UNITED
NATIONS

E

Economic and Social
Council

Distr.
GENERAL

CES/AC.71/1999/16
19 November 1998

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Meeting on the Management of Statistical Information Technology
(Geneva, Switzerland, 15-17 February 1999)

Topic (i): The impact of Internet on the statistical production and dissemination process

THE INTERNET AS A TOOL FOR CARRYING OUT THE CENSUS BUREAU'S MISSION

Submitted by United States Bureau of the Census[1]

I. OVERVIEW

1. The Census Bureau's mission statement says that “we will strive to be the preeminent collector and provider of timely, relevant, and quality data about the people and economy of the United States”.

2. We have adopted several strategies to help us carry out our mission successfully, including improving Customer Satisfaction and Public Cooperation. The first relates to how the public perceives the Bureau and its services, especially in terms of the quality, price and ease of access to data products. This strategy causes us to examine our data dissemination policies and methods. The second strategy--Public Cooperation--also relates to how the public perceives the Census Bureau, including how easy it is to participate in Bureau censuses and surveys, and how well the Bureau protects the confidentiality of individual responses. This strategy causes us to examine our data collection and security methods, as well as our dissemination plans.

[1] Prepared by Bradford R. Huther.

3. Using information technology to further Bureau goals is not new. The Census Bureau has long been a leader in information technology--not only in the processing of information, but in its dissemination, as well. Clearly, the Internet can be an integral part of the data collection and data dissemination methods we adopt, refine, and expand to increase customer satisfaction and public cooperation. In fact, the Internet has already become a central part of the Bureau's information dissemination practice. Using the Internet for data collection began a few years later, but we are making many strides to increase our understanding of the best way to use the Internet in this realm. Using the Internet to conduct these core business practices will only grow, as some of our ground-breaking initiatives attest.

II. DATA DISSEMINATION

Customer Service

4. Customer service has many components, such as ensuring that the data collected in censuses and surveys include the types of information needed by policymakers, state and local governments, academia and the private sector. The Bureau seeks customer advice and reaction to the types of information we collect and products we produce in several ways. These include conducting regular meetings with standing advisory committees and establishing a Marketing Services Office that profiles user requests and surveys users regarding product usefulness. This office also maintains the Customer Service Center, which responds to about 1,500 telephone and electronic mail requests for information each month.

5. Another major component of customer service is how the information is made available. Guidance on the dissemination of agency data products is found in long-standing Executive Branch policy, referred to as Office of Management and Budget Circular A-130. The policy states that "government information is a valuable national resource," and that "the free flow of information between the government and the public is essential to a democratic society." It therefore encourages data dissemination for no more than the cost of dissemination alone.

6. Census Bureau policies embody this practice. Dr. Barbara Bryant, a former Census Bureau Director, said it well: "Our data products are the return to the American people for the large investment they have made in the decennial census and other programs." The Bureau charges a modest fee for printed, diskette, tape, and CD-ROM products requested directly from the Bureau. However, our history also includes making data widely accessible at no cost by giving free copies of materials to a national network of libraries, data centers and Census Information Centers.

Current Census Bureau Internet Site

7. Today, the Internet allows us to continue this tradition, but in a more accessible and convenient manner. In fact, the National Performance Review,

an initiative of the Vice President, urges all federal agencies to increase efforts to deliver information and services to the public via electronic communication. The Bureau took that guidance to heart three years ago, when it announced plans to expand electronic dissemination of its data by using the Internet as a primary source for data release.

8. The Census Bureau's Internet site provides highlights of our latest censuses and surveys, detailed tables, and links to other sources of data, such as other U.S. Federal agencies, our State Data Center partners, and 86 International statistical organizations. Users also can order "traditional" products via the Internet. Our Internet site currently contains almost 61,000 distinct "pages" of information, including electronic versions of all printed reports published since January 1996. It currently receives almost 5 million "hits" (i.e., the number of times a page is accessed) a week. We are redesigning our Internet site to make information easier to find and are implementing standards to create a more consistent look, feel, and functionality. Information provided on our Internet site is available free of charge to any interested data user.

9. Two years ago, the Census Bureau created CenStats, a subscription service designed to complement the information already available on our Internet site. CenStats provides Internet access to products we now sell on CD-ROM and other specialized databases. For example, while 2-digit level commodity detail associated with our Foreign Trade Database is available on our public Internet site, CenStats users can access the 5-digit level of commodity codes.

