Journal of Official Statistics, Vol. 15, No. 3, 1999, pp. 353–366

The Future of Federal Statistics in the Information Age

Norman M. Bradburn[1]

It is a great honor for me to be invited to deliver the 1997 Hansen Memorial Lecture. I first came to know Morris when he became a trustee of NORC. When I became NORC Director in 1967, one of the first things I did was to come to Washington to get his advice on how to run an organization devoted to social statistics. He was always helpful, and I shall always treasure his advice. I think Morris would be pleased with my topic today because he was interested in much more than the statistical methods for which he is best known. He was a leader at the U.S. Census Bureau in focusing attention on issues of relevance, timeliness, and validity that I will be discussing today.

Statistics is one of the most important modes of knowledge about society. There have been some excellent works on the history of statistics and the evolution of statistical thinking, such as Steve Stigler's *History of Statistics* and Porter's *Rise of Statistical Thinking* (Stigler 1986; Porter 1986) and on specified institutions, such as Margo Anderson's wonderful history of the U.S. Bureau of the Census (Anderson 1988). But there has been relatively little written about what Paul Starr has called the ''sociology of statistics'' (Starr 1987). By the term ''sociology of statistics'' I mean viewing statistics as a mode of thinking about society, reflecting on the uses of statistics, particularly in public policy debates, and critically examining its institutional structure.

I would like to explore a little in the sociology of statistics. My central thesis is that in the information society there will be a very large increase in the demand for statistics to be used in debates about public policy issues. The federal statistical system, however, is not well situated to respond to this increase in demand. I shall concentrate on social and economic statistics, and most particularly on federal statistics, but I hope my remarks may have some relevance beyond this limited purview.

My talk will be divided into three sections. First, I will look at statistics as a mode of knowing about society and how it differs from other, competing ways of knowing. Second, I will examine the effect of new technologies on the way statistics are gathered and disseminated and the implications of these changes for the statistical enterprise. Finally, I shall reflect on the institutional structures needed to respond adequately to these developments.

My discussion is natually influenced by my experiences as part of an organization that conducts large-scale sample surveys for federal statistical agencies. It is also heavily conditioned by my work on numerous panels and committees that have served in advisory capacities to statistical agencies, most importantly by my membership on the Committee on National Statistics at the National Research Council. For reasons that I am sure those of you who work for federal agencies can appreciate, I want to make the usual disclaimer that

[1] University of Chicago, NORC, 1155 East 60th Street, Chicago, IL 60637-2799, U.S.A.

what I say today represents my own personal opinions and not those of CNSTAT or the NRC, or, I suppose I should add, of NORC or the University of Chicago.

1. Statistics As a Way of Knowing about Society

In his acceptance speech at the Democratic National Convention in August 1996, President Clinton cited 27 social and economic facts about the nation, comparing them to those that obtained when he was elected. These facts ranged over a great variety of topics, from better known indicators, such as the unemployment and inflation rates, to lesser known facts about the economy – for example, that 4.4 million more people owned their own homes – about social conditions – for example, that 1.8 million fewer persons were on welfare – about crime – for example, that 100,000 more police were on the streets – about health – for example, that the life expectancy of AIDS patients had doubled. Of course, President Clinton put forth these facts as an argument for the effectiveness of the policies of his administration. In contrast, Senator Dole cited no facts in his acceptance speech.

The more interesting observation to me, however, is that President Clinton was using quantitative statistical data in political argumentation and that his audience accepted this use as a valid means of argumentation. One might also marvel at the organizational infrastructure that was necessary to produce the data bases from which the numbers were derived and, more importantly for my thesis, the system that enabled Clinton's speech writers to recover the numbers in a timely fashion.

Since many of you work in agencies that are part of that infrastructure and may even have had a hand in producing or disseminating those numbers, it may not strike you as in any way remarkable that such numbers are produced and disseminated. And yet, two hundred years ago, the thought that one might want such numbers or that an audience would accept such argumentation would be foreign to all except the most advanced thinkers. A hundred years ago, few would have even thought that such numbers were important enough to produce or known how to produce them, even if wanted. Twenty-five, perhaps even ten years ago, assembling the numbers for the speech would have been a major undertaking. Statistics are so much a part of our modern thinking that we take them and their production for granted.

