

HODGES, KEN. 1995.

Life without the census; business demographics. *American Demographics.* October: 39.

HOLMES, STEVEN A. 1999.

Ruling Said to Raise Census Cost by $2 Billion. *New York Times*. February 24. A-18.

INNES, JUDITH ELEANOR. 1990.

Knowledge and Public Policy: The Search for Meaningful Indicators. Second expanded edition. New Brusnwick, NJ. Transaction Publishers.

JOHNSON, TOM. 1996.

Study finds pollution law works. *The Star-Ledger* (Newark: New Jersey). November 11.

JOHNSON, JEFF. 1996.

New Jersey pilot program eases permit path for companies that cut toxics. *Environmental Science & Technology News.* Vol. 30, No. 2.

JOINT AD HOC COMMITTEE ON GOVERNMENT STATISTICS. 1976.

Report of the Committee. *Statistical Reporter*. Washington, D.C.: U.S. Government Printing Office. September.

KATES, DON. B., HENRY E. SCHEFFER, PH.D., JOHN K. LATTIMER, M.D., GEORGE B. MURRAY, M.D., AND EDWIN H. CASSEM, M.D. 1995.

Guns and public health: epidemic of violence or pandemic of propaganda. Tennessee Law Review. Vol. 62 (Spring): 513-596.

KENWORTHY, TOM. 1995.

By Any Name, Biological Service Appears to Be Endangered Species. *The Washington Post*. June 27. A-15.

KINTNER, HALLIE. 1995.

Life without the census; big business. *American Demographics.* October: 41.

LANDEFELD, STEPHEN (DIRECTOR OF THE BUREAU OF ECONOMIC ANALYSIS). 1996.

Interview. July. Washington, D.C.

LETTER TO SENATORS MOYNIHAN AND KERREY. 1996.

In support of S. 2119, A Bill to Establish the Commission to Study the Federal Statistical System. 104th Congress, 2nd Session.

LI, C. M. 1962.

The Statistical System Of Communist China. Berkeley, Calif.: University of California Press.

LOHR, STEVE. 1996.

Administration renews efforts on prevention of repetitive motion injuries. *The New York Times.* December 11. A-24.

LOVE, JAMES P. 1992.

The Marketplace and Electronic Government Information. *Government Publications Review*. Vol. 19: 397-412.

LOVE, JAMES. 1995A.

Pricing Government Information. In *Agenda For Access.* Washington, D.C.: Bauman Foundation. ch. 10. Available at <http://www.rtk.net/S607>.

LOVE, JAMES. 1995B.

Four Years of Struggles to Free the Law. *Background Comments for Conference on Computers, Freedom and Privacy*, 1995 (CFP-95). Panel on Who Owns the Law. Friday, March 31. San Francisco. The text can be found at: http://lists.essential.org/1995/info-policy-notes/msg00131.html

LOVE, JAMES. 1995C.

SEC Press Release - EDGAR on Internet for Free. *Information Policy Note*. Washington, D.C.: Taxpayer Assets Project. August 28. http://lists.essential.org/1995/info-policy-notes/msg00157.html

LOVE, JAMES. 1995D.

SEC Edgar Database May Not Last. *Information Policy Note*. Washington, D.C.: Taxpayer Assets Project. October 3. http://lists.essential.org/1995/info-policy-notes/msg00160.html

LOVE, JAMES. 1995E.

Internet Community KO's Anti-FOIA Provision. *Information Policy Note*. February 15. http://lists.essential.org/1995/info-policy-notes/msg00120.html

MACHLUP, FRITZ. 1962.

The Production and Distribution of Knowledge in the United States. Princeton, N.J.: Princeton University Press.

MACLEAN, ALAIR AND PAUL ORUM. 1992.

Progress Report: Community Right to Know. Washington, D.C.: Working Group on Community Right-To-Know and OMB Watch. June-July. <www.rtknet.org/E2617T1005>

MARKOE, LAUREN. 1997.

10 firms honored for cutting pollution. *The Patriot Ledger* (Quincy, MA). December 4: 9.

MASSACHUSETTS. EXECUTIVE OFFICE OF ENVIRONMENTAL AFFAIRS. 1996.

Governor's Award Program for Outstanding Achievement in Toxics Use Reduction. May 15.

MATHEWS, JESSICA. 1996.

An opening for environmentalists. *The Washington Post*. April 15.

MCMILLEN, DAVID AND EDWARD SPAR. 1986.

Data sharing: a star or a comet? *Government Information Insider*. Spring: 10-12.

MILLER, GEORGE A. 1997.

State Scoring Welfare Successes. *The Daily Oklahoman*. March 6. Point of View, 4.

MILLS, FREDERICK C. AND CLARENCE D. LONG. 1949.

