

AUTOMATED DATA COLLECTION STRATEGIES FOR THE COVERED EMPLOYMENT AND WAGES (CEW) PROGRAM

Michael A Searson, U.S. Bureau of Labor Statistics
Suite 4840, 2 Massachusetts Avenue, NE, Washington, DC 20212

Key Words: survey costs; automation

Note: Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

ABSTRACT

With constant budget pressures and increasing workloads, the Covered Employment and Wages (CEW) program faces enormous challenges for the next decade. The CEW program is the cornerstone of the other statistical surveys of businesses at the Bureau of Labor Statistics (BLS) since it is used extensively as a sampling frame and population controls for employment levels. The CEW program uses the administrative records of the State's Unemployment Insurance (UI) system and supplements these data with information collected from two surveys. Since these data are administrative based, BLS has no option but to process all of these data. The CEW program is also known as the ES-202 program.

This paper focuses on efforts to automate and reduce burden and decrease data collection and processing costs of the surveys—the Annual Refiling Survey (ARS) and Multiple Worksite Report (MWR). The ARS is used to review and update the industrial and geographical codes initially assigned to an employer. The MWR is used to collect employment, wages, and business identification information for each worksite from large employers. This paper addresses the current use of a pilot Touch-tone Data Entry system for the ARS and future research projects as well as the electronic collection of the MWR data, and future web-based projects.

Background

The Covered Employment and Wages (CEW) program is one of six Federal/State Statistical programs operated by the Bureau of Labor Statistics (BLS) under a Cooperative Agreement with the State Workforce Agencies. Under this arrangement, the BLS provides funding to the States to collect, edit, review, and publish data for various economic series. The States follow the statistical methodology developed by BLS, data collection and editing procedures outlined in various program manuals and instructions, and the guidelines and deliverables noted in the Cooperative Agreement. The CEW program

includes the employment, wages, and business identification information for all employers subject to each State's Unemployment Insurance (UI) laws. The CEW data represent approximately 97% of all non-farm workers in the US.

The CEW program uses the administrative records of the UI system for its basic inputs, namely the new employer registration forms (called the Status Determination Forms (SDF)) and the forms used to collect tax information each quarter (called the Quarterly Contribution Report (QCR)). While these administrative records provide the vast majority of the information for the program, BLS developed additional forms to supplement these administrative records to meet the program's further statistical needs. These forms will be discussed in detail in this paper.

When it was implemented over 60 years ago, the primary purpose of the CEW program was to measure the extent of coverage of the UI program. Initial coverage of workers was originally limited to those employers having eight or more workers with no coverage for government or non-profit organizations. Since that time coverage has been greatly expanded and the uses of these data have increased dramatically as well. (See Farmer and Searson, "Use of Administrative Records in the BLS CEW Program.")

Status Determination Form (SDF)

During the initial registration process, the new employer files the SDF mentioned above. The primary purpose of this form is to determine whether an employer is liable for UI coverage under that State's UI law. In addition, the employer is requested to provide information on the principal economic activities that it expects to be engaged in that State. Also collected is information concerning the location of these economic activities. The specific county or counties are requested in addition to the physical location address(es) of these locations. A question on type of ownership (private sector, State or local government, or nonprofit organization) is also asked. Some States ask whether the employing unit performs services for other businesses or individuals versus other units of their own enterprise. The State CEW staff then assigns codes denoting their industrial, county, ownership and operating or supporting role status.

Please note that the new employer is requested to provide information on their expected business activities. Perhaps, the employer will manufacture two products, A and B, which are classified in different industries. Let's say that Product A will constitute 70% of expected revenue, whereas Product B will contribute 30% of the company's expected sales. Within the first six months of the new business, the demand for Product A declines, whereas the sales for Product B increases. At the end of the first year in business, Product B is now the dominant product with 65% of sales compared to A's 35%. Consequently, the CEW program now has a situation where this business is misclassified since Product B is now the dominant product and it is classified in a different industry. In addition, sales for Product B rise so rapidly that the business opens another location in a different county. Now, the CEW program has a situation where some of the employees are being reported in an incorrect county, too. These two examples illustrate the need to periodically review the industrial and county codes assigned to an employer, as well as its single/multi worksite status.

