Session 2

Achieving Timeliness in a "Real Time" World

Achieving Timeliness in a "Real Time" World
Charles Louis Kincannon, Director,
U.S. Census Bureau

Introductory Remarks

Data collectors, data disseminators, and data users contributed unique observations regarding the timeliness of data from the Federal statistical agencies. In particular, representatives from the data user community provided useful perspective of stakeholders' needs and uses for economic and demographic data. The obvious tension or inconsistency between the expectations of timeliness in a "real time" world and the requirements for accuracy and relevancy was immediately apparent. The "cost" to achieve the government standards of accuracy it appears to many, is often, timeliness.

Whether this is a tension or an inconsistency should be explored further. In many cases, it is an inconsistency, this means that Federal statistical agencies must do a better job with stakeholders to illuminate the entire data delivery process—from data collection to data use. However, there is often a tension between the requirements of data quality, or accuracy, and timeliness. If this tension can be resolved or overcome, data collectors, data disseminators, and data users may have to engage in a discussion of priorities. There is the possibility that the most important uses for some data may rank timeliness above restrictive measures of accuracy. Or, perhaps more likely, priorities established for processing certain data products with regard to others may need to be reassessed.

Achieving Timeliness in Real Time
John Kavaliunas, Chief, Marketing Services Office
U.S. Census Bureau

Abstract

The Census Bureau has made great strides in speeding up the process of releasing information. Not only have there been improvements in release dates, but new technology has enabled us to get the data into the hands of end users much more readily than ever before. What is on the horizon that will enable us to continue to meet rising user expectations?

Early Censuses

The results of the first census in 1790 were released as soon as the enumeration was completed, posted by the U.S. Marshals on tavern walls and in other public places. The results were sent to the President who sent a tabular statement of the results to Congress in late October, 1791. Of course there were only 3.9 million persons in the country at that time and the information collected was quite limited.

Throughout the 19th Century, additional questions were added to the census questionnaire and the size and the number of printed reports increased.

It took time to collect, tabulate, analyze and publish the increasing number of census reports. Timeliness of the data became a concern. General Francis A. Walker, Superintendent of the 1870 Census, wrote:

> So rapid are the internal changes of the country, oftentimes setting calculations at naught, so fierce and vast the growth of the Nation as a whole, that the hiatus in the statistical information at the command of the legislator, the pamphleteer, the journalist, and the social and political philosopher, becomes positively painful five or six years after the day of the census. (**Report of the Superintendent of the Ninth Census**, November 1, 1872.)

Things would get worse before they got better. By 1880, the number of census reports had risen to 22. This is a noteworthy number because it was more than all the reports printed from previous censuses combined. The first reports from the 1880 Census didn't appear till three years after the census was taken and the last reports weren't printed until 1888, prompting the Assistant Director of the Census Bureau, Frederick H. Wines, to write in 1899:

> Speed is to be greatly desired. Former censuses have required as much as nine years to complete the publication of their work, and their statistics have been to a certain extent out of date when they appeared. ("The Census of 1900," **Munsey's Magazine**, 1899).

The 20th Century

At the close of the 19th century, the introduction of new technology in the form of punch cards and electrical tabulating machines speeded up considerably the processing of decennial census results. Printed reports continued to be the mainstay of the dissemination process. However, the number of questions on the questionnaire continued to grow, as did the levels of geography for which the data were tabulated and, subsequently, the number of printed reports.

Census tracts, metropolitan districts, and the terms urban and rural first appeared in 1910.

Blocks were added in 1940; minor civil divisions and census county divisions became part of the standard geography in 1950, and data were first tabulated for block groups in 1970.

There were 1,003 individual reports in 10 different series published from the 1960 census, including the 421 reports in the report series, HC3 (*City Blocks*), which included data by block for all cities above 50,000 inhabitants, and for some 200 smaller places that had contracted for block statistics.

In the late 1960s, the Census Bureau experimented with releasing information on computer tapes and, by 1970, computer tape was a standard dissemination medium. While computer tapes speeded up the release process and much more data could be provided on tape than in a printed report, only large organizations like university research units, government agencies, and private companies could read the tapes and process the data. The public had to wait to hear about the data release and then find an organization that had the information.

The Democratization of Data

The introduction in the mid 1980s of personal computers and the adoption in 1985 of a new technology known as CD-ROM, made Census data even more accessible to the public. The widespread use of CD-ROMs to deliver 1990 Census data brought about, what then-director, Barbara Bryant, called the democratization of data.

But the revolution in information dissemination was just beginning. The real democratization of data didn't really occur until the introduction of the Internet, just a few years later. The Census Bureau launched its Internet site in May 1994. A little more than a year later, the Census Bureau announced that the Internet would be its primary means of data dissemination:

> The new dissemination plan will allow for quicker release of detailed data many people want. In the past, issuing tables and analyses in printed reports could add months to the process. And since we could only print a selection, users still might not get the data they wanted. A major advantage of this initiative is that it will allow users to receive data files on demand and to create their own reports rapidly... (CENSUS BUREAU EXPANDS ELECTRONIC DATA DISSEMINATION, Press Release dated August 9, 1995)

During the 1990s, the Census Bureau had already begun cutting back on the number of printed reports as well as the number of pages in the reports. In lieu of the traditional 200-300 page

reports, the Census Bureau began publishing short ***Briefs***, which summarized findings and included analysis, graphics and maps, but with a limited number of statistical tables. These tables were put on the Internet instead of in a report appendix. Another development was the use of Adobe Acrobat to convert reports into portable document format or *pdf*, enabling us to create web-based documents without having to go through the often lengthy printing process, which could add months to the release of the information. The planned number of printed pages from Census 2000 is about one-tenth of the 1990 census output, down from 400,000 pages to about 40,000.

