

# NADDI 2018

## Implementing DDI to Document the Consumer Expenditure Surveys

Reginald Noël  
Parvati Krishnamurty

5 April 2018



# The Consumer Expenditure Surveys

National surveys on U.S. households' spending

- Survey sponsor & data processing: **BLS**
- Data Collection: **Census Bureau**
- Two independent survey instruments: **Diary, Interview Survey**
- Data Processing: **BLS – across 3 subsystems**
  - Over 80 data processing jobs/edits for each survey of each year
  - Over 3,400 variables for each survey of each year
- Data Products release schedule: **BLS**
  - Tabular data (**rolling 4 quarters of data every 6 months**)
  - Public Use Microdata (**annually**)
  - Input to CPI (**quarterly**)





# This Presentation

The CE program is attempting to move towards a more efficient and comprehensive documentation of survey metadata. We provide 2 illustrations of current inefficiencies, and how we are attempting to use DDI to address them:

1. Questionnaire development
2. Investigating data inconsistencies



# 1. Questionnaire Development



# CE Questionnaire Development Process

- The Interview Survey is long and contains complex skip patterns designed to reduce respondent burden.
- Changes to questionnaire items such as wording, order, skip patterns, or new questions are often needed.
- CE provides detailed requirements, which Census then enters into their own specification system (SPIDER), and then programs the Blaise instrument. Changes are usually made every 2 years.
- CE Interview Survey instrument is being redesigned.



# Interview Survey Questionnaire Development

## Challenges

- No question bank
- No integrated mapping of questionnaire metadata
  - ▶ Multiple documents in different locations
- The changes document for each revision is in word/pdf/csv
  - ▶ Difficult to navigate: current full specifications ~ 4,000 pages
  - ▶ Iteratively making design changes is difficult

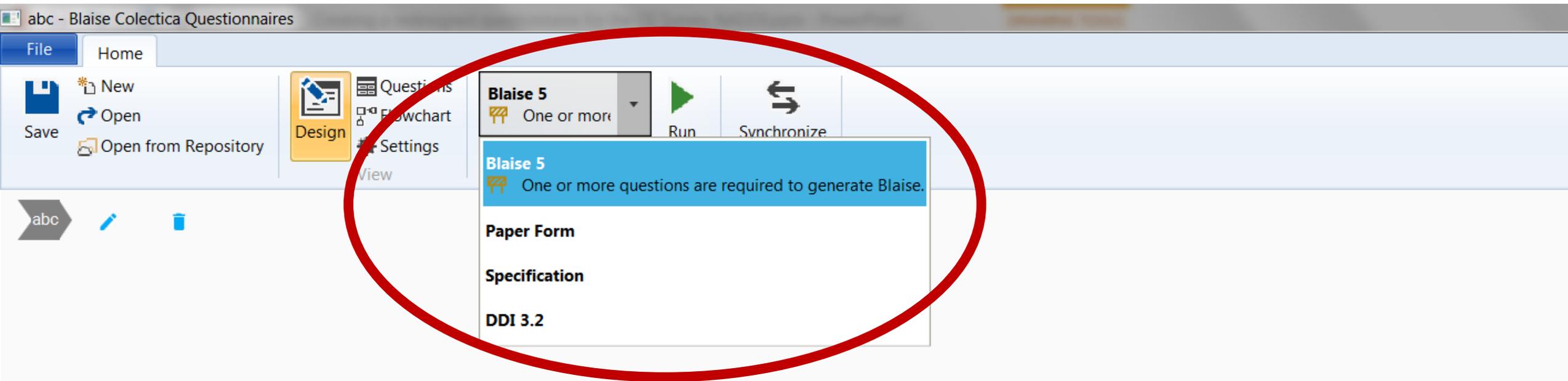
**Solution:** Testing the use of Colectica Questionnaires to create requirements for the new CAPI instrument, to be programmed into a Blaise instrument by Census Bureau.

# Colectica Questionnaires: Advantages

- Creates DDI based survey specifications
- Surveys can be published in multiple formats including Blaise 5
- Produces machine readable DDI 3.2 format xml files (source code for CAI systems)
- Rapid design iterations: Easy to make changes in the software and then generate new specifications
- Uses a repository of questions so we can retrieve questions and blocks
- Integrates with the rest of the Colectica software suite including Repository, Portal, and Designer



# Some advantages: Output formats



Add the first item

# Some advantages: Repositories

The screenshot shows the Blaise 5 software interface. The title bar reads 'abc - Blaise Colectica Questionnaires'. The ribbon includes 'File' and 'Home' tabs. The 'Home' ribbon contains several groups of icons: 'Save' (Save, Open, Open from Repository), 'Design' (Design, Questions, Flowchart, Settings), 'Publish' (Blaise 5, One or more, Run), and 'Repository' (Synchronize). Below the ribbon is a toolbar with 'abc', a pencil icon, and a trash icon. The main workspace is mostly empty, with a blue button labeled 'Add the first item' in the center. On the right side, there is a 'Add' panel with four sections: 'Static Content' (Descriptive Text), 'Standard Questions' (Multiple Choice, Text Entry, Numeric Entry, Date, Time), 'Structure' (Sequence, Edit Check), and 'Repository' (Question, Sequence). The 'Repository' section is circled in red. A green checkmark icon is visible in the bottom right corner.