Multiple Worksite Report (MWR)

Since the ES-202 program collects data on worksites, rather than employers, the BLS introduced the MWR in 1991. The MWR replaced individually designed State forms that collected employment and wages data from employers with workers located in different counties and/or industries. These data were collected on an industrial/county level basis and thus, required employers to summarize those situations where they had multiple locations in the same county that were performing the same economic activity. The MWR collects employment and wages data at the worksite level. The employer is also requested to provide the trade name for the worksite, its physical location address, and a worksite description uniquely identifying it in their payroll system.

Every quarter each State mails the MWR form to all employers meeting the CEW program criteria. The MWR lists all of the worksites identified by the employer on the prior quarter's MWR form. Any updates to the worksites' addresses, trade names, and/or worksite descriptions are noted on the next quarter's MWR form. The employer is requested to post the employment for each month of the quarter and the quarterly wages for each worksite. The employer should add new worksites and note those that are closed or sold to another employer. Any further updates to the business identification information for each worksite should also be noted on the MWR form.

Annual Refiling Survey (ARS)

To maintain the quality of the code assignments (industry and county), physical location addresses, and single/multi worksite status, the States currently mail a questionnaire to approximately one-third of their active UI employer accounts each year. This process is called the ARS. The

selection criterion is based on an employer's federal Employment Identification Number (EIN). By using the EIN as a selection factor, an added benefit is that all locations of the same enterprise (those using the same EIN) will be contacted the same year, regardless of where they have employees (different States). This factor will play a major role in one of our proposed efforts to reduce the employer reporting burden as well as cutting State costs to conduct these review activities.

BLS developed three forms for the States to use to conduct this survey. All of these forms have been reviewed and approved by the Office of Management and Budget. The first is called the Single Worksite Verification form (3023-NVS). It is mailed to all employers who have indicated that they are single worksite employers and who have a valid industrial code currently assigned. Most of the employers (approximately 2.1 million) received this form for the FY 2002 ARS.

The employer is first requested to review the address where the ARS form was sent, and asked to verify that this is the correct mailing address for the business to receive forms for statistical purposes. This address may be the address of the business, the accountant, or that of a firm providing UI claims services or payroll/tax filing services for the business. For a new employer, the address that the ARS is mailed to is typically the UI tax address. If the firm wants this form and others similar to it mailed to a different address in the future, then this is the employer's opportunity to advise the State of this information. If the employer provides a different address on the ARS form, then the next cycle of the ARS will use that mailing address. The updated address provided by the employer is referred to as the statistical mailing address. The CEW program has the ability to store three addresses for each employer—the UI tax address, the statistical mailing address, and the physical location address.

Next, the employer is asked to verify the physical location address that is pre-printed on the ARS form for its business. If it is incorrect, the employer is asked to provide the updated address. If no physical location address is on the CEW database for that employer, the field is left blank and the employer is requested to provide this information. The next question is intended to verify the county in which the business is located. The name of the county currently assigned to that employer is displayed and the employer is asked to verify its accuracy or provide the updated county information. The employer is also asked if the products being produced or services being rendered are for other businesses or individuals or exclusively for other units of the same enterprise.

The employer is also requested to answer a question dealing with the number of worksites located in that State under its UI account number. If there are other worksites within the State, then the employer is requested to list all of these worksites, their physical location addresses, and

the number of employees at each location. If the employer meets the criteria for filing an MWR, then CEW staff will send the employer an MWR form to complete. As noted earlier, the main purpose of the MWR is to dis-aggregate the employment and wages reported by the employer on the QCR to the proper industry and location so that the industrial and county integrity of the CEW program is maintained.

Finally, the employer is requested to review a brief description of the economic activities that are included in the industrial classification currently assigned. If the employer's staff agrees with this description, they simply check the "Yes" box and the ARS form is complete. If they disagree with the industrial description provided or are not sure that it is correct, then they are requested to provide information on their current economic activities. Space on the ARS form is provided to collect this information.

