Fostering modern and innovative data collection methods in several U.S. Bureau of Labor Statistics business surveys

Kenneth W. Robertson (U.S. Bureau of Labor Statistics)

Robertson.Ken@bls.gov

Abstract

The U.S. Bureau of Labor Statistics (BLS) manages many monthly, quarterly, and annual data collections. Within BLS, the Office of Industry Employment Statistics (OIES) is responsible for data acquisition efforts that result in over 40 million collections of business data each year.

The programs managed by OIES include the Current Employment Statistics (CES) survey, the Quarterly Census of Employment and Wages (QCEW), and the Job Openings and Labor Turnover Survey (JOLTS). The CES survey is a very large scale quick response establishment survey, collecting data from a sample of about 651,000 business and government establishments each month. The CES program produces estimates of employment, hours, and earnings by industry, for the nation, states, and metropolitan areas. These data, published about three weeks after the reference period, are among the first indicators of the health of the U.S. economy each month. The QCEW is a quarterly submission of data by the nearly 10 million establishments in the U.S., including monthly employment levels and quarterly wage totals. The JOLTS collects data from a sample of 16,000 establishments each month; data collected include job openings, hires, quits, layoffs and discharges, and other separations.

These data collections utilize mixed modes that are managed centrally and collected across several data collection centers. The methods utilized to collect these data have fundamentally changed over the past two decades. At the beginning of that period survey and administrative data were collected almost exclusively by mail, while data are now collected using multiple methods with most of the data collected electronically. The transformation of the data collection methodology required a substantial investment in managerial and human resource capabilities.

This paper describes the data collection management structure, the human resources utilized, and how tasks are allocated across personnel dimensions. The paper also describes management efforts that foster and lead to excellence and innovation in these programs. In addition, the paper describes recent data collection innovations and current explorations into the realm of alternative data sources.

<u>Background</u>

Data collection in the BLS[i] Office of Industry Employment Statistics is serious business. It is an ongoing effort every day, with nearly $70 million U.S. allocated to this business process each year in this office. These data collection efforts are focused on collecting business data to support the CES, JOLTS, and QCEW programs.

The CES[ii] program collects business data each month to support preliminary, second release, and final release estimates for national data, and preliminary and final release estimates for state and area data. These data are collected from a sample of about 651,000 business establishments (or worksites) each month. This represents about 7.8 million initial collection attempts each year. Additional collection attempts are made for those businesses who do not respond to these initial requests. For details on the management of the CES program, see Robertson (2016)[iii]. For substantial detail on the collection of data for the CES program, see Robertson and Hatch-Maxfield (2012)[iv]. Recent work on the best time of the day to remind CES respondents to report was described by Johnson (2016)[v].
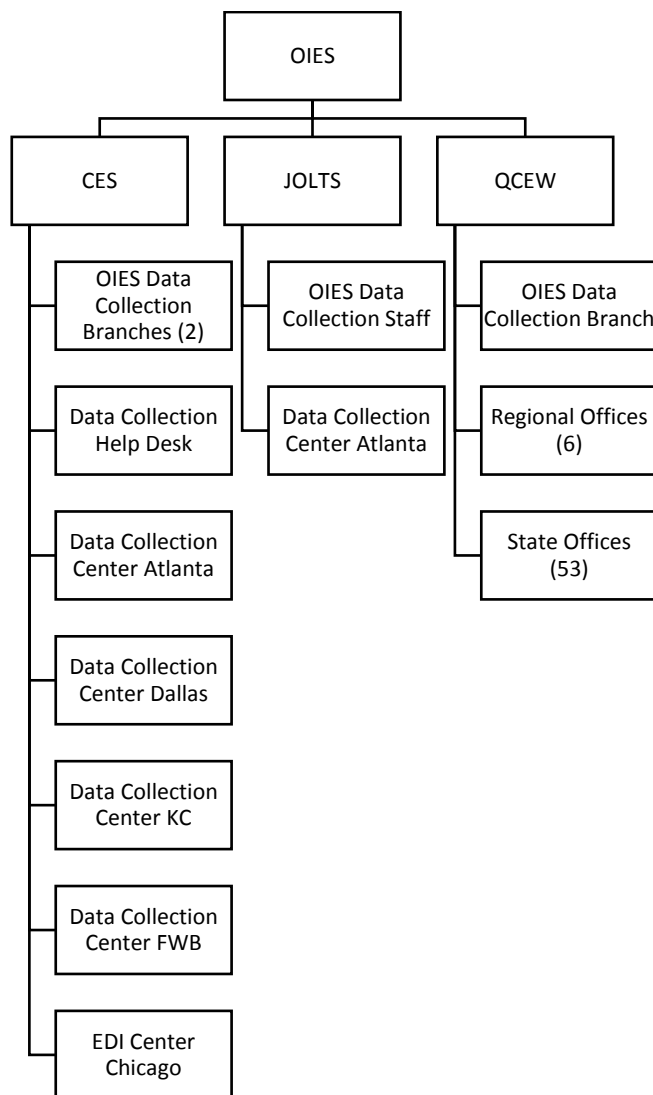
The JOLTS[vi] program collects business data each month to support preliminary and final estimates for national and regional data. A case study on the management of the JOLTS Data Collection Center is available in Calhoun and Jackson (2008)[vii]. These data are collected from a sample of about 16,000 business establishments each month. This represents about 192,000 initial collection attempts each year. As with CES, additional collection attempts are made for those businesses who do not respond to the initial requests.