Category	Item 1	Item 2
Static Content	Descriptive Text	
Standard Questions	Multiple Choice	Text Entry
	Numeric Entry	Date
	Time	
Structure	Sequence	Edit Check
Repository	Question	Sequence

# Colectica Questionnaire: Limitations

- Trial version cannot handle some of the more complex questionnaire design elements.
- Some examples are: grid questions, reusable lists of response options, globally adding fields for specifications, setting don't know and refusals globally, ability to store and cumulate display logic, rosters and loops, adding comments to questions.
- We got around the limitations by programming what we could in the software and providing a cover sheet for each section with additional information that could not be entered in the software.  
e.g. text fills, refusal/don't know options, instructions for looping, and grid questions.

# Cluster screener (grid question) in the CE

42 ? [F1]

Since the first of November, have you had expenses for any of the following, either for you or for someone outside your household?

♦ Read each item on list.

- 1. Fresh flowers or potted plants?
- 2. Professional photography?
- 3. Services of lawyers or other legal professionals?
- 4. Accounting fees?
- 5. Occupational expenses, such as union dues or professional licenses?
- 6. Gardening or lawn care services?
- 7. Housekeeping services?
- 8. Home security system service fees?
- 9. Other home services or small repair jobs around the house, not previously reported?
- 10. Moving, storage, or freight?
- 11. Stamp or coin collecting?
- 12. Lotteries or games of chance?
- 95. Continue list

Section 19, Part A - Miscellaneous Expenses

Item	Misc exp/Item	Description	Month	Who	Amount	C	Combined Codes	Store	Online	Location	Any more
6	Garden/lawn serv	Garden/lawn serv	1	1	\$500						2
7	Housekeep serv	Housekeep serv	2	1	\$300						2
7	Housekeep serv	Housekeep serv	1	1	\$200						2
8	Home security	Home security	2	1	\$50						2

# Grid question in Colectica Questionnaires

The screenshot displays the Colectica Questionnaires software interface. At the top, there is a ribbon menu with tabs for 'File' and 'Home'. The 'Home' tab is active, showing several groups of icons: 'Save' (Save, Open, Open from Repository), 'Design' (Design, Questions, Flowchart, Settings), 'Publish' (Blaise 5 dropdown, Run), and 'Repository' (Synchronize). Below the ribbon, there is a 'Household operations' section with a pencil icon and a trash icon. The main content area shows a grid question setup. The first row is titled 'OPSSEC' and contains the text: 'Now I will ask you about expenditures for household maintenance, repairs and service contracts. Please remember to include any payments you made online or had automatically deducted. Also, please include any shipping and handling charges with the cost of any item that was shipped.' The second row is titled 'OPSCSC' and contains the text: 'During [Month1], [Month 2], or [Month 3], did (you/you and any members of your household) have any expenses for--' followed by a list of six items: 1 Gardening or lawn care services?, 2 Housekeeping services?, 3 Termite or pest control services?, 4 Water softening services?, 5 Home security system service fees?, and 6 Moving, storage or freight?. Below this, there is a green bar with the text: 'Follow up questions Display if OPSCSC Contains Any 1 Gardening or lawncare services? or 2 Housekeeping services? or 3 Termite or pest control services? or 4 Water softening services? or 5 Home security system service fees? or 6 Moving, storage or fr'. The final row is titled 'Conditional' and contains a dropdown arrow pointing down, the text 'OPSCSCFL', and '12 items'. A green plus sign is visible in the bottom right corner.

File Home

Save New Open Open from Repository Design Questions Flowchart Settings View Publish Run Synchronize Repository

Household operations

OPSSEC

Now I will ask you about expenditures for household maintenance, repairs and service contracts. Please remember to include any payments you made online or had automatically deducted. Also, please include any shipping and handling charges with the cost of any item that was shipped.

OPSCSC

During [Month1], [Month 2], or [Month 3], did (you/you and any members of your household) have any expenses for--

- 1 Gardening or lawn care services?
- 2 Housekeeping services?
- 3 Termite or pest control services?
- 4 Water softening services?
- 5 Home security system service fees?
- 6 Moving, storage or freight?

Follow up questions  
Display if OPSCSC Contains Any 1 Gardening or lawncare services? or 2 Housekeeping services? or 3 Termite or pest control services? or 4 Water softening services? or 5 Home security system service fees? or 6 Moving, storage or fr

Conditional

OPSCSCFL  
12 items

## 2. Investigating Data Inconsistencies



# Fragmented metadata across time and processes

## Challenges

- Maintained separately
  - ▶ by surveys (Diary or Interview)
  - ▶ by years (... 2015, 2016, 2017 ...)
  - ▶ by 3 data processing subsystems
  - ▶ Initial Edit System (IES), Edit and Estimation System (EES), Microdata, and Tables
- Multiple points of manual input and alteration
- Many different folders, different file formats, separate database

**Solution:** Testing Colectica Designer, Repository, and Portal to put all metadata across time and subsystem onto one web-based portal

# 2. Investigating Data Inconsistencies

**Inquiry Request:** Research Universal Classification Code 580311

- What variables are used to map to this UCC?
- How is the associated question asked in the survey?
- Get descriptive statistics, i.e. average, frequencies, etc.