A number of factors impact the number of changes that are noted on the 3023-NVS form each year. The quality of a State's SDF is certainly a factor as the amount of space allotted to collect the economic activity, physical location address, and county information is critical to the assignment of accurate codes. Likewise the quality of the staff assigned within a State to assign these codes is another factor, as inexperienced staff are more likely to assign incorrect codes, even if the information is complete and accurate. With the recent trend toward one-stop new business registrations in the States (register for UI, Income Tax withholding and/or sales taxes at one time), a number of States no longer have control of the SDFs, have limited access to the information collected, or only see an image of the SDF. Reducing the number of errors on the initial assignment of codes for new employers and the omission of the physical location address increases the cost (e.g., data entry of the update information) of the ARS as these fields are more likely to be noted for correction during that process. All updates, however, are not errors because many employers do change the nature of their business activities over time, their business locations and/or expand their businesses to other locations. That is the purpose of the ARS—to note these changes and update the classifications and addresses of these businesses. The main point to be noted is that initial errors or data omissions do ultimately raise the cost of conducting the ARS.

One of the other ARS forms is the Verification form for Multi-worksites employers, the 3023-NVM. This form is similar in concept to the 3023-NVS except that it is mailed to all employers who have multiple worksites in the State. The main difference is that all of the worksites are listed and the employer is requested to review the physical location address for each worksite. Also, all of the worksites performing the same economic activity are grouped together, and the employer is asked if the industry description that is printed is applicable to all of these worksites. "Yes" and "No" boxes are provided for

each worksite. If the employer is a large mass retailer with 100 locations in a State, then it receives an ARS form listing all of these stores. If the employer has similar stores in other States, each State will mail the employer a set of ARS forms for their State.

The last ARS form, the 3023-NCA, was designed to deal with situations where the employer never filed an SDF with the State and thus is "unclassified" on an industrial basis. The county code for this employer is also probably unknown and is unclassified, too. The States are instructed to mail these forms on a flow basis (normally once a quarter) to these employers. States are instructed that the number of unclassified accounts should not exceed 0.5% of total employment. Again, the number of unclassified accounts is directly related to the same factors noted earlier dealing with errors being introduced in the initial assignment of codes during the new employer registration process.

Budget and Workload Issues

For the first quarter of 2002, the States provided data for 7.0 million employers and 8.2 million worksites. Since the CEW program represents a universe count of employers and the number of their workers and wages, the workload increases in an expanding economy. During the latest ten year period (first quarter 1992 to first quarter 2002), the number of employers increased 21.8% whereas the number of worksites increased even more (24.9%). During that same period, budgets for the program have not kept pace. BLS does not have the option of cutting the sample since this is a universe-based program. The only alternatives are to fully examine how some of these program activities are being conducted and develop possible cost-cutting proposals. In this way, BLS is looking towards modifying methods and using new technology to address this imbalance, at least in part to slow the inevitable erosion of data quality.

Approximately 114,000 employers are mailed an MWR form each quarter. In FY 2002, approximately 2.1 million employers were mailed an ARS form. BLS pays for the outgoing postage of these forms as well as providing a business reply envelope for their return to the State.

The factors affecting the costs of these surveys can be divided into a number of categories. Both surveys require that forms be printed. In addition, the ARS requires a cover letter and the MWR requires one for the first quarter report of each year. Depending on the degree of sophisticated equipment available, the State staff may also have to manually fold and stuff the cover letters and forms into envelopes. In other States, this is an automated process. Upon the employer's return of the forms to the State, the envelopes must be opened and sorted in various categories prior to additional processing. Many States are using bar codes or other types of scanning software to note the return of the survey form to prevent an inadvertent follow-up for non-response. In the ARS,

further processing would require a separation of those forms that require no additional review and those that do. The former would then be assigned a new response code requiring more data entry and ultimately filing. The latter would undergo further review and analysis, then ultimately require additional data entry and finally filing.

The MWR follows similar processes—printing, folding and stuffing, mailing, and upon return, opening and sorting into two categories—those requiring only data entry (no updates to existing worksites) and those requiring updates (new or closed worksites, corrections, etc.). The latter group would be reviewed and analyzed, updates assessed and then data entered.

The basic processes, then, are: printing, handling the form on the mail-out, handling of the returned forms, data review, and data entry. Thus, costs can be summarized as: printing, handling in and out, postage out and return, data review, data entry, and filing. To assess which potential alternative survey methods would provide the biggest return for the investment, one would need to develop a matrix of these activities and determine which methods impact the most activities. At the same time, one must remember that the costs of these activities may vary significantly. The methodology that has the most boxes checked may not be as effective as one that has fewer boxes checked if those are the ones that determine most of the State's survey costs. See the chart at the end of this paper for a graphical presentation of the cost savings associated with each method.