By the term ''statistics'' I mean the collecting of information in numerical form and organizing that information in systematic ways. ''Statistics as a way of knowing'' means using that numerical information to answer questions about phenomena. In so doing, we use techniques devised by statisticians. For our purposes we are interested in the use of statistics to answer questions about social and economic phenomena. For historical accuracy I should note that the term ''statistics'' was not originally restricted to quantitative information but referred to information about political states, whether or not it was numerical. Its modern sense emerged gradually, but by the end of the 19th century, ''statistics'' was synonymous with numerical information (Starr 1987; Porter 1986).

How else, one might ask, does one know about social and economic conditions? The traditional way, which is still often more powerful than statistics, is through personal experience or, more generally, vicarious personal experience, which includes the report of individual experiences by acquaintances or, perhaps more commonly, by the mass media. Statistical thinking requires skills like understanding concepts, knowing how to

relate different sets of numbers, and knowing why two sets of numbers that supposedly measure the same thing differ. Personal experience is immediate and does not involve numbers. It is experience of people as individuals who incidentally may have a number of social characteristics that would permit generalization. Attention centers on people with names and unique histories, not on frequencies of abstract categories where individual qualities have been averaged out. To put it somewhat pejoratively, it is a world of anecdotes rather than data. It is also a world of small and biased samples, and this is what makes it so suspect in the eyes of statistical thinkers. But it is a world of great phenomenological reality.

We all know the capacity of the anecdote to overpower the best statistical data. It is a much lamented fact that, for people who are not receptive to statistical arguments, a single dramatic example or description of a personal event can have greater persuasive power than careful quantitative analysis. Generalizing from a small, biased sample is a very human activity. We all engage in it from time to time, even when we know better, because we lack the time or resources to assemble the statistical data we might ideally want. But for some people, the anecdote is in principle a better basis for knowledge than statistics because it has a human face instead of a set of numbers. It seems more real and has more persuasive power.

Another way we know about society and how it works is through theory. While the hallmark of the modern scientific conception of theories is that they are capable of falsification, some people hold theories about human behavior and social institutions that are so general no data can disprove them. Or they believe in their theories so strongly that data cannot shake their convictions. While theories may be formed on the basis of personal experience or even statistical data, the theories themselves can become independent of the grounds on which they were formed and become the basis on which new knowledge is acquired. Because social experiments are difficult or often impossible to carry out, it is virtually impossible to test social and economic theories definitively. Thus, it is relatively easy for credible theories to override data.

Some people accept knowledge of society from authorities. Authority may be based on reputation, as with some media figures or experts. It may be based on faith in divinely inspired knowledge, on personal charisma, or on tradition. Whatever the basis for the authority, the authority figure is the source and legitimation for knowledge. Statistical data are of no use except to support the knowledge of the authority.

Of course, few people rely on only one way of knowing, and most people recognize the merits of all modes of knowing as appropriate about some topics. People vary considerably in their preferred modes of knowing. There are relatively few people for whom statistical thinking is the dominant mode of knowing, or at least not as many as we would like. Even in government and public policy, areas that historically gave rise to statistics, hence its name, the role of statistics and statistical thinking is not as great as we believe it should be. I believe that this situation will change as result of the spread of computer technology.

2. A Technological Revolution

We are in the early stage of a technological revolution that will dramatically increase the demand for statistics and their use in policy debates, as well as in many other areas of

society. This revolution is created by the advent of the Internet and the World Wide Web. Statistics that used to be difficult to access, or that could be found only in published volumes, can now be routinely consulted in a matter of minutes, can be downloaded to one's own computer or, in a growing number of instances, can be manipulated on line. The new web site FED STATS is a cornucopia of data on every conceivable subject, with more to come. In a few years we will regard these efforts as hopelessly primitive, as sites become faster, more complete and easier to use. We can easily drown in data.

It is impossible to foresee all the consequences of this enormous increase in the ease of obtaining access to vast data sources. Inevitably there will be many tendentious uses, as well as many mistakes. The mass media will ransack data files looking for quantitative data to buttress stories built around anecdotal evidence. One would hope that the general public will develop greater quantitative sophistication as demand increases for quantitative analyses based on these data. But I am not optimistic that this will happen. It is more likely that an industry will grow up to help unsophisticated users analyze and interpret data. As it becomes easier to compare data about the same phenomena from different sources, however, discrepancies will be more likely to be noticed, and there will be pressure for greater coordination in data definition and collection.