# 2. Investigating Data Inconsistencies

## ■ Current Production Method

- ▶ 1. Excel mapping file
- ▶ 2. Web-based P&C data dictionary
- ▶ 3. PDF public-use microdata dictionary
- ▶ 4. BLS website to get data
- ▶ 5. SAS enterprise to get stats

## ■ Using Colectica and DDI

- ▶ Colectica Repository and Portal through web-based prototype

# File 1: Excel mapping file for UCC

Excel interface showing a mapping file for UCC. The file name is `I_TimeAdjustmentParseFile_2013.xlsx`. The user is `Noel, Reginald - BLS`.

The spreadsheet displays columns A through S. The data is organized into rows, with a header row (row 3) and data rows (rows 4-1355). The columns are labeled as follows:

- A: comments line for columns
- B: GIFT
- E: NAME\_IN
- F: REC\_SEC
- I: SOURCE
- J: TIME\_ADJ
- K: UCC\_NO
- L: STARTDTE
- N: COMMENT
- O: IDCNT
- P: SEQCNT
- Q: ENDDTE
- R: TRANSDTE
- S: LASTUPDT

The data rows contain various values, including codes, dates, and counts. A red circle highlights the value `580311` in column K, row 1325. A red horizontal line is drawn across the entire row 1325.

	A	B	E	F	I	J	K	L	N	O	P	Q	R	S
2	comments line for columns													
3	CONDSPEC	GIFT	NAME_IN	REC_SEC	SOURCE	TIME_ADJ	UCC_NO	STARTDTE	COMMENT	IDCNT	SEQCNT	ENDDTE	TRANSDTE	LASTUPDT
1301	FURNPURY="171"	FURNGFTC="2	FURNPURX	FRAS08A	S08A	FURNMO	320221	20132	0	69	1			19123
1303		NA	QREP3MCX	FRBS08B	Recodes	&B	340630	20012	80	2	1			15063
1304		NA	QFRT3MCX	FRBS08B	Recodes	&A	340904	20012	80	3	1			15063
1314	HHICODE="2" & HHBCBS="1" & HHIFEET="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580112	19961	80	10	1	20131		19123
1316	HHICODE="2" & HHBCBS="1" & HHIFEET="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580114	19961	80	12	1	20131		19123
1318	HHICODE="1" & HHBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580312	19961	80	14	1			15063
1320	HHICODE="3" & HHBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580904	19961	80	16	1			15063
1322	HHICODE="4" & HHBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580906	19961	80	18	1			15063
1323	HHICODE="2" & HHBCBS="2" & HHIFEET="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580111	20032	80	19	1	20131		19123
1324	HHICODE="2" & HHBCBS="2" & HHIFEET="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580112	20032	80	20	1	20131		19123
1325	HHICODE="1" & HHBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580311	20032	80	21	1			16012
1326	HHICODE="3" & HHBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580905	20032	80	22	1			16012
1327	HHICODE="4" & HHBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580905	20032	80	23	1			16012
1328	HHICODE="2" & HHBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580115	20132	80	24	1			19123
1329	HHICODE="2" & HHBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580116	20132	80	25	1			19123
1333		NA	QCUMED1X	IHCS14C	Recodes	&C3	580901	19933	80	4	1			15063
1334		NA	QCUMED2X	IHCS14C	Recodes	&C2	580901	19933	80	5	1			15063
1335		NA	QCUMED3X	IHCS14C	Recodes	&C1	580901	19933	80	6	1			15063
1336		NA	QDRGPM1X	IHDS14D	Recodes	&C3	580907	20062	80	1	1			16890
1337		NA	QDRGPM2X	IHDS14D	Recodes	&C2	580907	20062	80	2	1			16890
1338		NA	QDRGPM3X	IHDS14D	Recodes	&C1	580907	20062	80	3	1			16890
1339	POLICYBYB="100"	NA	QLIFCMX	INBS13B	Recodes	&B	700110	19821	80	1	1			15063
1343	POLICYBYB="300" & INSPTYPE="1"	NA	QPR3MCMX	INBS13B	Recodes	&B	220121	19801	0	5	1			15063
1344	POLICYBYB="300" & INSPTYPE="2"	NA	QPR3MCMX	INBS13B	Recodes	&B	220122	19801	0	6	1			15063
1346	POLICYBYB="300" & INSPTYPE="3"	NA	QPR3MCMX	INBS13B	Recodes	&B	790600	19801	0	8	1	20131		19123
1348	POLICYBYB="400"	NA	QTN3MCMX	INBS13B	Recodes	&B	350110	19821	0	10	1			15063
1350	POLICYBYB="200"	NA	QVH3MCMX	INBS13B	Recodes	&B	500110	19821	0	12	1			15063
1351	POLICYBYB="500"	NA	QOTHCMX	INBS13B	Recodes	&B	002120	19992	80	13	1			15063
1352	POLICYBYB="110"	NA	QLNGTCMX	INBS13B	Recodes	&B	580400	20052	80	14	1			16602
1353	POLICYBYB="300" & INSPTYPE="3"	NA	QPR3MCMX	INBS13B	Recodes	&B	790600	20132	0	15	1			19123
1354	LSDCODE="100"	NA	QADDOWNX	LSDS10	Recodes	PMTMONTH	450313	19912	80	2	1	20131		19123
1355	LSDCODE="110"	NA	QADDOWNX	LSDS10	Recodes	PMTMONTH	450313	19912	80	2	1	20131		19123

At the bottom of the spreadsheet, there are status indicators: `UCC Tracker Updated` and `PARSE Updated`. The status bar at the very bottom shows `READY 1594 OF 2827 RECORDS FOUND` and `COUNT: 18`.