The Statistical Agencies of the Federal Government: A Report with Recommendations. Prepared for the Commission on Organization of the Executive Branch of the Government. Publications of NBER, No. 50. New York: National Bureau of Economic Research.

MORGAN, DAN. 1995.

Steep Learning Curve for Appropriators; House Newcomers Test Power Held by 'College of Cardinals.' *The Washington Post*. November 2. A-1

MOYNIHAN, SENATOR DANIEL PATRICK. 1996.

Statement in *Congressional Record*, Vol. 142, No. 134, September 25: S11296.

MYERS, DOWELL. 1995.

Life without the census; local government. *American Demographics.* October: 38-39.

NATIONAL RESEARCH COUNCIL (NRC). 1979.

Panel on Privacy and Confidentiality as Factors in Survey Response. *Privacy and Confidentiality as Factors in Survey Response.* Washington, D.C.: National Academy of Sciences.

NATIONAL RESEARCH COUNCIL (NRC) AND SOCIAL SCIENCE RESEARCH COUNCIL (SSRC). 1993.

Panel on Confidentiality and Data Access of the National Academy of Sciences. *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Washington, D.C.: National Academy Press.

NORWOOD, JANET. 1995.

Organizing to Count. Washington, D.C.: Urban Institute Press.

OBEY, DOUG AND ALBERT EISELE. 1996.
West: A study in special interest lobbying. The Hill. February 22: 1.

PAGE, PETER. 1996.
Gauging efforts to curb pollution. *The Times* (Trenton, NJ). October 28. A1, A10.

PEAR, ROBERT. 1996.
A computer gap is likely to slow welfare changes. *New York Times*. September 2: 1.

PLEWES, THOMAS J. (FORMER BLS ASSOCIATE COMMISSIONER). 1996.
Comments on Brady and Snow's presentation to the Committee on National Statistics, Lecture Room, National Academy of Sciences, Washington, D.C. October 25.

POPULATION RESOURCE CENTER (PRC) AND COUNCIL OF PROFESSIONAL ASSOCIATIONS OF FEDERAL STATISTICS (COPAFS). 1996.
Issues Facing the Federal Statistical System in an Era of Block Grants. Washington, D.C. PRC and COPAFS. January.

PRIMUS, WENDELL (FORMER HHS DEPUTY ASSISTANT SECRETARY). 1996.
Comments on Brady and Snow's presentation to the Committee on National Statistics, Lecture Room, National Academy of Sciences, Washington, D.C. October 25.

**REICHMAN, J. H. AND PAMELA SAMUELSON. 1999.
Intellectual Property Rights in Data: An Assault on the Worldwide Public Interest in Research and Development. Vanderbilt Law Review. Vol. 50 (January): get page numbers and put in fn in DW-2.

RICHE, MARTHA FARNSWORTH (DIRECTOR, BUREAU OF THE CENSUS AT TIME OF INTERVIEW). 1996.
Interview. June 12. Washington, D.C.

ROBINSON, J. GREGORY, PRITHWIS DAS GUPTA, AND BASHIR AHMED. 1990.
Evaluating the Quality of Estimates of Coverage Based on Demographic Analysis. Presentation at meeting of Population Association of America. May 3.

ROBINSON, J. GREGORY, BASHIR AHMED, PRITHWIS DAS GUPTA, AND KAREN WOODROW. 1991.
Estimating Coverage of the 1990 United States Census: Demographic Analysis. Presentation at meeting of American Statistical Association. August 20.

RUDDICK, MICHELLE M. 1996.
Characteristics of Federal-State Data Collection Systems: Perspectives from Federal and State Agencies. Working Paper for the Committee on National Statistics. Washington, D.C. National Research Council. September.

RULE, JAMES AND LAWRENCE HUNTER. 1996.
Privacy wrongs: corporations have more right to your data than you do. *The Washington Monthly*. November: 18-19. 17-20.

SAWICKI, DAVID S. AND WILLIAM J. CRAIG. 1996.
The democratization of data: bridging the gap for community groups. *Journal of the American Planning Association*. Vol. 62, No. 4 (Autumn): 512-523.

SAWICKI, DAVID S. AND PATRICE FLYNN. 1996.
Neighborhood Indicators: A Review of the the Literature and an Assessment of Conceptual and Methodological Issues. *Journal of the American Planning Association*. Vol. 62, No. 2 (Spring): 165-183.

SCHMICKLE, SHARON AND TOM HAMBURGER. 1995A.

West Publishing and the courts; U.S. justices took trips from West Publishing. *Star Tribune* (Minneapolis, MN) A-1. March 6. http://www.startribune.com/westpub/west.htm

SCHMICKLE, SHARON AND TOM HAMBURGER. 1995B.

West and the Supreme Court; Members accepted gifts and perks while acting on appeals worth millions to Minnesota firm. *Star Tribune* (Minneapolis, MN). A-16

SCHMICKLE, SHARON AND TOM HAMBURGER. 1995C.