Many of these likely consequences will benefit the federal statistical system in the long run, but will cause problems in the short run because they will increase the demand on statistical agencies to explain themselves and their activities to more people in shorter periods of time. Increased demand is also apt to increase criticism of agencies by raising the visibility of weak points that the agencies may already be aware of, or by pointing out problems that the agencies have not been aware of. This will be an added burden on their already strained resources, but, in the long run, it should lead to more support for the statistical enterprise as its products come to be regarded as even more important.

What are the main challenges that this greater accessibility of data presents to the federal statistical system? I believe that three principal challenges must be faced if the system is to successfully meet the demands placed on it by this technological revolution. None of them is particularly technological in itself, although the new technologies may permit solutions that would not be possible otherwise. The challenges are more conceptual and organizational in nature. The three challenges are discussed under the rubrics of Relevance, Validity, and Timeliness.

2.1. Relevance

Relevance refers to the questions to which the statistics gathered are the answers. How do we decide what data to collect and to make publicly available? As more people can access data more easily, and the demand for statistical data grows, controversy over what data to collect will also grow. With limited budgets, the statistical agencies will come under increasing public scrutiny, and decisions about what data to collect take on added importance.

The framing of information influences its perceived relevance. By ''framing'' I mean the psychological context within which something is understood. As we know from many psychological experiments and from the study of context effects in questionnaire design, the way questions are framed, affects how people interpret questions and what

types of thoughts, information, memories, etc. they deem relevant in answering the questions. So, too, does the way questions are framed affect the perceived relevance of statistics for policy issues.

Consider, for example, data about carbon dioxide. Among tables about manufacturing in the 1997 Statistical Abstract, we find that the United States manufactured 2,068 million cubic feet of carbon dioxide in 1982. Data on carbon dioxide production go back at least to 1970. After 1982, however, the series stops. On the other hand, in FED STATS, we find under data about the atmosphere that in 1982 the U.S. emitted 4,827.4 million metric tons of gas into the atmosphere. That number had risen to 5,167.1 million metric tons by 1993, the last year for which data are available. Clearly, something happened in the 1980's to change the framing of questions about carbon dioxide that resulted in changes in both the kind of statistics gathered and how they are reported.

What happened, of course, is that scientists raised the possibility that an increase in the concentration of carbon dioxide in the atmosphere due to increased emission from human activity is producing a ''greenhouse effect.'' The most serious consequence of this effect will be global warming, that is, an increase in the average temperature thoughout the world, with attendant changes in agricultural productivity in many parts of the world and in the distribution of people. While there has been a great deal of dispute about the extent and even reality of the ''greenhouse effect'' as a consequence of human activity, there is now sufficient evidence that something is going on with regard to world climate so that it is prudent for governments to begin to take action. Thus, there arose a demand for new kinds of data relevant to this newly framed problem.

This change in the framing of questions about carbon dioxide illustrates my central point about relevance, which is that the classification of data, indeed the very definition of what constitutes data, is a relative matter. Whether things are viewed as ''data'' worthy of being measured lies in the question being asked, not in the things themselves. Facts about carbon dioxide become data because someone asks questions about CO_2 and is able to convince someone else to allocate the resources necessary to collect information and to store it so that it can be accessed by those who want answers to the questions. Facts become data when they are available to those asking questions about things like carbon dioxide. Facts become ''policy-relevant data'' only when someone starts asking questions about policies to deal with a phenomenon, such as the ''greenhouse effect,'' and perceives that specific data are needed to shed light on the policy-relevant questions.

How do we decide what data to collect in order to meet policy and other information needs? There is no simple answer to this question. In order to carry out their functions, government agencies need information on an ongoing basis about their areas of responsibility. Fulfilling these needs is commonly the responsibility of government statistical agencies, although government officials may not rely exclusively on official statistics.