On the other hand, some alternative strategies may take a significant amount of dollars and BLS staff time to develop in order to produce long term savings. Thus, some of the methodologies may be easy to implement and not save that many dollars, but are worthwhile to pursue while other methodologies are being developed. In some cases, a mixing of methodologies may be appropriate to allow respondents a number of options to file their responses.

Another factor that must be included are the costs of conducting follow-ups for non-response. If a particular methodology requires extensive follow-ups either for non-response or clarifications, the anticipated savings may be significantly reduced.

Proposed Strategies

Status Determination Forms

BLS staff recently completed a review of all SDFs noting the areas where improvements were needed to improve the quality of the data collected. Since UI staff design these State forms for their own purposes, the CEW staff can only suggest changes. State CEW staff were instructed to open discussions with UI staff to review the recommended changes and develop a plan to incorporate

some of these changes in the next redesign of the SDF in their State. To assist in these efforts, BLS developed and provided the States with a PowerPoint slide show that illustrates State UI staff uses of CEW data and their impact on overall UI operations. Since the SDFs are not revised annually, it may take a few years to accomplish these objectives. To deal with the need for training for new staff performing industrial coding activities, BLS staff is developing a new Computer Based Training tool. It is an inter-active tutorial that describes basic industrial coding principles. After its completion, BLS is planning to develop an advanced course for more experienced coders, further strengthening coding knowledge and consistency of the States. These activities should reduce the number of initial coding inaccuracies and thus the number of updates resulting from the subsequent ARS.

Touch-Tone Data Entry (TDE)

In FY 2002 BLS returned to using a Touch-tone Data Entry system for the ARS. It had been tested in a few States in the late 1990s with mixed results. The TDE system was not suited to address the needs of the conversion from the 1997 Standard Classification System (SIC) to the North American Industry Classification System (NAICS) that began in October 1998. When re-introduced in October 2001 as a test in five States, the initial approach was modified so that only those employers meeting certain criteria were allowed to participate. The project's name was also changed to the Touch-Tone Response System (TRS). This selectivity feature raised the response rates and also reduced the number of phone calls requesting clarification on the procedures. Forty States plan to participate in this project in FY 2003. Assuming the same response rates are achieved next year, then the costs of conducting the ARS would be reduced by approximately \$200,000. For more detailed information on the TRS, please see "New Data Collections Using Touch-tone Data Entry" by S. Jakhu and M. Sauer. One limitation of the TRS is that any employer whose information on the ARS form requires an update can't use the TRS to respond. In these cases, the employer is instructed to return the completed ARS form to the State. Thus, improvements in the SDF process can pay additional dividends, as the number of TRS eligible employer accounts would also increase.

Fax

Allowing the employer to fax the ARS form back to the State could reduce the costs of using a business reply envelope. If this process proves effective, the second phase of a fax process could include the ability to create an image and ASCII file of the ARS form for each employer and fax it to them. The employer would then complete the ARS form and fax it back to BLS. This additional functionality would further reduce many of the processing steps noted earlier.

Internet (Web based Collection)

A more cost-effective approach is to develop a web-based ARS system that allows the employer to either use TRS or the web to indicate that its information is correct. The web-based system could also be designed to capture any updates that the employer indicates is necessary. In this scenario, the employer would be mailed an ARS form and offered these options. Included in the cover letter would be an account number and a password that the employer could use to access the BLS website. The employer would only be allowed to review its own information and the password would only be active for a fixed number of days (probably 60 to 90).

This methodology would reduce the ARS costs by eliminating the return postage, handling costs on the response, filing and data entry costs. In addition, it could also be designed to allow for some editing of the employer's response, if necessary. If the physical location address being provided is not sufficient for geocoding purposes, the web-based system could ask for additional information. However, the labor, systems component, and maintenance costs are not known, making "savings" amounts impossible to calculate at this time.