# File 1: Excel mapping file for UCC

2013 Interview Time Adjustment Expenditure Mapping file – (Excel file located in year-designated folder)

The screenshot shows an Excel spreadsheet titled 'L\_TimeAdjustmentParseFile\_2013.xlsx'. The data table has the following columns: A (comments line for columns), B (GIFT), E (NAME\_IN), F (REC\_SEC), I (SOURCE), J (TIME\_ADJ), K (UCC\_NO), and L (STARTDTE). Row 1325 is highlighted in red, and the cells 'HHICBCBS="2"' and 'QHI3MCX' are circled in red.

	A	B	E	F	I	J	K	L
2	comments line for columns	xx	xx	xx	xx	xx	xx	xx
3	CONDSPEC	GIFT	NAME_IN	REC_SEC	SOURCE	TIME_ADJ	UCC_NO	STARTDTE
1301	FURNPURY="171"	FURNGFTC="2"	FURNPURX	FRAS08A	S08A	FURNMO	320221	20132
1303		NA	QREP3MCX	FRBS08B	Recodes	&B	340630	20012
1304		NA	QFRT3MCX	FRBS08B	Recodes	&A	340904	20012
1314	HHICODE="2" & HHIBCBS="1" & HHIFEET="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580112	19961
1316	HHICODE="2" & HHIBCBS="1" & HHIFEET="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580114	19961
1318	HHICODE="1" & HHIBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580312	19961
1320	HHICODE="3" & HHIBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580904	19961
1322	HHICODE="4" & HHIBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580906	19961
1323	HHICODE="2" & HHIBCBS="2" & HHIFEET="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580111	20032
1324	HHICODE="2" & HHIBCBS="1" & HHIFEET="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580113	20032
1325	HHICODE="1" & HHIBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580311	20032
1326	HHICODE="3" & HHIBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580903	20032
1327	HHICODE="4" & HHIBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580905	20032
1328	HHICODE="2" & HHIBCBS="2"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580115	20132
1329	HHICODE="2" & HHIBCBS="1"	HHICOVQ="0"	QHI3MCX	IHBS14B	Recodes	&A	580116	20132
1333		NA	QCUMED1X	IHCS14C	Recodes	&C3	580901	19933
1334		NA	QCUMED2X	IHCS14C	Recodes	&C2	580901	19933
1335		NA	QCUMED3X	IHCS14C	Recodes	&C1	580901	19933
1336		NA	QDRGPM1X	IHDS14D	Recodes	&C3	580907	20062
1337		NA	QDRGPM2X	IHDS14D	Recodes	&C2	580907	20062



# File 2. Web-based P&C Data Dictionary

## Dictionary Search Page

---

### Search by Variable Information

Survey:

Dictionary:

RELN:

Field Name:

Description:

Phase:

# File 2. Web-based P&C Data Dictionary

## Variable QHI3MCX

---

**RELN:** EHB

**Survey Section:** 14B Hospitalization and Health Insurance

**Variable Description:**

Health insurance payments, reference period minus current month

**Recode:**

-

**Code Description:**

**Flag Name:** QHI3MCX\_

**Occurs:** 1

**Source:** Recodes

**Variable Phase:** 2n3

**P2 Attribute:** NUM8

**P3 Attribute:** NUM8

**P2 Primary Keys:** (FAMID,SEQNO)

**P3 Primary Keys:** (FAMID,SEQNO,ALCNO)

**Information queried from IY2013syb\_v26\_41.mdb**

[Back to Search Page](#)

[CE Web](#)

[DCES Home Page](#)

[Search Help](#)

# File 2. Web-based P&C Data Dictionary

## Variable HHICODE

**RELN:** EIHB

**Survey Section:** 14B Hospitalization and Health Insurance

**Variable Description:**

Insurance type code

**Recode:**

**Code Description:**

- 1 Health Maintenance
- 2 Fee for service plan
- 3 Commercial Medicare Supplement
- 4 Other Special purpose plan

**Flag Name:** HHICODE\_

**Occurs:** 1

**Source:** S14B HHICODE

**Variable Phase:** 2n3

**P2 Attribute:** CODED1\_R

**P3 Attribute:** CODED1\_R

**P2 Primary Keys:** (FAMID,SEQNO)

**P3 Primary Keys:** (FAMID,SEQNO,ALCNO)

Information queried from IY2013syb\_v26\_41.mdb

# File 2. Web-based P&C Data Dictionary

## Variable HHIBCBS

**RELN:** EHB

**Survey Section:** 14B Hospitalization and Health Insurance

**Variable Description:**

Is the insurance company Blue Cross/Blue Shield?