The relation between the responsibilities of government agencies and their responsibility to collect data about different aspects of society has been a source of continual disagreement. From the beginning of the United States, there has been a conflict between those who believe that the government should collect only data that are directly relevant to its mission and others who argue for a broader conception. Those who argue for a broader conception would have government collect data on aspects of society that may be of broad interest to citizens or legislators, but not necessarily of immediate use.

This argument goes back at least to the constitutional debates. In debating the provision for a decennial census, minimalists argued that the census should merely enumerate the population, because the purpose of the census was to establish the number of people for apportionment and tax purposes. Others, notably James Madison, argued for extending the census ''so as to embrace some other objects besides the bare enumeration of the inhabitants.'' He understood the low marginal cost of collecting extra data and argued for collecting information on what we would now consider social and demographic characteristics of the population and on economic conditions. Although Madison lost the argument in the first census, later censuses, beginning in the mid-19th century, expanded the number of topics covered in the decennial census. While the language of the debate in the late 18th century was different, the basic arguments are familiar to those following contemporary debates in the Congress over the necessity for the long form in the census.

''Official statistics'' are the output of the government statistical system. They have a special status in that they are used to support the formulation, implementation and evaluation of government policies. They are also increasingly used in budget, tax, social welfare, and energy models that are widely used by both the executive and legislative branches of the federal government to simulate the effects of proposed legislation and changes in government programs. It would be difficult to overestimate the importance of these models in contemporary government policy making. The quality of the statistical base that provides the numbers going into the models, thus, is of vital importance for the utility of the models. As Citro and Hanushek (1991, p. 24) put it in a comprehensive study of micro simulation modeling:

''Today, whatever the policy issue, 'the numbers' play a prominent role. Indeed, in Washington ... neither top administration officials nor members of Congress will move very far to develop legislation in the absence of detailed estimates of the cost and other effects of the proposed changes .... They treat the estimates not only as informative but often, in the case of costs, as binding.''

Because of their special status official statistics reflect both the concerns of government and the areas where government does not wish or dare to tread. As Paul Starr has pointed out (Starr 1987, p. 41), the political agenda and the official agenda are integrally bound to each other:

''The boundaries of official inquiry are the statistical counterparts to the boundaries of the political agenda, and it is an elementary point of political analysis that the control of such boundaries is a critical face of power. Just as statistical blackouts testify to deeper processes at work in a society, so do the patterns of statistical blind spots – that is, the anomalous lacunae in official facts. To make an official count of some phenomenon is often to confer recognition that the phenomenon is real and to risk that its measurement will embarass those in authority .... Even democratic societies decide that some subjects are too sensitive or volatile for political discussion – or for official statistics.''

Religion is perhaps the most obvious example of a topic that is off limits for official statistics in the United States. The controversies surrounding the implications of the constitutional separation of church and state extend to many different areas, even to a simple

enumeration of people's religious affiliation. The United States is the most religious of the developed countries of the world, but one would never know this from official statistics. We know it only from privately financed surveys or from academically based surveys that might have some government financing through research grants. Even in the case of the NSF-funded General Social Survey, most of the extensive questions on religion are privately financed.

Priorities in data collection must be continually examined for their relevance to questions being asked by users. Because of their special role in the policy process, government statistical agencies in particular must have a mechanism to review their programs and assess the needs for statistics in the areas of their responsibilities. I shall return to this point at the end of my talk.

2.2. Validity

Official statistics play a special role in the policy process and, therefore, should be of the highest quality. While government officials may rely on ''the numbers'' as played through models, they typically have little understanding or interest in such arcane matters as sampling and measurement error. In times of tight budgets, money for statistical agencies – and particularly for research in method improvement – is reduced without regard to the effect of the reduction on important data series that provide the basis for government decision making or the administration of existing programs.

Validity has two senses. One refers to the accuracy with which the measuring process measures what it sets out to measure. The second refers to the degree to which the measuring process adequately measures the concept that it purports to measure. For example, in assessing the validity of our statistics on poverty, one might focus on the accuracy of income measurement (validity in the first sense) or on the adequacy of the operational definition of poverty, that is, does the method that we use to measure poverty adequately conform to our notions of what we mean by ''poverty?'' The validity of the poverty index in this second sense has recently been the subject of an extensive CNSTAT report (Citro and Michael 1995).