Census 2000

On the one hand, the Census Bureau's release of Census 2000 information was somewhat comparable in timing to 1990. However, if we look at when the information was actually in the hands of the public, then the release of Census 2000 data is far and away the fastest ever. Technological advances such as the internet, the American FactFinder, File Transfer Protocol, and our ability to produce custom CD-ROMs enabled us to get the information to many more end users much faster than in 1990.

The number of end users of Census 2000 information is something we could not have imagined in 1992. Almost five million users visited the Census Bureau's Internet site during the month of October 2002 which, in terms of data releases, was an uneventful month. Compare this to the quarter of a million users who called or wrote the various Census Bureau call centers and regional offices in all of 1992.

Let's compare the timing of several key data releases.

Release of the Public Law Redistricting Data is mandated to be completed by April 1 of the year following the Census. While we met that deadline during both decades, it should be remembered that the Census 2000 version of the file, due to the multiple race tabulations and additional geography, is about 10 times greater than its 1990 counterpart.

Summary Tape File 1A was released on computer tape between April and August 1991, and on CD-ROM several months later (October- November 1991). Veterans of the 1990 Census data user community will remember that there were suffixes appended to the file names to indicate the geographic summary levels at which the data were provided. Summary Tape File 1A contained data down to the block-group level. Summary File 1B contained data for all 7 million blocks in the U.S. at that time and was released on computer tape in the fall of 1991. An extract version of this file on CD-ROM was finally made available in 1992, although due to recalls and other factors, data for some regions were not officially released until November of 1993 . The Census 2000 SF 1, containing data for blocks and block groups (i.e., no suffixed files) was released between June and August of 2001 with all states available on a single DVD in September 2001, some two months to 2 years earlier than its 1990 census counterparts.

Summary Tape File 3A, the first release of 1990 sample information, came out on computer tape in April and May of 1992; but the comparable CD-ROMs, all 61 of them, were not produced until that winter, with some not released until February 1993. The long-awaited 3B or ZIP Code

file was made available on tape or CD-ROM between April and June of 1993. By comparison, we released the entire SF 3 File in September of this year. The DVD is expected in late November, a half a year earlier than the comparable 1990 product. It should also be remembered that the Census 2000 version of this file is about 5 times larger than its 1990 counterpart (16,530 cells vs. 3,300).

The large 1990 Summary Tape Files 2 and 4 with detailed race, ethnic, and ancestry data, because of their complexity and size, were not produced on CD-ROM and were therefore only available to only a small number of state data centers and other groups that had the capacity to process these multi-reel files. In 2000, both files are much more accessible to end users, available to the public on the Internet through the American Factfinder, the file transfer protocol, and on CD-ROM and DVD.

Release Dates for Key Decennial Products		
Product	**1990**	**2000**
Redistricting Data	2/91-3/91	3/01
S(T)F 1	4/91-11/93	6/01-9/01
S(T)F 2	10/91-11/91	12/01-4/02
S(T)F 3	3/92-6/93	8/02-11/02
S(T)F 4	3/93-12/93	4/03-9/03(Planned)

Rising User Expectations

In a 1994 survey of users of Summary Tape Files 1 and 3 about three-fourths of respondents agreed with the statement that STF 1 on computer tape was available in a timely manner. However only 50 percent thought the STF 3 tape and CD-ROM products were available in a timely manner and more than one-third (36 percent) of CD-ROM users disagreed with that statement.

"Need to get the data out sooner. Business does not like to work with 3-year old data," wrote one survey respondent. "The quality of the product is excellent, but please try to work on release dates and delays," commented another.

So, how have users reacted to the timeliness of Census 2000 data? In a series of 12 focus groups with key customer segments in the winter of 2001-02 (that is, prior to the release of sample data), most participants said that the timing of 2000 release actually exceeded their expectations.

Nevertheless, they also said that while the release of data was considerably faster than in 1990, they wished the products could be released even sooner!

In Summary

Technological advances have shortened the time necessary for processing, but governments and society have demanded more data, more complex tabulations, and additional levels of geography. In 2000 the Census Bureau improved upon 1990 census release dates from several weeks to as much as a year or more. But perhaps more importantly, technological advances have put this information into the hands of more people than ever before.

With the timeliness of Census 2000 data releases exceeding public expectations, how do we meet the challenge of rising user expectations?

What will be the next technology to appear on the horizon? For the past two decades we've been able to quickly adapt new emerging technologies (CD-ROM and the Internet) to data dissemination. Will the Internet be even more pervasive in 2010 or should we look for still another technological advance?

Will the collection of data via the Internet improve the timely processing of information? Will real-time data collection result in real-time tabulation and dissemination?

And what about the re-engineered 2010 Census? Will annual data from the American Community Survey, available six to seven months after the end of the collection year and the collection and processing of responses to only a few basic short-form questions, make timeliness somewhat of a non-issue?

But these are topics for another presentation.