

INTEGRATING CATI CENTERS IN A DECENTRALIZED DATA COLLECTION ENVIRONMENT

**Kathleen R. Donohue, Richard L. Clayton, and
George S. Werking**

Richard L. Clayton, U.S. Bureau of Labor Statistics
2 Massachusetts Avenue N.E., Washington, D.C. 20212

I. INTRODUCTION:

Over the past decade, the Bureau of Labor Statistics has conducted extensive research and development in to automated data collection techniques, CATI the Current Employment Statistics (CES) survey. Concern over the quality of the estimates produced by the CES survey led to a rapid implementation of automated collection methods, particularly for large sample units which habitually reported too late for use in preliminary estimates. The accelerated implementation schedule led the BLS to establish two Data Collection Centers (DCC), in the Atlanta and Kansas City.

This paper addresses issues encountered in the establishment of the Data Collection Centers, including organizational structure, quality and efficiency of the sample conversion and data collection process, systems design, and future initiatives.

The Current Employment Statistics Survey:

The CES survey produces monthly estimates of employment, hours and earnings by industry for the nation, states, and geographical areas. Each month, survey forms are mailed to approximately 360,000 sample units. Currently, estimates for any given reference month are published three times; a preliminary estimate is released the first Friday of each month, concurrently with the results of the Current Population Survey. Revised estimates are published four weeks later, and final estimates in three more weeks. The preliminary estimate generates the most interest, but is based on only 50 to 55 percent of the sample.

The Federal/State Cooperative Relationship:

The CES survey is one of a number of Federal/State cooperative programs conducted by the BLS. In this arrangement, BLS establishes cooperative agreements with each of the State Employment Security Agencies to

conduct surveys according to the policies and procedures defined by BLS.

Historically, the states have conducted data collection activities and transmit clean microdata to Washington. Over time, there have been a small number of exceptions. Based on explicit needs for special handling and reporting arrangements, data from a few critical, large reporters are collected in Washington, and many large multi-unit businesses report directly to BLS regional offices as a means of facilitating reporting and providing central points of contact and follow-up.

II. BACKGROUND OF AUTOMATED COLLECTION IN THE CES:

The driving force behind data collection research was the need to reduce the size of revisions to monthly preliminary estimates. Beginning with Maine and Florida as test States, a CATI system was developed, research was conducted, and over time more States were added to the project. While considerable success was reported using CATI methodology, full-scale implementation of CATI as an ongoing collection method in a survey the size of the CES was determined to be prohibitively expensive (Werking and Clayton 1991).

To address the issue of costs, while maintaining the improvements in timeliness afforded by automated collection methods, a Touchtone Data Entry (TDE) system was developed as an alternative to CATI (Werking, Tupek and Clayton 1988, and Clayton and Harrell 1989). TDE proved to be effective and popular with both respondents and the State agencies responsible for collecting the data. However, while TDE is very timely and efficient, the personal contact afforded by CATI, which provided opportunities to educate the respondent and improve reporting behavior, was lost. To take advantage of the direct contact afforded by CATI, as well as the cost savings associated with TDE, a two-tiered rollover process was implemented.

Under this strategy, large, chronically late reporters are converted to CATI collection for approximately six months. During this period, interviewers impress on the respondent the importance of receiving their data on time, and resolve any definitional or other reporting issues. After the six month period, the respondents are "rolled" to ongoing TDE reporting. Each month, TDE reporters are provided with an "Advance Notice" postcard, and if necessary, a telephone nonresponse prompt call prior to the cut-off for preliminary estimates.

Goals of the Federal Economic Indicators Project:

Rising concern over revisions between preliminary and revised national estimates spurred implementation of automated data collection procedures. Research into the causes of the revisions established that the majority of the revisions were attributable to delinquent large reporters. Established success in the research and development phase of automated data collection, as well as continuing pressure to solve problems with revisions to the estimates, led to increased funding under the Boskin Initiative, resulting in implementation of the Federal Economic Indicators (FEI) project. The focus shifted to converting all eligible late, large units to automated collection as quickly as possible, to rapidly reduce revisions the preliminary estimates.

Creation of the Data Collection Centers:

Financial, administrative, organizational, and technical considerations led to the creation of the Data Collection Centers (DCCs). The most important issues concerned *timing* and *workload*. The Boskin Initiative created pressure for a rapid, accelerated conversion schedule of large numbers of reporters to augment the conversions by the states. As the DCCs established themselves as a viable alternative, many states chose to forego using CATI, preferring to have the DCCs handle the workload. In doing so, States eliminated their need for a small, specialized staff, which required skills, training, and back-up, for a short duration, high intensity effort. With States choosing to use the DCCs, resources were concentrated in two Data Collection Centers which manage the workload. While there are many advantages to this approach, several issues needed to be addressed to ensure

successful conversion and smooth dataflows among all the organizations involved.