Current controversies about three of the best known and most important measures in official statistics illustrate the two senses. The first example is the undercount in the census. Here the issue is accuracy. There is no controversy about the concept of a ''population'' or what the census is supposed to measure. It has been known for many years that there is an undercount of somewhat less than 2% overall, but considerably higher for some subgroups. While the size of the undercount is reliably known for the total population and for some of the major subgroups, we do not know the magnitude of the undercount for small geographic areas. But it is the small areas that are of interest for many policy purposes. By using statistical methods, such as sampling, the U.S. Census Bureau can make more accurate estimates of the population for the geographical areas of interest. While in previous decades there was considerable disagreement among statisticians and other social scientists about the appropriate methods for improving accuracy, there has recently been a convergence of views and a near, though not yet total, consensus that there are appropriate statistical methods to make the count more accurate. A number of reports from committees of experts have recommended that these methods be applied in U.S.

Census 2000 to produce a population estimate that is based on more than a simple enumeration of the population.

The controversy, of course, is not simply a technical one. The argument is also about the presumed effect of the use of sampling and other statistical methods on the total count in different geographical areas. There are two politically important outcomes of the census that might be affected by new methods. One affects the distribution of political power, that is, the boundaries of election districts in states and the number of seats allocated to different states in the House of Representatives. The other affects the distribution of money, that is, the amount of money distributed to the states, cities, or even smaller units like school districts by the federal government according to formulas based on population characteristics.

When the distribution of power or money is affected by numbers, it is unlikely that decisions about how to collect the numbers will be made purely on technical grounds; rather, political forces will be fully engaged in the decision. What is amazing about the current controversy about sampling in the census is how little analysis seems to have gone toward understanding the probable political consequences of increasing the accuracy of the census counts. Some attention has been given to the question of which states would be likely to gain or lose a seat if the undercount were reduced, although it is far from certain which states would lose and which would gain.

But the consequences that seem to elicit more concern are the effects of using sampling on the distribution of political power between Republicans and Democrats in the House of Representatives. The political effects of an improved census are, however, very difficult to predict. The most talked about putative effect is on the number of Republicans and Democrats in the House. In spite of received wisdom on this issue, it is difficult for me to see how changes in the undercount can, let alone will, have much effect on the balance of parties in the House. In the short run, the important variable determining the consequences of the decennial census on the balance of parties is which party controls the legislature in each state. State legislatures decide the boundaries of House districts, and the party that has a majority in the state legislature during the years when the district boundaries are redrawn is reputed to do a good job of seeing to it that the boundaries are drawn for maximal political advantage. Recent Supreme Court decisions about the gerrymandering of districts to produce predominatly minority districts have given greater flexibility to legislatures to dilute electoral strength based on any characteristic measured in the census. The makeup of the Congress in the first electrion after reapportionment will be far more a function of the distribution of control in the state legislatures than a function of any numerical changes of the magnitude that is apt to occur if the U.S. Census Bureau uses sampling.

In the longer run, the size and character of the undercount will have some effect on the distribution of funds to government or administrative units based on geography. But it is difficult to see that these factors will have much effect on the voting strengths of the different parties. After all, whether you vote or do not vote, or which party you vote for, is not determined by whether you are counted in the census. Improving the population count, unfortunately, will have no effect in increasing voter turnout.

My second example of a statistic that has received a lot of attention recently because of alleged mismeasurement is the Consumer Price Index (CPI). Here the issue is less about

how accurately we measure consumer prices, than about validity in the other sense of the term – that is, whether the CPI is a good measure of the cost-of-living (COL). The fact that the CPI is an index of prices and not of the cost-of-living surprises many people because, in the absence of a COL index, the CPI has been used as if it were a cost-of-living index. People have forgotten, if they ever knew, that the CPI is not meant to be used that way. It is used in many unionized wage agreements to adjust wages in periods between contract negotiations as if it were a COL index. It is similarly used to adjust many government payment levels, such as social security benefits, and to adjust income tax brackets. Because so many payments are tied to the CPI and because it is used by the financial markets as a principal indicator of inflation, it can affect many aspects of the economy, such as interest rates, stock market prices, and the federal deficit.