**Recode:**

**Code Description:**

1 Yes  
2 No

**Flag Name:** HHIBCBS\_

**Occurs:** 1

**Source:** S14B HHIBCBS

**Variable Phase:** 2n3

**P2 Attribute:** CODED1

**P3 Attribute:** CODED1

**P2 Primary Keys:** (FAMID,SEQNO)

**P3 Primary Keys:** (FAMID,SEQNO,ALCNO)

Information queried from IY2013syb\_v26\_41.mdb

[Back to Search Page](#)

[CE Web](#)

[DCES Home Page](#)

[Search Help](#)

# File 3. PDF Microdata Dictionary

## IHB – Detailed Expenditures Files (EXPB)

### *Hospitalization and Health Insurance*

#### 14 B Detailed Questions

VARIABLE NAME	DESCRIPTION	FLAG	FORMAT
---------------	-------------	------	--------

HHIRMPD	What period of time is covered by the regular payment? (HHIRMPD = 1 or 2 only) CODED 1 Week 2 2 weeks 3 Month 4 Quarter 5 6 months 6 Year 7 Other  CAPI Section 14	HHIR_MPD	CHAR(1)
---------	--	----------	---------

QH3MCX	Amount paid for health insurance premiums during the reference period  BLS derived	QH3MCX_	NUM(8)
--------	--	---------	--------

HHBCBS	Is insurance company Blue Cross/Blue Shield? CODED 1 Yes 2 No  CAPI Section 14	HHBCBS_	CHAR(1)
--------	---	---------	---------

HHICODE	What type of insurance plan is it? CODED 1 Health maintenance organization 2 Fee for service plan 3 Commercial Medicare supplement 4 Other special purpose plan  CAPI Section 14	HHICODE_	CHAR(1)
---------	---	----------	---------

HHIPOS	Under normal circumstances, if you go to a doctor who is not part of your plan without a referral, will your insurance pay for the cost? (HHICODE = 1 only) CODED 1 Yes 2 No  CAPI Section 14	HHIPOS_	CHAR(1)
--------	--	---------	---------

# File 4. BLS website to download dataset

## Consumer Expenditure Survey

SHARE ON: [f](#) [t](#) [in](#) | CE | FONT SIZE:

- BROWSE CE
- CE HOME
- CE OVERVIEW
- CE NEWS RELEASES
- CE DATABASES
- CE TABLES
- CE PUBLIC USE MICRODATA
- CE SURVEY MATERIALS
- CE PUBLICATIONS
- CE FAQs
- CE EVENTS IN YOUR AREA
- CONTACT CE

- SEARCH CE
- CE TOPICS
- PUBLICATIONS
- GEOGRAPHY
- GLOSSARY OF TERMS
- INFORMATION FOR CE RESPONDENTS
- POVERTY RESEARCH
- DATA COMPARISONS
- CE REDESIGN
- METHODOLOGY

### PUMD Data Files

From 1996 forward, public-use microdata (PUMD) are available in four formats: SAS, SPSS, STATA and Comma Delimited (ASCII). Each format presents the data for each year in two zipped files. Data collected with the Interview Survey are in the "Interview" file and those with the Diary Survey are in the "Diary" file.

Pre-1996 PUMD are available as Comma Delimited files. The [Inter-university Consortium for Political and Social Research](#) (ICPSR) provides the data for free to its members. Here is their [member list](#). BLS also sells the data on [USB flash drives \(PDF\)](#).

If you are new to CE PUMD data, you may want to explore the data availability with the [Getting Started Guide](#). For those using income data in their analyses, we also recommend reading the [User's Guide to Income Imputation in the CE](#). Additionally, we supply training programs and sample data ([ZIP](#)).

The **NEW** [State Weights for New Jersey](#) are available with their [documentation](#).

The next dataset will be released on September 11, 2018. To be notified when new datasets become available, please sign up for the [CEX e-mail updates](#). If you have comments, suggestions, and questions about this page and its contents, please [contact us](#).

If you are having difficulty downloading the data, please see the [instructions on how to download zip files](#).

	SAS	SPSS	STATA	CSV
<b>SAS</b>				
2016	<a href="#">Interview (zip)</a> (91.6 MB)   <a href="#">Diary (zip)</a> (10.5 MB)   <a href="#">Documentation</a>			
2015	<a href="#">Interview (zip)</a> (87.7 MB)   <a href="#">Diary (zip)</a> (10.5 MB)   <a href="#">Documentation</a>			
2014	<a href="#">Interview (zip)</a> (99.1 MB)   <a href="#">Diary (zip)</a> (12.3 MB)   <a href="#">Documentation</a>			
2013	<a href="#">Interview (zip)</a> (101.1 MB)   <a href="#">Diary (zip)</a> (11.5 MB)   <a href="#">Documentation</a>			