III. DATA COLLECTION CENTERS:

The DCCs are unique within BLS, in that they are located in the regional offices, but administratively are part of the national office. The DCC managers must work through the various regional offices to set up communications and procedures with each of the individual States entering the DCC, including on-site State visits. While data are collected in the Centers, responsibility for the quality of the data, as well as dissemination of the data to BLS and the States, remain with the DCC manager.

While the DCC managers maintain broad discretion in operating the Centers, policy decisions and basic procedures, as well as systems support, reside in the national office. In addition, cooperative agreements with the states are negotiated and monitored by the regional offices, according to instructions from BLS-Washington. Any issues regarding a state must be resolved through the appropriate regional office.

This organizational structure demands that effective communications be established between all the parties involved--several organizations within BLS-Washington, the Data Collection Centers, the regional offices, and the states. To enhance communications, several electronic mail groups have been established to address specific areas of concern, including technical support. Bi-weekly meetings involving representatives from all the organizations involved in the national office are held to review the status of projects, software and procedures enhancements, research efforts and resolve any issues. Every two or three months, the DCC managers travel to Washington for meetings, software demonstrations, and are provided with the opportunity to present their issues and concerns.

As a new state begins working with the DCC, the DCC manager visits the state and works closely to discuss operational procedures and ensure a smooth transition and ongoing operations. These on-site, face-to-face meetings have proven especially effectively in ensuring strong communications and working relationships. In addition, a variety of regional and State meetings often provide opportunities for all of the involved organizations to present

issues and policy, and elicit feedback from the field.

Organization, Staffing, and Workload in the DCCs:

The DCC manager is a BLS employee who reports directly to the Office of Field Operations in BLS-Washington. All other staff, which consists of supervisors and interviewers, are contract personnel. At full capacity, each DCC will include twenty interviewers, two supervisors and a manager. On-site technical support is also provided by existing regional office staff.

Due to the nature of data collection in the CES survey, two weeks out of each month are "peak" collection periods, when the bulk of the CATI calls are scheduled. The remaining weeks are devoted to preparing and mailing the informational packages to newly converted units, file maintenance and other activities. During a typical month, interviewers work a full 40 hours during the peak collection weeks, and 10 to 20 hours during non-peak collection weeks.

Interviewers are hired on a flow basis, with additional staff hired as new states sign contracts to participate in the program. Training of the interviewers is the responsibility of the DCC managers, and is quite extensive. The training program in the Atlanta DCC, for example, extends over a week and covers not only interviewing techniques and hands-on training of the CATI hardware and software, but also extensive coverage of the definitions and edit and screening processes used in the CES program. Virtually all of the interviewers hired thus far are college graduates. Many of these, for a variety of reasons, prefer the flexible work schedule offered by the job.

A full caseload for an individual interviewer is 360 cases. Thus, when the DCCs are operating at full capacity, approximately 7,200 establishments will be collected on CATI in each center in any given month. Since data are collected on CATI for a period of six months before the respondents are converted to TDE and returned to the State, a flow of nearly 2400 sample units per month will move in and out of the Centers. At this rate of conversion, it is estimated that it will take three to four years to convert the an a sufficient number of large reporters to TDE.

Results Achieved in the DCCs:

The success of CATI collection conducted by the States varied considerably, with response rates ranging from 67% to 95%. and an overall response rate averaging around 80 to 85%. By comparison, the DCCs have consistently achieved collection of over 90% of the sample prior to first closing. Also, the variation of response rates across States has narrowed to only a few percentage points. By any standard, centralized data collection has resulted in the highest response rates yet seen in the CES CATI program, and is an unqualified success. The chart below shows first closing response rates broken out by mail, State CATI collection, and DCC CATI collection.

Several reasons may be posited for the improved performance of the Centers. State agencies must handle a wide variety of tasks associated with the CES program, including sample solicitation, processing a large mail sample, generating estimates, and responding to inquiries from the public. The addition of the CATI workload required new staff, new computer skills, and new management attention. In addition, the staff assigned to collect data on CATI often had other duties; CATI collection was not the primary focus of their jobs. By contrast, the DCCs are devoted strictly to CATI operations. The staff are well trained, and chose to work in the field of telephone interviewing. The DCC managers have the time and resources to research any problems, and work constantly to improve procedures.

IV. SYSTEMS DEVELOPMENT:

The Federal/State environment historically has encouraged states to develop their own systems to meet their individual needs. Over the past several years concerted efforts have been made to standardize State processing systems, which have met with considerable success. For standard survey processing, currently, 38 states operate ACES, a mainframe processing system developed by several States and supported by the Iowa State Agency. In addition, BLS-Washington maintains its own system. While the necessity of integrating systems had been recognized, and considerable progress had been made, the FEI project pushed these issues into the forefront. In order for the DCCs to collect data and disseminate it to BLS as well as the States by the required deadlines, the CATI system had to be revised, an umbrella file management system developed, and the

telecommunications among the DCCs, BLS-Washington, and the participating States had to be vastly improved.

