

Using DDI and GSBPM Together at the US Bureau of Labor Statistics

Daniel Gillman

Information Scientist

US Bureau of Labor Statistics

Modern Stats World Conference

Thursday 12 April



Overview

- Consumer Expenditure Survey (CE)
- Quick DDI Review
- Usage of DDI for CE Metadata
- BLSBPM
- Usage of BLSBPM for CE Metadata



Consumer Expenditure Survey (CE)

- Measures how US people and households spend money
- Conducted by BLS
- Data collected by Census Bureau
- Tabular data reported every 6 months
 - ▶ Based on past 12 months
- Public Use Microdata released yearly
 - ▶ Microdata are person and household level data
- Also used as input to Consumer Price Index (CPI)



CE – Survey Design

■ Two surveys:

▶ Interview

- Respondents interviewed for 4 quarters (every 3 months)
- Questioned about previous 3 months
- Data collected every month through rotating sample
- Includes large or recurring expenses (e.g., cars, rent)

▶ Diary

- Diary issued (almost) weekly, sent to each respondent once
- Covers expenses for 2 consecutive one week periods
- Each week handled separately
- Includes small, frequent expenses (e.g., utility bill, groceries)

▶ Many expenses asked in both

CE – Questionnaires

■ Diary

- ▶ 2 week listing of purchases for small items
- ▶ <https://www.bls.gov/cex/csx801p.pdf>

■ Interview

- ▶ Complex detailed questionnaire
- ▶ More than an hour to complete
- ▶ <https://www.bls.gov/cex/capi/2015/cecapihome.htm>

CE – Processing

■ Production – 4 phases or subsystems

▶ 1) Data Collection

- Performed at US Census Bureau

▶ 2) Initial Edit

- Computational Edits, Recodes, Consistency Edits

▶ 3) Estimation Edit

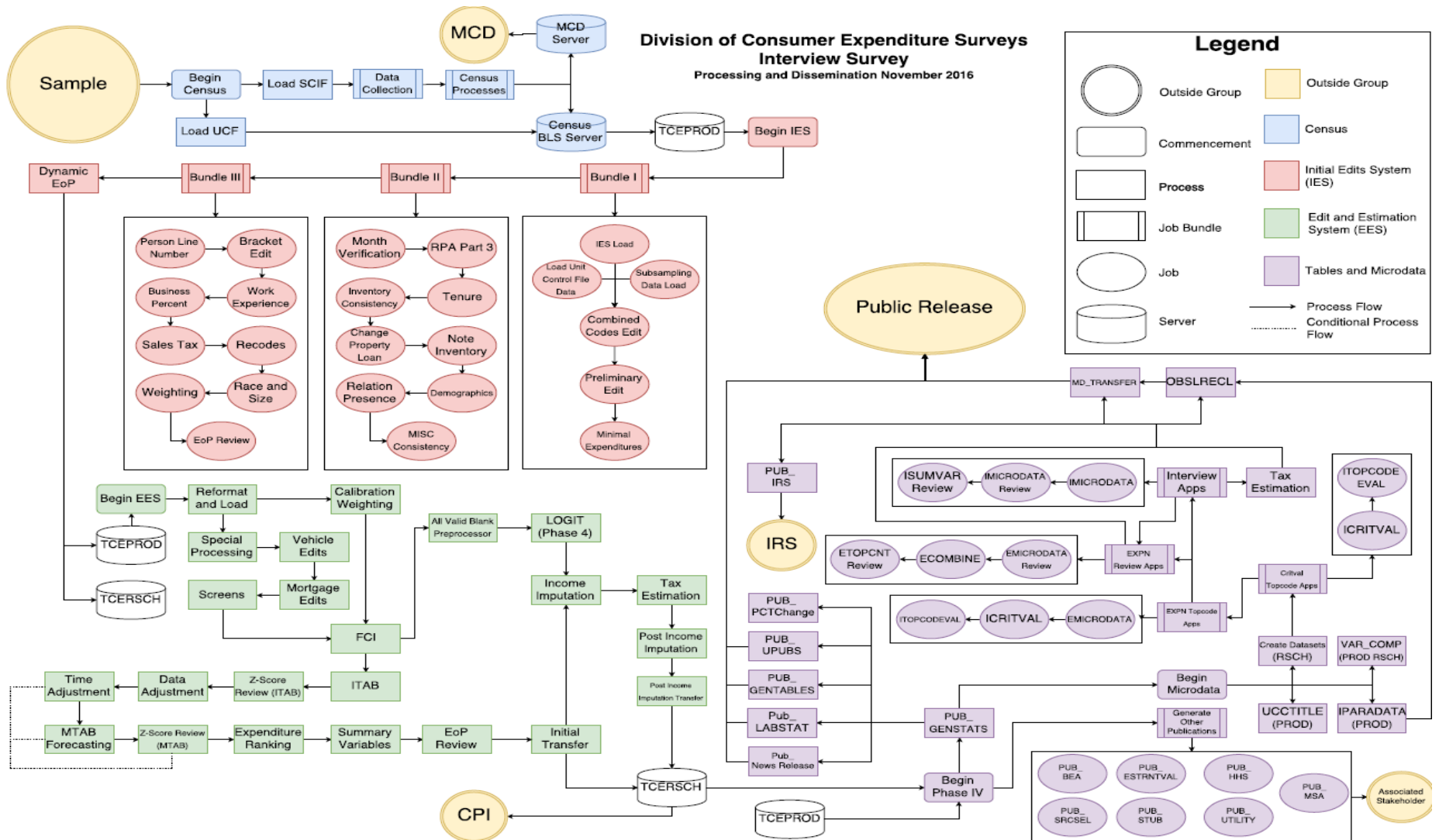
- Additional edits, adjustments, imputation, estimation

