Data Sharing Between the Bureau of Labor Statistics and the Census Bureau

BLS Data Synchronization Project

Sharon Stang
Economist
Division of Administrative Statistics/
Office of Employment and Unemployment Statistics

FedCASIC 2016
May 4, 2016
A long time ago, in a galaxy far, far away....
EPISODE 1

THE PHANTOM MENACE

Turmoil has engulfed the galactic statistical community. The disparities between NAICS codes on the Census Bureau and Bureau of Labor Statistics business registers is under intense scrutiny.

Hoping to resolve the matter peacefully, a workshop to discuss interagency data sharing was organized by the National Research Council on behalf of the Bureau of Economic Analysis.
History

- CIPSEA 2002
- Data sharing between BEA, BLS, and the Census Bureau (2005)
- Protect confidentiality and improve the quality of data and statistics
- Benefits agencies and data users
EPISODE II

ATTACK OF THE CLONES

While much time has passed, there is lingering unrest in the galactic statistical community. A legislative blockade has made it difficult for the individual agencies to share their administrative data sets for the common good of the Republic.

Seven years after their first meeting, the Census Bureau and the Bureau of Labor Statistics forge a pact to share their multi-unit data, and embark on a quest to rid the galaxy of NAICS coding discrepancies.
2012 - MOU Between BLS and Census

- Share certain business data of multi-location companies.
- Research and fix industry coding discrepancies
Data Sources

- Census Business Register (BR)
  - 2007/2012 Economic Censuses
  - Annual Company Organization Survey (COS)
  - Annual Survey of Manufacturers (ASM)
  - Service Annual Surveys (SAS)

- BLS Quarterly Census of Employment and Wages (QCEW)
  - 2007 Annual data and 2011/2012 Q1-Q2 data
  - Quarterly State Unemployment Insurance (UI) filings
  - Annual Refiling Survey (ARS)
Shared Data

■ Census

▶ Data free of Title 26 Restrictions
▶ For multi-location companies
  - Business name and address
  - State, county, and place geocodes
  - EIN
  - NAICS code
  - NAICS Source Code
  ❖ 2007 March Employment
  ❖ 2007 Annual Payroll

■ BLS

▶ Multi-unit company name and address
▶ Affiliated establishment information
  - Name and address
  - EIN
  - NAICS code
  - ARS year
  - Operational Status
  - Quarterly Employment
  - Quarterly and annual payroll

❖ 2012 Economic Census data available mid-2014.
BLS Data Synchronization Project

1. Bureau of Labor Statistics business register, based on UI records with other survey sources
2a. Match records
2b. Independently verify NAICS codes based on publicly available information and respondent contact
3. No IRS Data
   - Bureau of Labor Statistics and States’ business registers
   - Independently Verified NAICS Codes
   - Census Bureau business register

Matching Data Elements:
- Business name
- Trade name
- Mailing Address
- Physical location address
- Tax identification number
- Principal industry activity
- Number of employees
- Total wages
- Gross receipts
What are Multis?

- Census – Multiple establishments with the same federal Employer Identification Number (EIN).

- BLS – Multiple establishments with the same Unemployment Insurance Account Number (UIN) within a STATE.
QCEW/Census 2012* file matching

Matching Criteria: EIN, State FIPS, Company Name, Physical Location Address (PLA), and County/Township Code - after standardization.

- Worksites on the 2012 BR file: 1,915,448
- Number of matched worksites: 905,853

- Worksites on the 2012-Q1 QCEW file: 9,247,203
- Number of matched QCEW worksites: 884,597

- Total Matches: 857,743
- One to One matches: 817,209
- QCEW NAICS = Census NAICS 611,944 (75%)
- QCEW NAICS <> Census NAICS 205,265 (25%)

* 2012 Economic Census data made available in May 2014.
Exclusion Criteria

- Establishments with less than 50 in QCEW employment in March 2012
- Establishments with an absolute employment difference greater than 30% between the March 2012 Census employment and the March 2012 QCEW employment
- Establishments with no sector mismatch between Census and QCEW

- 7,790 in-scope establishments
  - 100+ employment, sector mismatch, and Census/QCEW Employment difference ≤ 30% --- 3,820 establishments
  - Employment between 50-99 --- 3,970 establishments (lower priority but in-scope)
Census 2007/2012 File Comparison

- **5,972** originally identified 2007 in-scope establishments

- **7,990** in-scope establishments in 2012
  - **3,256** (42%) were also in the original 2007 matched establishments.
  - **4,534** (58%) were newly identified as sector mismatches.
EPISODE III

REVENGE OF THE SITH

*Jedi Knights have been dispatched to all corners of the galaxy, to root out the causes of NAICS discrepancies in the data files.*

*Once identified, the brave Jedi will coordinate a two-pronged attack to resolve these conflicts and return data consistency to the galaxy.*
NAICS Adjudication Steps

- BLS regional office staff were provided with the QCEW establishment information only and asked to research, and contact the establishments as necessary, to either verify or update the QCEW NAICS codes.