Central Collection of Large Employers for ARS

BLS is also initializing a centralized ARS data collection system for the larger employers that are operating many locations in many States. For FY 2003, staff will develop the procedures that will be used in succeeding years to eliminate the need to mail ARS forms to these large employers. A limited number of firms will be selected for FY 2003 with the number of firms selected being expanded in succeeding years once the procedures and system modifications are set. When the CEW program was initiated in the 1930's, the US economy was dominated by locally owned firms, particularly in the retail trade sector of our economy. With the arrival of the mass merchandise retail firms from the 1960's forward, large national chains now dominate most of retail trade. Since most of these stores perform the same economic activity, it doesn't seem necessary to mail the ARS forms to these employers. Most BLS and State staff are familiar with their products. In addition, their websites supply enough detail to determine their correct industry code. Consequently, these employers will not be mailed a series of ARS 3023-NVM forms listing all of their worksites in each State. This process should reduce the employer reporting burden as well as reducing State costs for printing, handling, postage, and filing.

Centralized Printing:

One other cost-cutting option being explored for the ARS is the potential use of an efficient centralized facility to conduct many of the ARS tasks. During FY 2003, Phase I of this option is being tested with five states. This initial phase is restricted to simply printing the ARS forms for these states and returning them for further processing and mailing. In Phase II the facility could print, fold and stuff the appropriate ARS forms and cover letters; mail using pre-sorted first class cost-savings; and possibly include the return of the ARS questionnaires. All States could use the services of this facility.

MWR --Electronic Data Interchange Initiative

With regard to the MWR, the BLS has been soliciting large multi-State employers to its Electronic Data Interchange Center in Chicago since its inception in 1995. The strategy being employed for large firms can be reviewed in detail in M. Searson's "Strategies to Implement Electronic Collection of Multiple Worksite Report Data." That strategy emphasizes the addition of electronic reporting in the systems of those firms selling payroll/tax-filing software. The purchasers of these software products are normally large multi-State employers with multiple locations in most States. Similarly, BLS staff has also been working with firms providing payroll/tax-filing services for their clients. These firms would then offer the MWR electronic reporting as an additional service for their clients. The extra advantage for BLS in this scenario is that these firms provide these services for large, medium, and small employers.

Internet (Web-based Collection)

BLS staff responsible for the MWR data collection efforts are also developing a web-based Internet product. Within the past 12 months, staff in 10 States conducted a fact-gathering project that noted the typical errors, data omissions, problems, etc. over a six-month reporting period. BLS staff developed a proposed functionality for the new system and asked the States to determine if this met their needs. After two successive quarters of reviewing this proposed functionality, the States and BLS staff agreed that the system would be designed in two phases. The first phase would be restricted to a "bare bones" system whose sole purpose would be to collect the MWR data. The second phase would use the problems noted in phase one to determine the types of employer reporting issues that could be addressed in a more systematic manner and possibly resolved during the data collection process. For example, an employer's failure to add new worksites could be addressed by merely prompting the employer with a question on this issue before the report was noted as final. Likewise, a significant change in the level of employment and/or wages for a worksite could also be brought to the attention of the employer.

Potential ARS Collection Methods and their Impact on Survey Processes, Activities, and Costs

Method	Process											
	Print	Handling	Postage		Handling Returned Forms							
			Out	Return	Open	Scan	Sort	Data Review	Assign Response Code	File	Data Entry Update	
TDE (TRS)				■	■	■	■	■	■	■	■	
Fax Phase I				■	■							
Fax Phase II	■	■	■	■	■	■	■	■	■	■	■	
Web Collection				■	■	■	■	■	■	■	■	■
Central Collection	■	■	■	■	■	■	■				■	
Centralized Print Facility Phase I	■											
Centralized Print Facility Phase II	■	■	■		■		■	■	■	■	■	

Legend ■ Results in cost savings

The author wishes to acknowledge the assistance of Dori Allard of BLS in reviewing this paper and providing numerous constructive suggestions on its content and layout.

REFERENCES

Farmer, T. and Searson, M. (1995), "Use of Administrative Records in the Bureau of Labor Statistics' Covered Employment and Wages (ES-202) Program," 1995 Bureau of the Census Annual Research Conference, Washington, D.C., March 1995.

Jakhu, S. and Sauer M., "New Data Collections Using Touch-tone Data Entry," American Statistical Association, New York, 2002.

Searson, M., "Strategies to Implement Electronic Collection of Multiple Worksite Report Data," Federal Committee on Statistical Policy, Washington, DC (November 2001)