In spite of the CPI's importance to so much of the economic life of the country, the U.S. Bureau of Labor Statistics (BLS) has had trouble getting users to understand what it really is and what needs to be improved. The BLS has been engaged for years in a program of research that is leading to a revision of the CPI. The Bureau's plans for an orderly program of revision were interrupted by a seemingly casual remark from the Chairman of the Federal Reserve Board during testimony before Congress. He noted that the CPI overstated inflation with the result that the federal deficit was larger than it would be if the CPI properly measured inflation. The Chairman's statement became big news, and politicians began to bring pressure on the BLS to immediately adjust their numbers downward by one percentage point. Much to the Bureau's credit, the Commissioner of Labor Statistics resisted this pressure and refused to make any arbitrary adjustments despite dire warnings from some prominent Congressmen. Congress did appoint a panel of distinguished economists to review the information available from past research. This panel made recommendations for changes in some measurements that go into the CPI. They also noted that the CPI was not a real cost-of-living index and suggested that a real COL index should be produced. Some of these recommendations were in line with what the BLS was working on and will be incorporated into the revised CPI next year.

It had been the practice of the BLS to incorporate methodological and conceptual changes in the CPI during a major revision usually carried out about every ten years. Ironically, the BLS had requested resources a number of years ago to start work on a major revision of the CPI but was turned down. Many of the issues in the current controversy would have been considered as a part of that revision. If the request had been granted when the BLS first made it, the revisions would have been completed by the time the politicians took up the cry for immediate change.

The challenge of designing a real COL index is a daunting one and cannot be done by simply adjusting the CPI arbitrarily because they are fundamentally different concepts and require different measurement methods. Perhaps the difference is now better understood by those who have used the CPI as if it were a COL index, and serious work on a COL index can be funded.

The third example currently in the limelight illustrates the ambiguities of concepts and their measures – that is, the question of measuring race and ethnicity. Race and ethnicity are concepts that have no consensual meaning. They pose the most difficult of measurement problems, that is, trying to devise measures of a concept that are unambiguous in their application but, at the same time, reflect the meaning that most, if not all, people

give them when there is no agreed upon meaning to the concept. U.S. Office of Management and Budget (OMB) Directive 15 states the federal standards for reporting on race and ethnicity, but it does not try to specify the actual wording of the questions used to classify people.

Race and ethnicity are essentially matters of social categorization and self-identity. For obvious reasons, the government does not want to ascribe racial or ethnic identities to individuals on the basis of some objective, but arbitrary criteria. Thus, self-report is the only viable way to get the data. We know, however, that the way questions are asked – for example, the wording of the questions, the number and type of response categories offered, and the order of the questions – can affect respondents' answers. Thus, a common minimum standard for reporting race and ethnicity will not guarantee consistent estimates if the questions are asked in different ways in different surveys.

Recently there has been agitation to allow respondents to give multiple answers to race and ethnicity questions in government surveys and to add a multiracial reporting category to the five categories stipulated by Directive 15. While empirical tests of alternative versions of race and ethnicity questions suggest that only a very small proportion of the population would now classify themselves as ''multiracial,'' some people believe that this proportion will grow in the future. Some of those who now regard themselves as multiracial resent the lack of a category that they can use to identify themselves. Some have suggested that the absence of a multiracial category supports racism and the perpetuation of race as an important social category.

An interagency committee to review Directive 15 has recommended that ''multiracial'' should not be included in a list given to respondents in questions to elicit racial identity, but that instead, respondents should be allowed multiple responses to a racial identity question. This recommendation has the merit of preserving detailed information about the respondents while still allowing them to indicate a multiracial identity. Of course, it has the disadvantage that it makes the problem of reporting the data complex because one must devise reporting rules for different combinations of racial and ethnic categories.

One must tread a delicate line here between the Scylla of using response categories that do not match respondents' own frame of reference, particularly in such emotionally and politically charged questions as those of race and ethnicity, and the Charybdis of a system that does not allow data users to answer the questions for which the data are collected. If ''multiracial'' is an analytic category that is important for social scientists and policy analysts to use in answering scientific and policy questions, then it would be a severe loss to make it impossible for them to understand what is going on in society. My own recommendation is to use ''multiracial'' as a reporting category, but not as a measurement category, that is, when summary measures are to be used, to report people who select more than one racial category as ''multiracial.'' We live in a society that is becoming slowly but surely, increasingly multiracial, and our official statistics should not only recognize that fact, but also chart the movement in that direction. To do otherwise is to undermine the validity of the measure.