- A second identical blind review of the establishments was conducted by a different regional office.

- Results of the two independent reviews were then evaluated.
NAICS Adjudication Results

- 1,466 establishments were adjudicated in 2014.

Results:

- 464 matched the QCEW NAICS Sector. **Candidates for Census NAICS change.**
- 247 matched the Census NAICS Sector. **Candidates for QCEW NAICS change.**
- 53 did not match the QCEW or Census NAICS Sectors. **Candidates for Census and QCEW NAICS Change.**
- 702 had different suggested NAICS sectors in both reviews.
What We learned

- Matching microdata records is hard
- Some discrepancies fix themselves over time
- Discrepancies exist because some establishments are just hard to code
- This work is time-consuming and resource intensive
- There will likely always be discrepancies given how the programs are structured
EPISODE IV

A NEW HOPE

Armed with extensive new information on the scope of the NAICS coding discrepancies across the galaxy, the Bureau of Labor Statistics has begun planning its counter-attack.

This macro-level attack heralds an awakening of the Force within the galaxy as Census and the Bureau of Labor Statistics work tirelessly toward the goal of consistent, protected, reconcilable business register data.
BLS Data Synchronization Project: Phase -2

- Take a macro-level approach
- Identify significant sub-sector level differences between the BR and QCEW
- Research why discrepancies exist
- Develop strategies to address discrepancies based on their root causes
<table>
<thead>
<tr>
<th>NAICS Sub-Sector</th>
<th>2012 CBP and QCEW Published Data</th>
<th>Establishments</th>
<th>Employment</th>
<th>Wages (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QCEW</td>
<td>CBP</td>
<td>Absolute Value Difference</td>
<td>% Diff</td>
</tr>
<tr>
<td>NAICS 561 Administrative and support services</td>
<td>449,229</td>
<td>361,733</td>
<td>87,496</td>
<td>19.48%</td>
</tr>
<tr>
<td>NAICS 611 Educational services</td>
<td>99,704</td>
<td>68,276</td>
<td>31,428</td>
<td>31.52%</td>
</tr>
<tr>
<td>NAICS 551 Management of companies and enterprises</td>
<td>55,835</td>
<td>52,380</td>
<td>3,455</td>
<td>6.19%</td>
</tr>
<tr>
<td>NAICS 622 Hospitals</td>
<td>8,214</td>
<td>6,546</td>
<td>1,668</td>
<td>20.31%</td>
</tr>
<tr>
<td>NAICS 423 Merchant wholesalers, durable goods</td>
<td>248,211</td>
<td>244,963</td>
<td>3,248</td>
<td>1.31%</td>
</tr>
<tr>
<td>NAICS 425 Electronic markets and agents and brokers</td>
<td>230,707</td>
<td>44,247</td>
<td>186,460</td>
<td>80.82%</td>
</tr>
<tr>
<td>NAICS 813 Membership associations and organizations</td>
<td>140,214</td>
<td>106,911</td>
<td>33,303</td>
<td>23.75%</td>
</tr>
<tr>
<td>NAICS 448 Clothing and clothing accessories stores</td>
<td>129,378</td>
<td>147,502</td>
<td>18,124</td>
<td>14.01%</td>
</tr>
<tr>
<td>NAICS 517 Telecommunications</td>
<td>43,507</td>
<td>50,576</td>
<td>7,069</td>
<td>16.25%</td>
</tr>
<tr>
<td>NAICS 713 Amusements, gambling, and recreation</td>
<td>71,692</td>
<td>71,562</td>
<td>130</td>
<td>0.18%</td>
</tr>
<tr>
<td>NAICS 334 Computer and electronic product manufacturing</td>
<td>18,654</td>
<td>13,218</td>
<td>5,436</td>
<td>29.14%</td>
</tr>
<tr>
<td>NAICS 518 Data processing, hosting and related services</td>
<td>14,501</td>
<td>14,319</td>
<td>182</td>
<td>1.26%</td>
</tr>
<tr>
<td>NAICS 454 Nonstore retailers</td>
<td>40,805</td>
<td>63,691</td>
<td>23,666</td>
<td>58.89%</td>
</tr>
<tr>
<td>NAICS 511 Publishing industries, except internet</td>
<td>34,599</td>
<td>26,886</td>
<td>7,713</td>
<td>22.29%</td>
</tr>
<tr>
<td>NAICS 213 Support activities for mining</td>
<td>16,386</td>
<td>14,080</td>
<td>2,306</td>
<td>14.07%</td>
</tr>
<tr>
<td>NAICS 523 Securities, commodity contracts, investments</td>
<td>94,116</td>
<td>98,109</td>
<td>3,993</td>
<td>4.24%</td>
</tr>
</tbody>
</table>
Contact Information

Sharon Stang
Economist
DASLT/OEUS
www.bls.gov/cew
202-691-5789
stang.sharon@bls.gov