▶ 4) Dissemination

- Measures; Tables; Microdata; Disclosure Avoidance

■ CE Process Flow Diagram

CE – Processing



CE Documentation

■ Needs

- ▶ Complete survey processing
- ▶ Single system for managing variables
 - Through life-cycle
 - Across years
 - Across surveys (Interview and Diary)
 - Support processing
 - Support dissemination

■ Currently, variables managed through

- ▶ MS Access databases
- ▶ One per processing subsystem
 - With little coordination

Data Documentation Initiative

■ DDI

▶ Developed under DDI Alliance

- Secretariat – ICPSR
- ~ 50 members

■ 2 main standards

▶ DDI Codebook 2.5

- Single data sets or studies

▶ DDI Lifecycle 3.2

- Survey life cycle <- Selected by CE project

▶ Both based on XML

Data Documentation Initiative

- Implementation details
 - ▶ Selected Colectica suite of tools
 - Designer
 - Repository
 - Portal
 - Questionnaire
- Tools shift burden
 - ▶ From needing to know XML
 - ▶ To content



DDI Implementation

- Examples
 - ▶ Processing
 - ▶ Variables
 - ▶ Lineage
 - ▶ Code lists

CE Process Flow

-  [Common Metadata](#)
-  [BLS CE Instruments](#)

- Flowchart
- Details**

Start

Census

IES

Bundle 1

Bundle 2

Bundle 3

Dynamic EoP Dynamic end of processing review

EES


Tables and Microdata

Label Highest Grade Completed?


Statistics

Code Comparison


Correspondence Tree

 EDUCA


Highest Grade Completed?

 EDUCA


Highest Grade Completed?

 2012 EES Subsystem -
MEMB.EDUCA


Highest Grade Completed?

 2012 IES Subsystem -
MEMB.EDUCA


Highest Grade Completed?

 EDUCA

Highest Grade Completed?

 2013 IES Subsystem -
MEMB.EDUCA

What is the highest level of education that member has/you have completed?

 2013 EES Subsystem -
MEMB.EDUCA

What is the highest level of education that member has/you have completed?

Highest Grade Completed?

Interview
 2012
 memi131 **6**

- Variable
- Details
- Lineage**
- Concordance Variables

memi123 - EDUCA
What is the highest level of school the member has completed or the highest degree the member has received?

2012 EES Subsystem - MEMB.EDUCA
Highest Grade Completed?

2012 IES Subsystem - MEMB.EDUCA
Highest Grade Completed?

educa2012
WHAT IS THE HIGHEST LEVEL OF SCHOOL (NAME) HAS/YOU HAVE COMPLETED OR THE HIGHEST DEG

memi131 - EDUCA
What is the highest level of school the member has completed or the highest degree the member has received?

2012 EES Subsystem - MEMB.EDUCA
Highest Grade Completed?

BLSBPM

- Based GSBPM



Result: GSBPM Version 5.0

❖ Released December 2013

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

BLSBPM

■ GSBPM

- ▶ 8 phases

■ BLSBPM

- ▶ 10 phases

- ▶ Specify Needs

- ▶ Design Survey

- ▶ Construct Frame

- ▶ Construct Sample

- ▶ Collect Data

- ▶ Review and Edit Collected Data

- ▶ Calculate Estimates

- ▶ Analyze Estimates

- ▶ Disseminate Data

- ▶ Archive



BLSBPM

- Several phases specify parallel sub-processes
 - ▶ Construct sample
 - ▶ Collect data
 - ▶ Analyze estimates
- Model used to
 - ▶ Classify development projects
 - ▶ Classify existing systems
 - ▶ Aid management and reduce redundancy

Used Together

- CE Management – 2 independent staffs
 - Subject matter
 - IT support and development
- Need coordinated documentation
- DDI provides needed specificity
 - ▶ Limited process model in Lifecycle 3.2
 - ▶ However, can be used hierarchically
 - ▶ Ability to describe processing at all levels

Used Together

- Map development to DDI documentation
 - ▶ CE Process diagram provides slots
 - ▶ BLSBPM provides classification of those slots
 - ▶ DDI documentation serves as IT documentation as well
 - Helps IT management
 - Even though system is centered around CE
- DDI implementation
 - ▶ Under development
 - ▶ Iterative process

Conclusion

- BLS using BLSBPM
 - ▶ Usage expansion planned
- CE using DDI and Colectica software
 - ▶ Usage expansion begun
 - ▶ More detail as time goes on
 - ▶ Integrated process documentation the goal
 - With all variables
 - Through statistical life cycle

Contact Information

Daniel Gillman

Office of Survey Methods Research

US Bureau of Labor Statistics

www.bls.gov/OSMR

+1.202.691.7523

Gillman.Daniel@BLS.Gov