In this section I have tried to point out some of the challenges that face the statistical system regarding the validity of its data, both from the point of view of accuracy and from the more difficult point of view of conceptual adequacy. I now turn to the question of timeliness.

2.3. Timeliness

Timeliness refers not only to the gap between the time of data collection and their availability but also to the periodicity of data collection operations and to the revisions in the measures to reflect changes in society that affect validity. Technology may provide means for decreasing the amount of time necessary to collect and process data, but technology may not be able to decrease the time sufficiently to satisfy an audience that has almost instant access to whatever is available.

If my prediction is correct that the availability of official statistics on the World Wide Web will greatly increase their use for policy and advocacy purposes, then, inevitably, the demand for more timely statistics will also increase. When accessing data takes a long time and they come out in hard copy, people are willing to believe that it takes a long time to produce them. But when we can get access to them almost instantaneously, we are impatient for more up-to-date points in a data series or more recent statistics on most topics. When I have to rely on looking up numbers in a book, I am willing to accept the fact that in 1997 I can get data on carbon dioxide emissions only through 1993. When I can access data on the computer in one minute, I find it intolerable that more current data on emissions are not there.

The advent of computer-assisted data collection goes some way toward reducing the elapsed time between data collection and data dissemination, but not by as much as some had hoped. The increased use of computers in every aspect of data collection, processing and dissemination has not produced major reductions in the cost of compiling statistics, although it has reduced the time for data processing somewhat. At the same time, however, computer-assisted data collection has increased our ability to collect data more accurately by increasing the complexity of the data collection process. It has also tempted agencies into creating more sophisticated and larger data bases in order to improve the statistical output. Both of these trends work against making significant progress in timeliness.

Technology, moreover, has little to contribute toward the decision of agencies to change measures to reflect social changes that underlie the construction of the measure. While all agencies periodically revise their measures that are constituents of time series, the frequency of revision is a matter of judgment, as well as budget. In 1994, the U.S. Bureau of Labor Statistics made a major revision in the way it asked the employment status questions on the Current Population Survey, namely changing the operational definition of employment and unemployment. This was the first major revision in the questionnaire in 27 years. As I mentioned earlier, the data collection methods for the Consumer Price Index are modified about every ten years, and it will be revised beginning next year. OMB Directive 15 was issued 20 years ago, and its revision should be out soon. There is no magic formula to specify when a definition or a question should be changed. For important measures, change requires considerable research into the effects of the change on times series or other historical statistics and, where appropriate, time series must be readjusted to be consistent with the new measures. As Starr (1987, p. 50) has pointed out:

‘‘Although statistical systems need continually to be adjusted to take account of such (structural) changes, they have powerful inertial forces at work. The more categories and questions are revised, the less comparable are data for different periods. Structural

revisions are also costly to undertake. Hence there are strong conservative tendencies that may lead to a lag in the recognition of structural change.''

Again, as data become more available and used in policy debates by advocates of different policies, measures will come under closer scrutiny more often, and pressure for bringing measures into line with changes in social and economic conditions will increase.

3. The Organization of the Federal Statistical System

I now turn to the institutional structure necessary to respond to these challenges. How well organized are we to meet these challenges? As this audience well knows, the United States has a decentralized federal statistical system. Several years ago in his Morris Hansen lecture, Ivan Fellegi reviewed the strengths and weaknesses of centralized and decentralized systems and made the case for a centralized system. While I would not argue that the United States should move entirely to a centralized system, our current system is not well-suited to meet the challenges of relevance, validity and timeliness that I have argued constitute the challenges of the near future.

Our present system can be strong in the areas where there are clear functional interests and responsibilities, assuming that there is adequate budgetary support. We have statistical agencies with specific responsibilities for areas such as health, education, transportation, justice, agriculture, science, and the economy. The system as a whole is coordinated by an Office of Statistical Policy in the Office of Management and Budget headed by the Chief Statistician. The staff resources available to the Chief Statistician, however, are woefully insufficient to carry out the heavy responsibilities of the office. Thus, the Office has to rely on Interagency Committees as a principal mechanism to study in depth particular problems such as revising the definition of a Metropolitan Statistical Area or the way in which data on race and ethnicity are to be collected and reported.