# File 4. BLS website to download dataset

Extract all files

Name	Type	Compressed size	Password p...	Size	Ratio	Date modified
 apb13.sas7bdat	SAS Data Set	331 KB	No	3,456 KB	91%	9/2/2016 12:06 PM
 cla13.sas7bdat	SAS Data Set	1,293 KB	No	11,584 KB	89%	9/2/2016 12:06 PM
 cld13.sas7bdat	SAS Data Set	25 KB	No	256 KB	91%	9/2/2016 12:06 PM
 cnt13.sas7bdat	SAS Data Set	251 KB	No	1,600 KB	85%	9/2/2016 12:06 PM
 cra13.sas7bdat	SAS Data Set	18 KB	No	192 KB	91%	9/2/2016 12:06 PM
 crb13.sas7bdat	SAS Data Set	205 KB	No	2,752 KB	93%	9/2/2016 12:06 PM
 eda13.sas7bdat	SAS Data Set	135 KB	No	1,280 KB	90%	9/2/2016 12:06 PM
 eqb13.sas7bdat	SAS Data Set	55 KB	No	512 KB	90%	9/2/2016 12:06 PM
 fra13.sas7bdat	SAS Data Set	265 KB	No	2,304 KB	89%	9/2/2016 12:06 PM
 frb13.sas7bdat	SAS Data Set	10 KB	No	128 KB	93%	9/2/2016 12:06 PM
 hel13.sas7bdat	SAS Data Set	32 KB	No	256 KB	88%	9/2/2016 12:06 PM
 <b>ihb13.sas7bdat</b>	SAS Data Set	409 KB	No	2,624 KB	85%	9/2/2016 12:06 PM
 ihc13.sas7bdat	SAS Data Set	166 KB	No	1,280 KB	88%	9/2/2016 12:06 PM
 ihd13.sas7bdat	SAS Data Set	97 KB	No	704 KB	87%	9/2/2016 12:06 PM
 inb13.sas7bdat	SAS Data Set	965 KB	No	12,672 KB	93%	9/2/2016 12:06 PM
 lsd13.sas7bdat	SAS Data Set	50 KB	No	384 KB	88%	9/2/2016 12:06 PM
 mdb13.sas7bdat	SAS Data Set	489 KB	No	4,672 KB	90%	9/2/2016 12:06 PM
 mdc13.sas7bdat	SAS Data Set	18 KB	No	192 KB	92%	9/2/2016 12:06 PM
 mis13.sas7bdat	SAS Data Set	477 KB	No	4,480 KB	90%	9/2/2016 12:06 PM
 mor13.sas7bdat	SAS Data Set	705 KB	No	3,264 KB	79%	9/2/2016 12:06 PM
 opb13.sas7bdat	SAS Data Set	430 KB	No	3,648 KB	89%	9/2/2016 12:06 PM
 opd13.sas7bdat	SAS Data Set	10 KB	No	128 KB	93%	9/2/2016 12:06 PM
 oph13.sas7bdat	SAS Data Set	25 KB	No	192 KB	87%	9/2/2016 12:06 PM
 opi13.sas7bdat	SAS Data Set	889 KB	No	13,312 KB	94%	9/2/2016 12:06 PM
 ovb13.sas7bdat	SAS Data Set	1,972 KB	No	18,304 KB	90%	9/2/2016 12:06 PM
 ovc13.sas7bdat	SAS Data Set	29 KB	No	256 KB	89%	9/2/2016 12:06 PM

# File 5. SAS to get dataset descriptive statistics

The screenshot displays the SAS Enterprise Guide interface. The 'Tasks' menu is open, with 'Describe' selected, leading to a sub-menu where 'Summary Statistics...' is highlighted. The background shows a data table with columns: IHIR\_MPD, QHI3MCX, QHI3MCX\_, HHIBCBS, HHIBCBS\_, and HHICODE. The 'QHI3MCX' column is highlighted in blue.

	IHIR_MPD	QHI3MCX	QHI3MCX_	HHIBCBS	HHIBCBS_	HHICODE
1		0	D	1	D	2
2		0	D	2	D	2
3		543	D	1	D	1
4		80	D	1	D	1
5		80	D	1	D	1
6		.	A	2	D	1
7		.	A	1	D	4
8		.	A	1	D	1
9		.	A	2	D	2
10		158	D	2	D	1
11		107	D	1	D	2
12		481	F	1	D	1
13		.	A	1	D	3
14		400	D	2	D	1
15	2	600	D	2	D	4
16	1	800	D	2	D	1
17	2	1041	D	2	D	2
18	2	.	A	2	D	3
19		.	A	2	D	4
20	2	3205	D	2	D	2
21	2	0	D	2	D	1
22	1	270	D	2	D	1
23	2	740	D	2	D	3
24	2	1131	D	2	D	3
25	2	1020	D	2	D	4
26	1	0	D	2	D	1
27		.	A	2	D	1
28	1	312	D	1	D	1
29	1	593	D	2	D	1
30	1	44	D	2	D	4
31	1	174	D	2	D	4
32	1	46	D	2	D	4
33	1	350	F	1	D	2
34	1	96	D	2	D	4

# 1. Colectica Portal: Prototype Home Page

 UNITED STATES DEPARTMENT OF LABOR

 BUREAU OF LABOR STATISTICS

 BLS CE

Search

Explore

Basket 0

Help ▾

bls@example.org ▾

## Home

Welcome. This appears to be your first visit. Click the button below to take a quick tour.

Tour this Page

### Welcome

Welcome to the BLC CE DDI Metadata Proof of Concept system.