The QCEW[viii] program is a quarterly statistical business register constructed from administrative records. The raw data are a mandatory report in the U.S. for the purposes of securing funding from businesses to support employees who are laid off. This unemployment insurance (UI)[ix] is primarily supported by taxes each state collects from businesses as part of this quarterly reporting. The UI data are then obtained by the QCEW staff at each of the State Workforce Agencies. The QCEW state staff reviews the data for accuracy, and edits and adds information where needed in order to develop a high quality business register containing quarterly information on the nearly 10 million business establishments in the U.S. As part of the collection effort, BLS manages a quarterly data collection to get *worksite data* for businesses who report to the UI program as an *aggregate multi-site* employer; this is called the Multiple Worksite Report[x]. BLS also conducts an annual data collection, where one third of all employers with more than three employees are asked to verify their industry and address; this is called the Annual Refiling Survey[xi] (ARS). The data collected for the QCEW program represent well over 40 million data collection attempts. However, a large part of this collection is mandated by law, and that part is collected and enforced by a non-statistical agency managing each states UI program. The collection of the 40 million raw data records from the state UI office is a collaborative affair – because the state and national UI programs uses the data outputs from the QCEW program as quality checks for the administrative data the UI program is collecting.

Data Collection Management Structure

The management of OIES statistical programs follows the requirements of the U.S. Office of Management and Budget. For a description of these requirements see Robertson and Eltinge (2013)[xii].

The management of data collection in the OIES spans multiple groups. Monitoring data collection is a critical component of senior management work. Data collection is something that OIES senior managers monitor every day for our monthly data collections, and frequently for our quarterly and annual data collections. Within the office there are dedicated resources for whom data collection management and improvement is the sole focus of their job. The structure of this Data Collection Management Team is shown below.

```
                          ┌──────────────┐
                          │     OIES     │
                          └──────┬───────┘
          ┌──────────────────────┼──────────────────────┐
    ┌───────────┐          ┌───────────┐          ┌───────────┐
    │    CES    │          │   JOLTS   │          │   QCEW    │
    └─────┬─────┘          └─────┬─────┘          └─────┬─────┘
          │                      │                      │
  ┌───────────────┐      ┌───────────────┐      ┌───────────────┐
  │  OIES Data    │      │  OIES Data    │      │  OIES Data    │
  │  Collection   │      │ Collection    │      │ Collection    │
  │  Branches (2) │      │    Staff      │      │   Branch      │
  └───────────────┘      └───────────────┘      └───────────────┘
  ┌───────────────┐      ┌───────────────┐      ┌───────────────┐
  │ Data Collection│     │ Data Collection│     │ Regional      │
  │   Help Desk   │      │ Center Atlanta │     │ Offices (6)   │
  └───────────────┘      └───────────────┘      └───────────────┘
  ┌───────────────┐                             ┌───────────────┐
  │ Data Collection│                            │ State Offices │
  │ Center Atlanta │                            │     (53)      │
  └───────────────┘                             └───────────────┘
  ┌───────────────┐
  │ Data Collection│
  │ Center Dallas  │
  └───────────────┘
  ┌───────────────┐
  │ Data Collection│
  │  Center KC    │
  └───────────────┘
  ┌───────────────┐
  │ Data Collection│
  │  Center FWB   │
  └───────────────┘
  ┌───────────────┐
  │  EDI Center   │
  │   Chicago     │
  └───────────────┘
```