These Interagency Committees do excellent work, but they are ad hoc and by their very nature limited to a particular topic. Since they have no continuing existence, except for one committee on statistical methodology, they are not well placed to think about broader issues confronting the entire federal system or to anticipate changes of the sort I have been talking about today. There needs to be some forum where challenges such as I have discussed can be addressed on an ongoing basis and responded to before they become politicized.

In her wise book, *Organizing to Count* (1995), Janet Norwood argues for combining the U.S. Bureau of the Census, the U.S. Bureau of Labor Statistics, and the U.S. Bureau of Economic Analysis, and for lodging the statistical coordination function in a Central Statistical Board. Such a combination would go far toward giving us a system that, at least for a broad range of economic statistics, would be capable of responding to the challenges of the information society.

There is, however, an important hole that would still be left if such an entity were formed, for it is a hole that exists in the present system. A wide variety of social and demographic statistics are now collected by a number of agencies – the U.S. Bureau of the Census, the National Institute on Aging, the National Center for Child Health and Human Development, the National Science Foundation, the U.S. Bureau of Labor Statistics, and the National Centers for Health and Education Statistics. But no agency that has

program responsibility and relies on social and demographic statistics to plan, implement, and evaluate its programs also has responsibility for a coordinated system of social and demographic statistics. A crucial ingredient of any reorganization of the federal statistical system would be a National Center for Social and Demographic Statistics, or at least its functional equivalent. This responsibility could be exercised by a Central Statistical Board if it were given the mandate and necessary authority. If this responsibility cannot be accommodated in an entity such as Janet Norwood has suggested, then I would suggest a new statistical agency lodged in the Department of Health and Human Services, whose programs and policy interests more closely span the areas in which statistics are now the weakest. This could be accomplished by expanding the mandate of NCHS to more closely reflect the mandate of DHHS or, if health statistics are judged to be a specialized topic requiring its own agency, then by establishing an entirely new statistical agency. Such a solution, however, would likely be less effective than a more central Board charged with overseeing an integrated statistical system.

It is not obvious to most policy makers that the organization of statistical agencies has a profound impact on the policy-making process, because the pervasive effect of statistics is not visible to most people, even those who use them in policy deliberations. But statistical agencies do play an extremely important role in government in spite of the fact that the vital importance of ''the numbers'' is not widely recognized except at exceptional times when someone calls attention to them in dramatic ways. I have argued that the role of statistical agencies and their products will become more visible in the future as statistics become more accessible to a wider audience and that the system of official statistics will come under increasing scrutiny as a consequence. If the system is to respond constructively to the changes in the demand for information, then it must be organized in a way that can cope with these demands, and not just retreat into a defensive posture. Perhaps the greatest challenge to the federal statistical system is how to organize itself to meet the changes brought about by the technological revolution in data accessibility. It will take enlightened leadership to bring about changes, but the current widespread concern for reinventing government offers a favorable time to accomplish change.

3. References

Anderson, M. (1988). The American Census: A Social History. New Haven, Conn.: Yale University Press.

Citro, C. and Hanushek, E. (1992). Micro Simulation Modeling. Washington, D.C.: National Academy Press.

Citro, C. and Michael, R. (1995). Measuring Poverty: A New Approach. Washington, D.C.: National Academy Press.

Coleman, J., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., and York, R.L. (1996). Equality of Educational Opportunity. Washington, D.C.: U.S. Department of Health, Education and Welfare, U.S. Government Printing Office.

Perlman, M. (1987). Political Purpose and the National Accounts. In The Politics of Numbers, eds. W. Alonso and P. Starr. New York: Russell Sage Foundation.

Porter, T.M. (1986). The Rise of Statistical Thinking: 1820–1900. Princeton, N.J.: Princeton University Press.

Starr, P. (1987). The Sociology of Official Statistics. In The Politics of Numbers, eds. W. Alonso and P. Starr. New York: Russell Sage Foundation.

Stigler, S.M. (1986). The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, MA: Belknap Press of Harvard University Press.

Received January 1999