10. Another early Bureau project already on-line is the Federal Electronic Research and Review Extraction Tool. FERRET is a computer search tool developed by the Census Bureau and its sister agency the Bureau of Labor Statistics. It is designed to assist data users at many levels of sophistication in accessing the results of two major Bureau surveys--the Current Population Survey and the Survey of Income and Program Participation.

11. We believe that our efforts at listening to our customers is paying off, both in terms of the growing list of awards our Internet site has received, but more importantly, the increasing number of users of our Internet services. First, the site has received several federal awards, and was named once again as one of the top 100 sites by PC Magazine in 1998. Second, surveys of customer preferences suggest that many traditional census data users are complementing their use of traditional products with use of our Internet services. An increasing number are expressing a preference for on-line dissemination. Third, a growing number of new users are accessing our Internet site every year (this includes a fairly steady rate of increase among international users, who comprise about four percent of all users).

The American FactFinder

12. To serve our customers even better, the building of our "next

generation" of Internet services is well underway. This next generation system was originally conceived in 1995 as the Data Access and Dissemination System (DADS). When fully implemented, the system will provide more cost effective and wider dissemination, greater and more timely data access, and a simpler user interface that will include "point and click" commands and thematic maps. DADS will be cross-cutting and comprehensive in scope, and will change the way the Census Bureau tabulates, publishes, and disseminates all demographic and economic data.

13. The first public release of this new system, called the American FactFinder, will be in the first quarter of 1999. American FactFinder will allow internal Census Bureau employees and external data users to formulate queries against many different databases, and retrieve "just the information they want," subject to the data on the files in the system. The American FactFinder will include information from the 1990 Census of Population and Housing, American Community Survey, 1997 Economic Census, and eventually Census 2000.

14. Information will be made available in three "tiers" of information, with each providing more detailed information than the last and each supported by increasingly sophisticated technology. These tiers will include predefined data that users will be able to print, order, or download in electronic format. Users will also be able to access the data bases and formulate customized queries. These queries will be available to both summary and detailed-level data.

15. The system is heavily dependent on a comprehensive Census Bureau-wide "metadata"--"information about the information"--architecture. It will help a diverse set of users--from school children, to social science researchers, to business analysts--quickly find the information they are seeking. Users will be able to select geographic locations by pointing on a map, and visualize information graphically through "thematic maps."

16. As the American FactFinder evolves, the Bureau must continue addressing several important issues. The first is the risk of data disclosure. The Census Bureau is obligated under the law and through long-standing professional ethics to protect the confidentiality of every respondent. This takes not only the form of keeping responses private, and applying security measures, but also applying disclosure avoidance techniques to avoid identifying an individual through unique characteristics, even without using a name. This issue is very relevant to our FactFinder system. Access to detailed-level information will be subject to extensive checks through implementation of strict confidentiality filters. These filters will examine user queries to the microdata (detailed-level access) and the tabulated results that are returned, and will ensure that the confidentiality of individual user data is protected, as required by law.

17. Perhaps the most difficult aspect of the disclosure issue is perception. We are going to great lengths to guard against any disclosure. However, making user-defined tabulations from a microdata file available could be

perceived as a breach of confidentiality. We are working on ways to avoid this difficulty including a public airing of the data dissemination plans and disclosure protections in advance of census 2000.

18. A second issue of concern with increased access via the Internet is potential decreased access to census data by those without computers or those who simply prefer paper copies. Libraries in the United States are also getting more and more computerized. But for many libraries, the number of work stations is limited, making it difficult for them to serve all their customers as more and more information from all sources is put on the Internet. By early in the next century, we expect that almost all data users will have some kind of Internet access. Even so, we are working with our State Data Center program to make sure that those without direct Internet access have an opportunity to use census data.

19. A third issue is the scalability of the system to respond to the anticipated large volume of users. The Census Bureau has tried to balance the cost of building a system to meet the needs of many users with the level of service needed to meet the need of some, more sophisticated, users. One way we have addressed this concern is through the use of "tiers" of information. The three tiers have been implemented like an inverted pyramid. We anticipate that many of the users will be able to get the information they are looking for from the lower tiers of the pyramid--tiers 1 and 2. This will limit the number of complex, custom queries that may be made against tier 3 microdata and should help maintain a consistent level of performance.

20. A fourth issues concerns cost. Current federal policy on electronic dissemination of data allows for dissemination cost recovery, but does not allow for developmental costs. Therefore, investments in American FactFinder must be funded by the agency's appropriated funds. However, we anticipate that there will be fees for some transactions, such as tier 3 queries, within that system once operational.

