



# A Complex Use Case - Documenting the Consumer Expenditure Survey at BLS

Daniel Gillman, Evan Hubener, Reginald Noel, Bryan Rigg, Arcenis Rojas, Lucilla Tan, Taylor Wilson

U.S. Bureau of Labor Statistics,

## ABSTRACT

The Consumer Expenditure Survey (CE) is a Bureau of Labor Statistics (BLS) program that measures how US families spend their money. These data are also input to the CPI. BLS selected DDI-3.2 to document CE, including the entire life-cycle.

CE is conducted as 2 separate surveys, Interview and Diary. The data are combined during processing and packaged in 2 ways, one for CE dissemination and the other for CPI. Changes in design occur every odd numbered year. Yearly estimates are created every 6 months, PUMS issued yearly, and data sent to CPI monthly. CE processing is divided into 4 phases: 1) - sample selection and collection; 2) - initial edit subsystem; 3) - estimation and edit subsystem, with data sent to CPI; and 4) - final edits, tables, microdata. Data are processed in packages by expenditure type.

A documentation system needs to handle all these features. For development, BLS is conducting a phased approach, adding complexity from phase to phase. The incremental systems are designed to establish that DDI and the Colectica system are sufficiently sophisticated to account for each feature of CE. This paper will go into detail about the particulars of the CE survey, describe progress made, and plans for the future.

## CONSUMER EXPENDITURE SURVEYS

Measures how US people and households spend money

Conducted by BLS; data collected by Census

Consists of 2 surveys

- Interview (Quarterly)
  - Conducted every month on rotating panel
  - Interview for past 3 months
  - Includes large or recurring expenses (e.g., rent)
- Diary (2 Week)
  - Distributed weekly
  - 2 consecutive one week diaries
  - Includes small, frequent expenses (e.g., groceries)

Processing

- 4 subsystems
  - Data Collection Subsystem
  - Initial Edit Subsystem
  - Edit and Estimation Subsystem
  - Dissemination Subsystem

## MOTIVATION

Complete survey processing redesign

Currently, variables managed through

- Independent MS Access databases
  - One per subsystem
  - Tracking variables across DBs is very hard

Single system for managing variables

- Across surveys (Interview and Diary)
- Throughout life-cycle
  - Including dissemination
- Over years

## GOALS

Want to produce a system to show:

- How similar variables change over time
  - Changes occur in odd years
- How similar variables change over life-cycle
  - Including code list differences
  - Groupings - expenditure and UCC
- How similar variables differ across surveys
  - Interview and Diary designs are
    - Similar in some ways
    - Different in others
- Entire life-cycle
  - Questions to final variables and tables
  - Include all production subsystems
- Full processing steps
  - Links to variables as inputs/outputs
  - Show flow through and between each subsystem
- Instrument design
  - Including wording and skip flow
  - Ability to input / output transferrable questionnaire

## PATH to SOLUTION

Need metadata system

- Selected Data Documentation Initiative (DDI)
  - Version 3.2 - Lifecycle
- Selected DDI commercial software
  - Colectica Designer and Repository / Portal

Develop system iteratively

Start small

- Series of pilot systems
- Two built so far
- Third being designed

## RESULTS

CE documentation needs

- Pilot 1
  - Showed DDI is sufficient to
    - Document CE data
    - Document CE processing
    - Show differences across time and surveys
- Pilot 2
  - Showed Colectica Repository/Portal can handle
    - Manage metadata across time
    - Display metadata in customizable ways
  - Details to follow

## PILOT 2 RESULTS

Time / resource limited

Keep system simple

Show 2012 and 2013 metadata

Illustrate with 2 variables

- Education
  - Hospitalization and Health Insurance
- EIHB expenditure set

## PILOT 2 RESULTS

Compare variables

Two main means

- Correspondence Tree
  - Across surveys
  - Over time
  - Through lifecycle
- Code Comparison
  - Observe changes to categorization
    - Substantive
    - Gratuitous

Education code comparison

- Highest grade completed
- Substantive differences
  - 2012
    - 21 categories
    - One for each grade up to 12
    - Nursery/kindergarten not available choices
  - 2013
    - 8 categories
    - Nursery/kindergarten/elementary versus high school
    - One category for professional/masters/PhD

## PILOT 2 RESULTS

- Gratuitous differences
  - Category labels change
    - Added blanks
    - “12<sup>th</sup>” versus “twelfth”
    - “professional school degree” versus “professional degree”
    - “high school (grades 9-12), no degree” versus “high school (grades 9-12, no degree)”
    - “high school graduate - high school diploma or the equivalent (GED)” versus “high school graduate - high school diploma or equivalent (GED)”

## CONTACT

Dan Gillman  
Information Scientist  
[www.bls.gov/osmr](http://www.bls.gov/osmr)  
202-691-7523  
[Gillman.Daniel@BLS.gov](mailto:Gillman.Daniel@BLS.gov)