### Quick Links

Process Flow

2013 UCC Codes

Interview Series

EDUCA

QHI3MCX



2  
Series



8  
Questions



5,915  
Conceptual Variables

# 2. Find and click the UCC in question

UNITED STATES DEPARTMENT OF LABOR

BUREAU OF LABOR STATISTICS

BLS CE

Search

Explore

Basket 0

Appears Within

Common Metadata

- 530411 Taxi fares and limousine services on trips
- 530412 Taxi fares and limousine services
- 530510 Intercity train fares
- 530901 Ship fares
- 530902 School bus
- HEALTH Healthcare
  - HLTHINSR Health insurance
    - COMHLTIN Commercial health insurance
      - 580111 Traditional fee for service health plan (not BCBS) (thru Q20131)
      - 580113 Preferred provider health plan (not BCBS) (thru Q20131)
      - 580115 Fee for service health plan (not BCBS) (new UCC Q20132)
    - BCBS Blue Cross, Blue Shield
      - 580112 Traditional fee for service health plan (BCBS) (thru Q20131)
      - 580114 Preferred provider health plan (BCBS) (thru Q20131)
      - 580116 Fee for service health plan (BCBS) (new UCC Q20132)
      - 580312 Health maintenance organization (BCBS)
      - 580904 Commercial medicare supplement (BCBS)
      - 580906 Other health insurance (BCBS)
      - 580311 Health maintenance organization (not BCBS)
      - 580907 Medicare prescription drug premium
    - COMEDOTH Commercial medicare supplements and other health insurance
      - 580903 Commercial medicare supplement (not BCBS)
      - 580905 Other health insurance (not BCBS)
    - 580400 Long term care insurance
  - MEDSERVS Medical services
    - 560110 Physician's services
    - 560210 Dental services
    - 560310 Eyecare services
    - 560400 Service by professionals other than physician
    - 560330 Lab tests, x-rays
    - 570111 Hospital room and services
    - 570240 Medical care in retirement community
    - 570220 Care in convalescent or nursing home
    - 570230 Other medical care services
  - 540000 Prescription drugs

# 3. Obtain desired information

UNITED STATES DEPARTMENT OF LABOR

BUREAU OF LABOR STATISTICS

BLS CE Search Explore Basket 0

Help

bls@example.org



Appears Within

## Health maintenance organization (not BCBS)

2013 UCC Classification

Name 580311

Label Health maintenance organization (not BCBS)

Is Missing False

Conditional Variables

- EIHB.HHIBCBS - Is the insurance company Blue Cross/Blue Shield?
- EIHB.HHICODE - Insurance type code

Value Source Variables EIHB.QHI3MCX - Health insurance payments, reference period minus current month