21. The last issue is the need to complete a technical and business strategy for implementation of the original DADS vision and future expansion of American FactFinder. To accomplish this, the Census Bureau implemented a cross-functional organizational team in the summer of 1998 to formulate a road map for full implementation. This group is the Integrated Information System Redesign Project and it will deliver a final set of recommendations to Census Bureau management in the first half of 1999.

III. DATA COLLECTION

Public Cooperation

22. Public cooperation is an essential strategy for achieving our mission, since voluntary response is the cornerstone of high quality results. Even for censuses and surveys identified as mandatory, refusal is possible, and if not for the persuasiveness of field representatives, other census personnel, and

census materials, those refusals will often result in loss of data. The Census Bureau, as well as other data collection organizations, has faced an environment of declining response rates to all types of surveys--whether via the mail, telephone or in person. A good deal of effort is expended identifying methods just to maintain our current response rates. For example, a comprehensive mailing strategy was tested and adopted for Census 2000 that, if all components are implemented, should increase response rates by about 10 percentage points over that achieved by using a more traditional mailing strategy. However, this gain is likely to just offset the loss due to decreasing cooperation since the 1990 census.

23. The Internet provides an opportunity to enhance public cooperation in data collection by providing an additional response mode, which some users may find more convenient. Secondarily, it also offers opportunities for improving data quality and increasingly the efficiency of data processing. Bureau experience with the Internet for data collection is somewhat limited, but a number of initiatives currently underway suggest that the Internet could be a major data collection instrument within a few short years. Our challenge is to learn from these targeted initiatives to ensure successful implementation of this technology.

Electronic Data Collection for Surveys of Businesses

24. In the recent past, economic surveys were considered better candidates for applying some technologies because firms are more likely than households to have personal computers, and other automated equipment. The Census Bureau's approach to applying technology in the economic census and survey area was guided by a desire to use technology as a means of reducing reporting burden, while also improving data quality, and census Bureau processing efficiency.

25. Although the Census Bureau's use of the Internet for data collection is currently limited, we offer a wide variety of electronic reporting modes for many of its Economic censuses and surveys. These include Computer Assisted Telephone Interviewing, Fax, and Electronic Data Interchange. A newer, promising technology is Computerized Self-Administered Questionnaires, called CSAQ (pronounced see sack). Diskette CSAQ via the mail is being used in several ongoing surveys, and companies given the option are increasingly selecting that response option. In addition, we currently receive a limited number of responses via the Internet from companies participating in the Automated Export Reporting Program, developed in conjunction with the U.S. Customs Service. This information is used in our foreign trade statistics program.

26. In order to explore moving to an Internet application of CSAQ, respondents in a pilot program associated with a 1996 Industrial Research and Development Survey were asked, via a mailed screener questionnaire, about their interest in using the Internet for reporting, as well as their Internet browser capability. Based on their responses, 50 companies were selected. We subsequently provided them with a Web-based CSAQ as their response vehicle for

that R&D survey.

27. This experience taught us several important lessons about how to pursue further Internet-based data collection efforts. First, the ideal software remains elusive. Outside experts have recommended that we use commercial off-the-shelf software to develop CSAQs and other Computer Assisted Interviewing technology. None were available that met our requirements. So the code we selected was (HTML/Java Script). It provided the benefits of real-time editing and a screen-based design. However, this combination with the necessary security measures eliminated many potential candidates from our participant pool.

28. Second, with Internet questionnaires being so new, Usability testing in our cognitive laboratory needed to be done to determine the best screen design and navigational method. A screen-based design using Next and Previous buttons for navigation allowed us to implement a reasonable questionnaire design. However, we learned that seemingly simple items, like the respondent misplacing the letter containing their username and password, can significantly increase the amount of assistance required to complete an Internet-based questionnaire.

29. Third, Internet response may be motivated by factors quite different from a more traditional response mode. The amount of follow-up required to get the 34 (out of 50) completed responses was disheartening, particularly given that the sample of 50 were self-selected. Related, the types of assistance required to facilitate a successful response may be quite different.

30. These findings have caused us to investigate several uses of the Internet as a response option for companies and other organizations. We are currently conducting research to quantify business concern regarding Internet reporting and also to ensure that we have not overlooked any special requirements or features. We also are in the midst of developing an Internet application for another survey, the Company Organization Survey. Finally, we have also developed an Internet application for a demographic survey, the Library Media Center Survey. This was released to over 900 public and private school respondents in late October. For these surveys and future economic and demographic surveys, a substantial security system to protect the confidentiality of the data collected via the Internet has been developed.