There is management at each of the levels depicted in this organizational chart. The work of the teams in the lower boxes is monitored and directed by the staff in the OIES office. That is, each of the teams labeled OIES Data Collection (Branches, Staff, or Branch) have an oversight role in directing and monitoring the work of the data collection centers, in coordinating the data collection work of the Regional Office staff and the data collection and editing work of the State Offices. The Regional Office staff take on several roles; they monitor State Office deliverables, they are responsible for personally collecting data from selected large businesses, and they serve as trainers for the state office staff.

Data Collection Human Resources

The human resources for data collection in these OIES programs spans those same groups and locations.

Program								
CES			**JOLTS**			**QCEW**		
Site	Managers	Staff	Site	Managers	Staff	Site	Managers	Staff
OIES CES Data Collection Branches (2)	2	10	OIES JOLTS Data Collection Staff	0	1	OIES QCEW Data Collection Branch	1	7
Data Collection Help Desk	1	9	Data Collection Center Atlanta	5	26	Regional Office (6)	3	24
Data Collection Center Atlanta	10	62				State Office (53)	53	342
Data Collection Center Dallas	8	39						
Data Collection Center KC	17	89						
Data Collection Center FWB	11	79						
EDI Center Chicago	6	28						
IT Resources	3	24	IT Resources	0	1	IT Resources	1	15
TOTAL	58	340	TOTAL	5	28	TOTAL	58	388

This is a grand total of 877 supervisors and staff dedicated to data collection and editing activities for these programs.

The category labeled as Managers includes both federal and contract supervisors, and staff whose primary function is quality assurance (QA). The federal staff do not directly manage the contract staff; what they do is monitor the contract deliverables and work with the contract managers to ensure that the many contract requirements are met.

The QA staff look at data collection from a holistic point of view. For example: they review response rates for different categories of data to direct response efforts as collection progresses; they run aggregate data checks to identify emerging trends in response errors; they listen to random interviews to ensure adherence to protocols; and, they run comprehensive data checks to ensure that there are no indicators of potential data falsification.

How tasks are allocated across personnel dimensions

Both CES and JOLTS use a Computer Assisted Telephone Interviewing (CATI) system called TopCATI (or WebCATI) to initiate new businesses into these surveys. TopCATI is also used to collect data while training these respondents to report for the correct survey concepts and payroll periods, prior to training them to self-report on the web. The TopCATI system allows a supervisor – typically a contract supervisor – to assign cases to individual contract data collectors and to monitor the resulting response rates by interviewer. The software also allows the federal QA staff to monitor overall collection, and to monitor collection rates by industry or for groups of respondents. At the national level, other software is used to generate management reports to monitor the overall collection process. Principle among these are SAS[xiii] programs that are used to determine the status of the data collection and generate MS Excel[TM] spreadsheets and graphs at a summary level each day.

Within each Data Collection Center, tasks are allocated depending on the part of the monthly collection cycle we are in. For two weeks, the primary focus of each interviewer is to collect data from recently initiated businesses, and following back with businesses whose reported data seem unusual. For the other two weeks each month, new businesses are being enrolled into the survey – making the initial calls to identify the correct respondent within the business, getting the first initiation package ready and mailed, and then recontacting the respondent to walk them through the survey materials and collect their first monthly report.

Within OIES, we have several meetings each year where all of the federal data collection managers meet to share experiences and learn from each other, discuss any proposed changes to procedures, and to ensure that they are all on the same page with respect to official policies and procedures. OIES managers also make a significant effort to visit each Data Collection Center once per year, in order to share information with the federal and contract staff, and to hear their issues and concerns.