High stakes and hot competition; In face of change, West Publishing fights to maintain its lead in legal publishing. *Star Tribune* (Minneapolis, MN). March 6. A-1. http://www.startribune.com/westpub/west.htm

SKERRY, PETER. 1992.

The Census Wars. *The Public Interest*. No. 106 (Winter): 17-31.

SKERRY, PETER. 1996.

Many American Dilemmas: The Statistical Politics of Counting by Race and Ethnicity. *Brookings Review*. Summer: 36-39.

SPAR, EDWARD J. 1996.

The federal statistical system in an era of block grants. *Business Economics*. Vol 2., No. 31. April: 7-11.

STARR, PAUL AND ROSS CARSON. 1987.

Who will have the numbers? The rise of the statistical services industry and the politics of public data. In *The Politics of Numbers*, ed. William Alonso and Paul Starr (New York: Russell Sage Foundation, 1987), pp. 434-35.

STEELE, SHARI. 1996.

Letter to Mr. Keith M. Kupferschmid, Commissioner of Patents and Trademarks, November 22. Published at www.eff.org/pub/Intellectual_property/eff_wipo_19961122.comments.

STOLBERG, SHERYL. 1996.

U.S. welfare rosters are down 15%, Clinton says; Assistance: President attributes elimination of 2.1 million people from rolls in four years to his pledge to reform public aid. California's higher rate bucks national trend. *Los Angeles Times*. December 8. A-1.

SUPLEE, CURT. 1996.

House to consider "ergo rider" restraints on OSHA. *The Washington Post*. July 11. A-4.

THOMAS, RICHARD K. 1995.

Life without the census; health care. *American Demographics.* October: 40-41.

THOMPSON, GRANT P. 1994.

Making Public Libraries Centers of Electronic Dissemination. In *Agenda for Access.* Washington, D.C.: Bauman Foundation. Ch. 7. Available at <http://www.rtk.net/S607>.

UNITED STATES. BIPARTISAN COMMISSION ON ENTITLEMENT AND TAX REFORM. 1995.

Final Report to the President. Washington, D.C.: U.S. Government Printing Office. http://fedbbs.access.gpo.gov/libs/wh_ntitl.htm

UNITED STATES. BUREAU OF THE CENSUS. POPULATION DIVISION. 1996.

Findings on Questions on Race and Hispanic Origin Tested in the 1996 National Content Survey. Working Paper No. 16. Washington, D.C.: Bureau of the Census. December.

UNITED STATES. BUREAU OF THE CENSUS. 1997A.

Report to Congress: The Plan for Census 2000. Washington, D.C.: Bureau of the Census. July.

UNITED STATES. BUREAU OF THE CENSUS. 1997B.

Census Bureau Submits Subjects for Census 2000 to Congress. Press release. Washington, D.C.: Bureau of the Census. March 31.

UNITED STATES. CONGRESS. HOUSE. COMMITTEE ON GOVERNMENT REFORM AND OVERSIGHT. 1996.

Sa*mpling and Statistical Adjustment in the Decennial Census: Fundamental Flaws.* 104th Congress, 2nd Session. Washington, D.C.: U.S. Government Printing Office. September 24.

UNITED STATES. CONGRESS. SENATE. 1996.

A Bill to Establish the Commission to Study the Federal Statistical System. S. 2119, 104th Congress, 2nd Session. September 25.

UNITED STATES. CONGRESS. SENATE. COMMITTEE ON FINANCE. ADVISORY COMMISSION TO STUDY THE CONSUMER PRICE INDEX (BOSKIN COMMISSION). 1996.

Toward a More Accurate Measure of the Cost of Living: final report to the Finance Committee. Washington, D.C.: U.S. Government Printing Office.

UNITED STATES. DEPARTMENT OF COMMERCE (DOC). OFFICE OF THE INSPECTOR GENERAL. 1994.

2000 Decennial Census Data Capture Needs Management Action. Washington, D.C.: *U.S. Government Printing Office.* September.

UNITED STATES. DEPARTMENT OF COMMERCE (DOC). OFFICE OF THE INSPECTOR GENERAL. 1995.

Inadequate Design and Decision-Making Process Could Place 2000 Decennial at Risk. Washington, D.C.: *U.S. Government Printing Office.* November.

UNITED STATES. DEPARTMENT OF TRANSPORATION (DOT). BUREAU OF TRANSPORTATION STATISTICS. 1996.

Implications of Continuous Measurement for the Uses of Census Data in Transportation Planning. Washington, D.C. April. description and ordering information at: <http://www.bts.gov/programs/btsprod/implications.html>

UNITED STATES. ENVIRONMENTAL PROTECTION AGENCY (EPA). 1995.