31. Ultimately, the use of an Internet questionnaire, as well as the design needed to support it, may be quite different between an economic and a demographic survey. An economic survey often requires historical data transmitted to the respondent; a lengthy, intermittent survey completion time; and relatively little concern about nonresponse bias due to access to a computer. A demographic survey rarely requires historical files or more than one session to complete, but may create significant coverage concerns due to access to a computer with an up-to-date browser. The coverage concerns would be overcome by offering the Internet questionnaire as an optional method of reporting with paper as a backup.

32. While we definitely see the Internet as a tool to reduce respondent burden, and to improve data quality and efficiency, we have a number of things to learn before we using it broadly across many of our censuses and surveys.

Electronic Data Collection for Census 2000

33. The Census Bureau is developing a means for the public to respond to the 2000 decennial census via the Internet. Currently, our plans include allowing any household receiving a short form (mailed to about five of every six housing units) to respond by accessing a specially designed site, entering the identification number associated with the questionnaire they were mailed, and completing an "HTML" short form questionnaire.

34. We believe that the key to successful implementation of this massive, one-time application is to ensure simplicity of use without non-essential "flashiness." We plan to make the Internet address easily available to respondents, either on the questionnaire itself or on a mailing insert. Then, we will make the questionnaire on-screen look as similar as possible to the one mailed to the respondent's home. By using only simple HTML-Form and avoiding features that may cause problems across different web browsers, like Java or Java Script, we hope to make filling out the Internet form and returning it a simple task for respondents.

35. On the other hand, we are taking some steps that could be said to limit the ease of response, due to concerns about keeping information secure, and matching and "unduplicating" multiple responses from the same housing unit.

36. First, to ensure that we protect respondent confidentiality effectively, we are using a "double encryption" approach. The completed questionnaire file will be sent back to the bureau in a secure, encrypted form (using HTTPS, 128-bit encryption). Then, once received by the Bureau's server, it will be encrypted again prior to storage on a disk. Unfortunately, the initial encryption measure limits those who can access our site by requiring a relatively new web browser; however, this trade-off appears necessary to maintain the high level of security we are committed to maintaining.

37. Second, to ensure that we can accurately link a respondent to an address, and an address to a piece of census geography, each questionnaire mailed to a U.S. household will contain an identification number. Responses without such a number may require extensive follow-up efforts, which could delay timely completion of census activities. They could also introduce error into the counts if erroneous assignments are made. Therefore, respondents will not be able to provide responses via the Internet without first providing the identification number from the questionnaire mailed to their home.

38. Our past research gives us reason to believe that the Internet will not be a primary response mode for those residents we generally consider hard-to-count. Therefore, we do not expect to have a negative affect on census coverage by virtue of this decision, since a respondent attempting

unsuccessfully to reply via the Internet is almost certain to be counted in another census operation.

39. Once received by the Bureau, an extract of the information (i.e., all but some detailed information such as time of receipt) will be formatted into a record as similar as possible to those being generated by the Data Capture System 2000. This system will capture, via scanning technology, paper-based responses completed through traditional census techniques. From that point on, Internet responses become part of the main census processing stream.

40. Open issues with this Internet application include whether a requirement to offer response options for the long form (sent to the remaining one in six households) will be levied. If so, both the short and long form response options are likely to be offered via a "down-loadable, executable" file structure, such as that being considered in the economic survey area. A series of security issues not applicable to the HTML environment are introduced here. These include concerns about where the respondent stores the file, since the Bureau cannot control access to those locations.

41. Another open question is the level of public response to this response option. Our research to date has focused primarily on the response attractiveness of alternatives such as the telephone, or a paper questionnaire in a language other than English. Given the growing number of individuals with access to the Internet, we could see a ground-swell of interest in this response option by the year 2000.

IV. CONCLUDING THOUGHTS

42. Wide-accessibility to the Internet presents us with an enormous opportunity to re-engineer our business processes in order to better serve the public. Early Internet initiatives at the Census Bureau tended to be "grass roots" rather than centrally designed. Our DADS data dissemination project is our first major opportunity to significantly redesigned data dissemination in light of the Internet. Data collection is clearly our next opportunity to use our experiences to create a means of interacting with the American public that goes beyond an "add-on" to the current panoply of data collection modes for any given census or survey. While our economic survey area has led the way in exploring Internet use for interacting with business, our Census 2000 experience provides our first opportunity to reach the public directly with an Internet-based Census 2000 questionnaire. Those lessons learned are sure to give direction to our efforts in the first years of the coming millennium.