Management efforts to foster and lead excellence and innovation

The OIES managers are expected to develop a culture of excellence, and to foster staff development efforts that lead their staff to have the same high commitment to excellence that the managers have. They do this by working on several different fronts to get this result.

A first critical area is a focus on hiring outstanding motivated candidates when we have vacancies. All of our federal employee hires have a relevant college degree; either a degree in economics or statistics. Many of our hires are either recent college graduates, or veterans who bring leadership experience and a strong service motivation with the relevant academic credentials.

Next, we develop a culture that provides safe opportunities to fail at new ideas. Staff who propose new ideas are given space to develop the ideas. Those ideas that have promise are more fully researched, developed, tested further, and if warranted the idea is implemented. Many fail, and that's OK. It is useful to have a list of ideas that we've explored that we know are not improvements, so when we are asked what we've done to improve the survey we have both a list of implemented improvements and a list of researched ideas that were ultimately rejected. Ideas that are implemented typically have demonstrated that they either lower cost without harming data quality, they improve efficiency of operations, or they do both. Ideas that are implemented earn the staff who proposed it recognition from their peers and supervisors, as having thought up a great idea and working the problem until it was implemented. When available, we typically award such staff with modest awards and with a permanent recognition in their personnel file.

We devote significant efforts to provide both in-house and academic training for staff. Many of our staff enter the federal service with a Bachelor's degree and then work toward a Master's degree in Economics or Statistics. These efforts give us staff who are highly motivated and appropriately educated to understand and improve the quality of our complex statistical processes, including the ever expanding literature on data collection theory and practice.

We require staff to work on teams to accomplish many projects. This is an outgrowth of the effort to find improvements to our methods. Once a new idea is accepted by the manager for work, a team is formed to further explore and develop the idea. This process encourages each staff member to learn to function as a productive team member. BLS has a number of documents that assist managers and staff members in understanding how to organize and facilitate teamwork.

A last key to fostering a culture that inspires innovation is one that also contributes to the development of new leaders. In practice I try to have two co-leaders for each team. The first should be an experienced leader, either a supervisor or a senior staff member who has had some leadership training and experience. The second should be a less experienced staff

member, perhaps with limited leadership training but no hands-on experience. The latter co-leader is expected to learn from the first, and thereby gain experience. Operating in this manner expands leadership training opportunities and experience, creating a more confident and trained cadre of potential leaders.

Recent data collection innovations

(1) Spreadsheet tool for online data entry

For many years, CES had mailed to several hundred respondents who had more than a handful but fewer than hundreds of worksites, a specifically formatted spreadsheet for them to report their monthly CES data on. The spreadsheet was updated by the respondent and then uploaded to a BLS website each month. Each year a fresh spreadsheet was sent out to these companies. The spreadsheet was very popular among these respondents, but, many of these respondents customized the spreadsheet to make it easier for them to report the data or to add customized information useful to them. These customizations made the automatic reading of the spreadsheet and the extraction of the reported data fail; creating work each month for OIES staff. In November 2013, CES staff began a project to address this problem. The respondents were surveyed to identify what their preferences were for reporting in this manner. About a third of the surveyed respondents provided input. For several years, prototype designs were explored, to meet the needs of these respondents.   In 2017, a new collection tool[xiv] was developed and tested among a small group of participants to test usability. The production release of this tool is imminent, scheduled for September 2018. The tool will allow users to directly enter data into the web-based spreadsheet-like tool, or they can copy and paste columns from their own spreadsheets into this tool. A graphic of the tool is provided below.

BUREAU OF LABOR STATISTICS
Current Employment Statistics

Update Respondent Information | Help | Logout

Data Entry

The online spreadsheet below offers a number of features to facilitate data entry. View hints for entering data.