Evolution of CATI Systems In the CES:

Initial research work led to the development of a CATI instrument using the CASES software by the University of California at Berkeley. This CATI system was screen based, and provided a wealth of management information that was critical during the research phase of CATI. Taking advantage of the new technology developed during the 1980's, BLS developed a forms-based system written in Foxpro, which was implemented in 1990. Given the nature of the CES survey, which consists of collecting five or six numeric data items, the Labor Information Processing System (LIPS) design simplified the collection instrument, provided for access to 16 months of historical microdata for longitudinal editing, and streamlined the management of the databases and files. LIPS is currently used by all participating States and the DCCs.

Umbrella System:

The stand-alone LIPS CATI system was originally designed to meet the needs of CATI collection in each State. Database and file management functions, import and export of cases, and creation of reports and management information are done on the individual collection PC, with files copied to diskettes for transmission and back-up. In order to accommodate the special needs of the DCCs in the operation of a multiple workstation environment, the File Allocation and Transfer for the Labor Information Processing System (FAT-LIPS) is currently under development. This "umbrella" system makes use of a server, one 386 PC, and one 486 mb fixed drive to provide the functionality needed by the DCC managers to effectively manage a multiple workstation environment.

The FAT-LIPS environment links the 20 individual workstations to a central server, gathering all of the databases from the CATI PCs and transferring one large database to the FAT-LIPS machine. The software on this machine in enable the DCC managers to run all necessary status reports and files of management information for the participating states, as well as BLS-Washington.

The FAT-LIPS system will permit DCC managers to allocate cases based on any field

or combination of fields in the database, such as by State, industry, size, or number of months on CATI. Cases can be also reallocated at any time. At any time, all necessary functions can revert back to the individual collection PCs. This system, when available, will greatly reduce the operational burdens now facing the DCC managers, and provide needed flexibility for efficiently managing the increasing workload in the Centers.

Reversing Dataflows for Improved Timeliness:

Given the decentralized Federal/State nature of the CES survey, and the time-critical movement of both microdata and registry information among the DCCs, the States, and the national office, a new scheme for the effective transfer of data was needed. Prior to the establishment of the DCCs, virtually all data were collected and transmitted by the States. Within the State, all data collected, whether by mail, CATI, or TDE, are key-entered or uploaded to the mainframe processing system, where it is edited and screened prior to transmission to BLS. For two reasons, it was decided that data collected in the DCCs would be transmitted directly to BLS, rather than returning to the States for transmission. First, the overriding goal of the project is improved timeliness. By transmitting directly to BLS, data collected up to the last minute in the DCCs can be received and used in calculating national estimates. Second, LIPS edits and screens all data collected by CATI using the actual criteria used in State ACES systems. Thus, for most States, data collected in the DCCs are subject to the same parameters resident in the State systems, and therefore meet program requirements for receipt of "clean" microdata. The goal was to reverse the flow of data. Rather than have the DCCs send data back to the States for transmission to BLS, the DCC would transmit data directly to BLS, which would then disseminate it to the States.

An Electronic Mailbox:

A data retrieval and dissemination system was already under development. This system, called DataBox, provides the functionality needed to reverse the dataflows. The DCCs transmit data directly from a PC outfitted with a BARRHASP board to the DataBox. The original file is then sorted by State, and the appropriate records are

loaded into each individual State's "mailbox". At this point, the data are available for State use.

Using this approach, the DCC's workload is reduced by transmitting data to one place, rather than to each individual State. By reversing the dataflows, time which would have been spent getting data from the DCCs to the States for transmission to Washington is instead used for data collection, improving response rates by 2% to 7% each month.

IV. THE FUTURE:

The Data Collection Centers were established to perform a specific function in order to achieve a defined goal. After this project has been fully implemented, a number of initiatives may be carried out through the DCCs. High first closing response rates must be consistently achieved, which could involve defining certain sample members as "permanent CATI" units, centralizing TDE collection and non-response prompting activities in the DCCs, research into direct sample solicitation into CATI and/or TDE reporting, or other projects to reduce and control response error. Clearly, the DCCs offer the opportunity for continued research and development in the area of automated data collection techniques, and for the conduct of special surveys.

In summary, the Data Collection Centers represent a new facet in CES data collection operations, adding great potential for pursuing a number of research, development, and operational projects not easily accomplished in the past.

IV. References:

Werking, G. S., and Clayton, R. L. (1990), "Enhancing Data Quality Through the Use of Mixed Mode Collection," *Survey Methodology*, June 1991, Vol. 17, No.1.

Werking, G., Tupek, A., and Clayton, R. (1988), "CATI and Touchtone Self-Response Applications for Establishment Surveys," *Journal of Official Statistics*, Vol. 4, No. 4, 349-362.

Clayton, R. L., and Harrell, L. J. (1989), "Developing a Cost Model for Alternative Data Collection Methods: Mail, CATI and TDE," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

**First Closing Response Rates: Mail vs. CATI
Units with 100 and Greater Employment**