Report to President Clinton: Expansion of Community Right-to-Know Reporting to Include Chemical Use Data: Phase III of the Toxics Release Inventory. Washington, D.C.: U.S. EPA. October. http://gate.cruzio.com/~meuser/scruztri/docs/c2.htm

UNITED STATES. ENVIRONMENTAL PROTECTION AGENCY (EPA). OFFICE OF POLLUTION PREVENTION AND TOXICS. 1996A.

Chemicals in the Environment: Public Access Information, TSCA at Twenty: The Toxic Substances Control Act at Twenty. (7407) EPA 749-R-96-001. Fall. <http://www.epa.gov/opptintr/cie/issue04j.htm>

UNITED STATES. ENVIRONMENTAL PROTECTION AGENCY (EPA). OFFICE OF SUSTAINABLE ECOSYSTEMS AND COMMUNITIES. 1996B.

An Assessment of EPA Regional Offices' Community-Based Environmental Protection Need: Final Report. Washington, D.C.: April.

UNITED STATES. EXECUTIVE OFFICE OF THE PRESIDENT. OFFICE OF MANAGEMENT AND BUDGET (OMB). 1997.

Statistical Programs of the United States Government: Fiscal Year 1997, Washington, D.C.: U.S. Government Printing Office.

UNITED STATES. EXECUTIVE OFFICE OF THE PRESIDENT. OFFICE OF MANAGEMENT AND BUDGET (OMB). 1999.

Draft Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity. Report by the Tabulation Working Group of the Interagency Committee for the Review of Standards for Data on Race and Ethnicity. February 17. Washington, D.C.: Office of Management and Budget. This document can be downloaded as a PDF file from: http://www1.whitehouse.gov/OMB/inforeg/index.html

***UNITED STATES. FEDERAL STATISTICAL REORGANIZATION PROJECT [BONNEN PROJECT]. 1980.

Report. Washington, D.C.: U.S. Government Printing Office.

UNITED STATES. GENERAL ACCOUNTING OFFICE (GAO). 1988.

Block Grants: Federal-State Cooperation in Developing National Data Collection Strategies: Report to Congressional Requesters. GAO/HRD-89-2. Washington, D.C.: U.S. Government Printing Office. November.

UNITED STATES. GENERAL ACCOUNTING OFFICE (GAO). 1992.

Decennial Census: 1990 Results Show Need for Fundamental Reform. GAO/GGD92-94. Washington, D.C.: U.S. Government Printing Office. U.S. Government Printing Office. June.

UNITED STATES. GENERAL ACCOUNTING OFFICE (GAO). 1995A.

Economic Statistics: Status Report on the Initiative to Improve Economic Statistics. GAO/GGD-95-98. Washington, D.C.: U.S. Government Printing Office. July.

UNITED STATES. GENERAL ACCOUNTING OFFICE (GAO). 1995B.

Decennial Census: Fundamental Design Decisions Merit Congressional Attention. Testimony of *L. Nye Stevens* before the Subcommittee on National Security, International Affairs, and Criminal Justice, Committee on Governmental Reform and Oversight, House of Representatives. GAO/T-GGD-96-37. Washington, D.C.: U.S. Government Printing Office. October.

UNITED STATES. PRESIDENT'S COMMISSION ON FEDERAL STATISTICS 1971.

Report of the President's Commission. Two volumes. Washington, D.C.: U.S. Government Printing Office.

USEEM, JERRY. 1996.

U.S. Data Worst in the World--and Getting Worse. *Inc. Magazine*. October: 26.

VOBEJDA, BARBARA AND JUDITH HAVEMANN. 1997.

Do shorter welfare rolls spell success? *The Washington Post*. January 5. A1

WACHTER, KENNETH W. AND DAVID A. FREEDMAN. 1996.

Testimony before the Committee on Government Reform and Oversight, House of Representatives. *Census 2000: Putting Our Money Where It Counts*. 104th Congress, Second Session. Washington, D.C.: U.S. Government Printing Office. February 29: 94-97.

WALLICH, PAUL. 1996.

Flying Blind. *Scientific American* January: 32.

WARDELL, KEITH. 1995.

Life without the census; direct marketing. *American Demographics.* October: 39-40.

WEISS, JANET A. AND JUDITH E. GRUBER. 1984.

Deterring Discrimination with Data. *Policy Sciences*. Vol. 17, No. 1 (May): 49-66.

WESTIN, ALAN F. NO DATE.

Privacy & American Business. Newsletter published by the Center for Social & Legal Research and available from Two University Plaza, Suite 414, Hackensack, NJ 07601; telephone (201) 996-1154.

WHITE, ANDREW A. AND KEITH F. RUST, ED. 1996.

Sampling in the 2000 Census: Interim Report I. Panel to Evaluate Alternative Census Methodologies, Committee on National Statistics, National Research Council, Washington, D.C.: National Academy Press. June.