- **Copy & paste** across cells (use keyboard commands Ctrl+'C' for copy and Ctrl+'V' for paste).
- **Sort rows** ascending/descending (click on the column header).
- **Expand** location information (click '+' next to 'Location Information').

| June 2018 | July 2018 | **August 2018** | Please report data for the pay period that includes the 12th of the month. |

	Location ⊕	Pay Frequency ⊖		All Employees					⊖	Production/Non-supervisory Employees			⊖
	Location	Payroll	Commissions	Count	Women	Payroll (excluding commissions)	Commissions (paid at least monthly)	Hours (including overtime)	Count	Payroll (excluding commissions)	Commissions (paid at least monthly)	Hours (including overtime)
1		▼	▼									
2		▼	▼									
3		▼	▼									
4		▼	▼									
5		▼	▼									
6		▼	▼									
7		▼	▼									
8		▼	▼									

Continue

(2) Continuing the move toward email blasts for prompting web reporters

The BLS has been utilizing internet data collection for many years; the CES program was an early adopter in this realm and has been collecting data via the web since 1996. As part of the web collection process, we utilize email prompts to encourage respondents to report. We call these email blasts because, depending on the program, we may be sending out tens or hundreds of thousands of emails in a short time span as part of this collection process.

The QCEW's Annual Refiling Survey began web collection through the BLS Internet Data Collection Facility (IDCF) in 2012. The ARS surveys about 1.2 million reporters per year. In 2012, 6.2 percent of the ARS respondents reported via web. In 2014, the collection was 27.5 percent web[xv]. In 2017, the ARS was a paperless survey, and it achieved a 77.0 percent unit response rate; this transformation has dramatically reduced the costs of this very large survey.

For the QCEW's Multiple Worksite Report, web data collection is limited to those reporters who sign up to participate in web collection. Email blasts to this group has been immensely successful. For example, in the first quarter of 2018, more than 89 percent of companies that signed up for web reporting reported on time based simply on email blast prompts – saving thousands of dollars in printing and postage costs and providing BLS with 59 thousand company reports.

<u>Current explorations into alternative data sources</u>

OIES has produced two recent products that are the result of using alternative data sources. The first matches the BLS QCEW with a public data set from the U.S. Internal Revenue Service that identifies all 501 (c)(3) non-profit business entities in the U.S. This match produces employment and wages associated with non-profit businesses[xvi], and allows a comparison of those businesses with the more traditional for-profit business. The second data product is a match of the BLS QCEW with data from the U.S. Bureau of Economic Analysis on foreign direct investment in U.S. companies. This matched dataset is further matched to the BLS Occupational Employment Statistics survey. The resulting data allow for profiling of occupational staffing patterns, wages, and employment for firms with significant foreign investment and for those without significant foreign investment.

Several new projects are just getting underway to take advantage of publicly available data. Two of these are associated with the QCEW's Annual Refiling Survey. As mentioned previously, the ARS surveys nearly one third of all U.S. worksites each year, asking them to verify their physical location address and asking them to verify or update their industry. The way this survey works, we do not spend any additional resources on a worksite that has had its industry and address verified (i.e. it is unchanged from the previous report); only worksites with changes warrant further review. The two new processes are expected to result in a high percentage of verifications with no further review needed. The first project takes advantage of multi-worksite business respondents who want to report this information for all of their worksites at once by supplying a file with all of the data. This will allow us to reduce their respondent burden and to reduce our survey costs; a win-win for everyone! The second is to identify companies who tend to keep a full list of like-business addresses on the company website, and scrape that data instead of mailing a survey form to hundreds or thousands of worksites. For example, many restaurants keep their restaurant addresses on the web so that customers can find locations while traveling. A third venture is to explore the use of high-quality third-party sources for this same data. For example, the U.S. Federal Depositary Insurance Corporation maintains an online database of all U.S. banks, their location, and their detailed industry classification. Leveraging this data source is likely to provide consistently high quality location and industry classification data. Since we are verifying industry and location, we will flag those cases that differ from our current data for follow up to confirm the change. The potential to save survey resources and respondent burden through these efforts is significant.

<u>Final Thoughts</u>

Data collection is serious business. It requires a substantial share of the funding for a survey program. In order for data collection to be done well it requires constant management attention. This process needs managers and staff who are well versed in the state-of-the-art

methodologies available for data collection, and it requires a willingness to explore new methods to maintain high response rates and limit response burden where possible. Given declining response rates for more traditional methods, data collection requires these new methods to keep response rates high and to ensure high quality data. Where the provenance of alternative data is well known, and where the data owner has given consent, it may be found to serve an important role in the overall data collection. As a method to obtain data to reduce respondent burden this seems like a great choice.

Alternative data can, and should be used where possible to create new data products. This allows the creation of new data products without incurring an additional cost of data collection, and without further burdening the data provider. This is a value added leveraging of work the respondent already did. If done to create a needed product then this is simply good cost-effective governance of federal funds.

I do have one serious note of caution about the use of alternative data. Official statistics, especially those presented as time series, require consistency of inputs to ensure that the estimated changes are economic in nature, and are not spurious changes induced by a change in a third party input. For this reason, while I am a wholehearted fan of using alternative data sources as substitutes for microdata reports or data items where it makes sense, I am considerably wary of using non-governmental alternative data sources as model inputs into official statistics.

Opinions expressed in this paper are those of the author and do not reflect official policy of the U.S. Bureau of Labor Statistics.

<u>References</u>

[i] The U.S. Bureau of Labor Statistics website is found at https://www.bls.gov/

[ii] The Current Employment Statistics website is located at https://www.bls.gov/ces/ for national data and at https://www.bls.gov/sae/ for state and metropolitan area data.

[iii] Kenneth W. Robertson (2016). *Management of Business Surveys and Production of Economic Statistics; Perspectives from the U.S. Bureau of Labor Statistics' Current Employment Statistics Survey.* International Conference on Establishment Surveys V. http://ww2.amstat.org/meetings/ices/2016/proceedings/154_ices15Final00095.pdf

[iv] Kenneth W. Robertson and Julie Hatch-Maxfield (2012). Data Collection in the U.S. Bureau of Labor Statistics' Current Employment Statistics Survey. United Nations Economic Commission for Europe, Conference of European Statisticians, Seminar on New Frontiers for Statistical Data Collection.

http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2012/mtg2/WP20.pdf

[v] Nicholas A. Johnson (2016). *Maximizing Web Survey Response: Research from the Current Employment Statistics (CES) Survey*. International Conference on Establishment Surveys V. http://ww2.amstat.org/meetings/ices/2016/proceedings/209_ices15Final00214.pdf

[vi] The Job Openings and Labor Turnover website is at https://www.bls.gov/jlt/

[vii] Paul R. Calhoun and Laura W. Jackson (2008). Bureau of Labor Statistics Monthly Labor Review, *Organization and Management of a Data Collection Center*. https://www.bls.gov/ore/pdf/st010050.pdf

[viii] The Quarterly Census of Employment and Wages website is at https://www.bls.gov/cew/ and its related Business Employment Dynamics website is found at https://www.bls.gov/bdm/

[ix] Unemployment Insurance is explained on the Department of Labor's website at https://www.dol.gov/general/topic/unemployment-insurance

[x] Information on the Multiple Worksite Report is found at https://www.bls.gov/cew/cewmwr00.htm

[xi] Information about the Annual Refiling Survey can be found at https://www.bls.gov/respondents/ars/

[xii] Robertson, Kenneth W. and Eltinge, John L. (2013) Compilation and Management of Labor Market Statistics for Macroeconomic Analysis at the U.S. Bureau of Labor Statistics. Malaysia MyStats conference 2013. http://www.bnm.gov.my/documents/conference_vol/2013_MyStats/np/Paper_Session%202(A)_John%20Eltinge%20(BLS).pdf

[xiii] SAS is a high-end commercial-off-the-shelf statistical analysis software package for "Analytics Software & Solutions", see https://www.sas.com/en_us/home.html

[xiv] The spreadsheet-like web data collection tool is based on a tool called the "handsontable". See https://handsontable.com/

[xv] Sharon Stang and Emily Thomas (2016), *Web Collection in the Quarterly Census of Employment and Wages Program*, International Conference on Establishment Surveys V. http://ww2.amstat.org/meetings/ices/2016/proceedings/072_ices15Final00299.pdf

[xvi] Data on employment and wages by industry in U.S. non-profit businesses can be found at https://www.bls.gov/bdm/nonprofits/nonprofits.htm