# 4. Get additional information

Appears Within

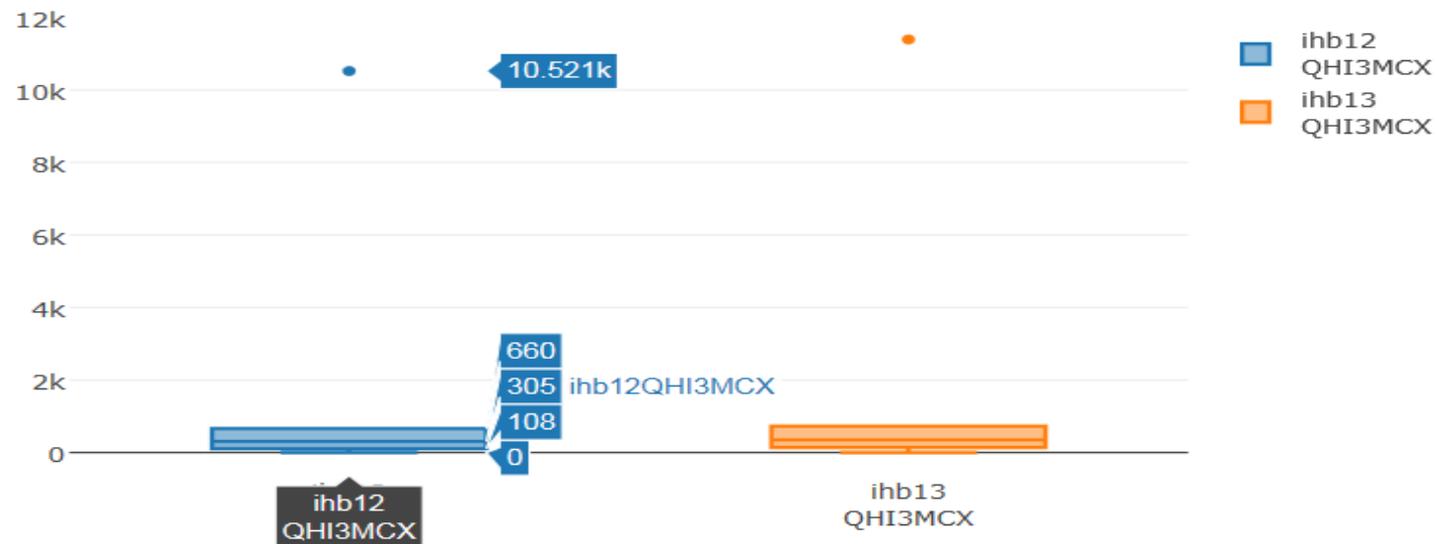
- Interview
- 2013
- ihb13

## Amount paid for health insurance premiums during the reference period

Variable Details Lineage Concordance Variables

Statistics Code Comparison Correspondence Tree

Health insurance payments, reference period minus current month



# 4A. Get additional info from different years

Health insurance payments, reference period minus current month

Interview

2012 2

Conceptual Variable

Name QHI3MCX

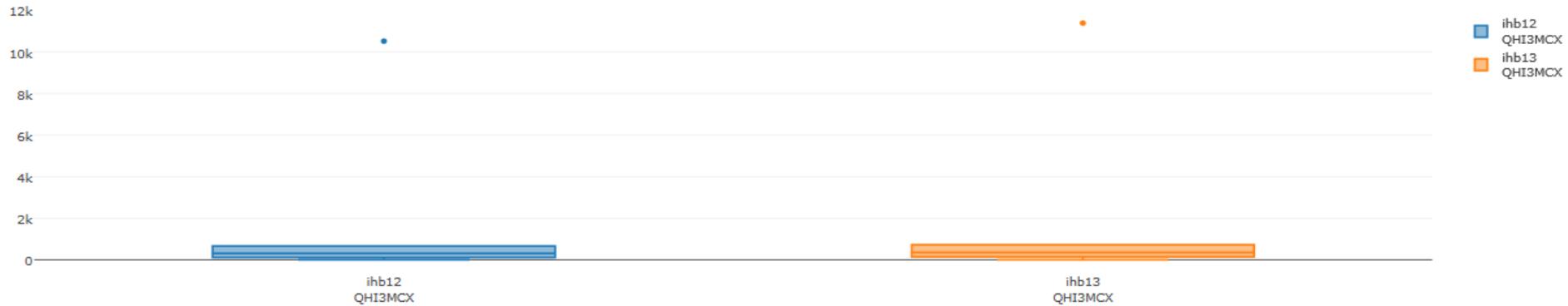
Label Health insurance payments, reference period minus current month

Statistics

Code Comparison

Correspondence Tree

Health insurance payments, reference period minus current month



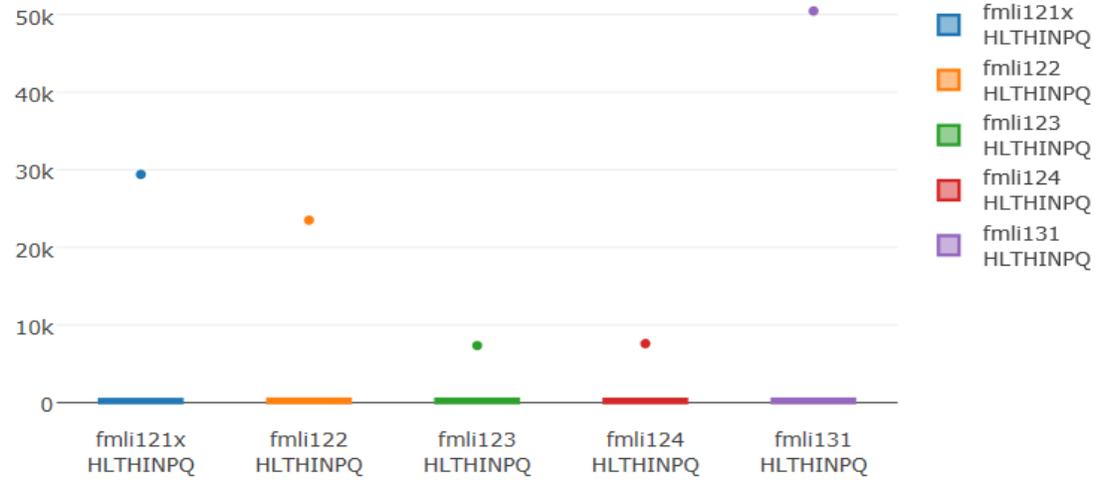
Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
<input type="checkbox"/> ihb12	QHI3MCX	25091	8755	0	108	305	660	10,521	501	655
<input type="checkbox"/> ihb13	QHI3MCX	22949	9311	0	139	351	728	11,392	576	799

# 4B. Get additional info for different subsystems

Appears Within

- Interview
- 2012
- fml123

Health insurance last quarter



Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
<input type="checkbox"/> fml121x	HLTHINPQ	6838	0	0	0	160	428	29,401	331	642
<input type="checkbox"/> fml122	HLTHINPQ	6715	0	0	0	160	465	23,501	351	632
<input type="checkbox"/> fml123	HLTHINPQ	6689	0	0	0	167	464	7,355	357	595
<input type="checkbox"/> fml124	HLTHINPQ	6751	0	0	0	160	443	7,610	337	541
<input type="checkbox"/> fml131	HLTHINPQ	6769	0	0	0	150	452	50,450	347	834

# A User's Perspective

## Colectica can help us by:

- Limiting the different folders and files need to research
- Saving on the amount of time to look up info
- Making all metadata readily accessible in one platform
- Facilitating easy access to descriptive data



# Consumer Expenditure Surveys

Contact Us:

[krishnamurty.parvati@bls.gov](mailto:krishnamurty.parvati@bls.gov)

[noel.reginald@bls.gov](mailto:noel.reginald@bls.gov)

CE DDI Team

Daniel Gillman

Evan Hubener

Parvati Krishnamurty

Reginald Noël

Bryan Rigg

Arcenis Rojas

Lucilla Tan

Taylor Wilson

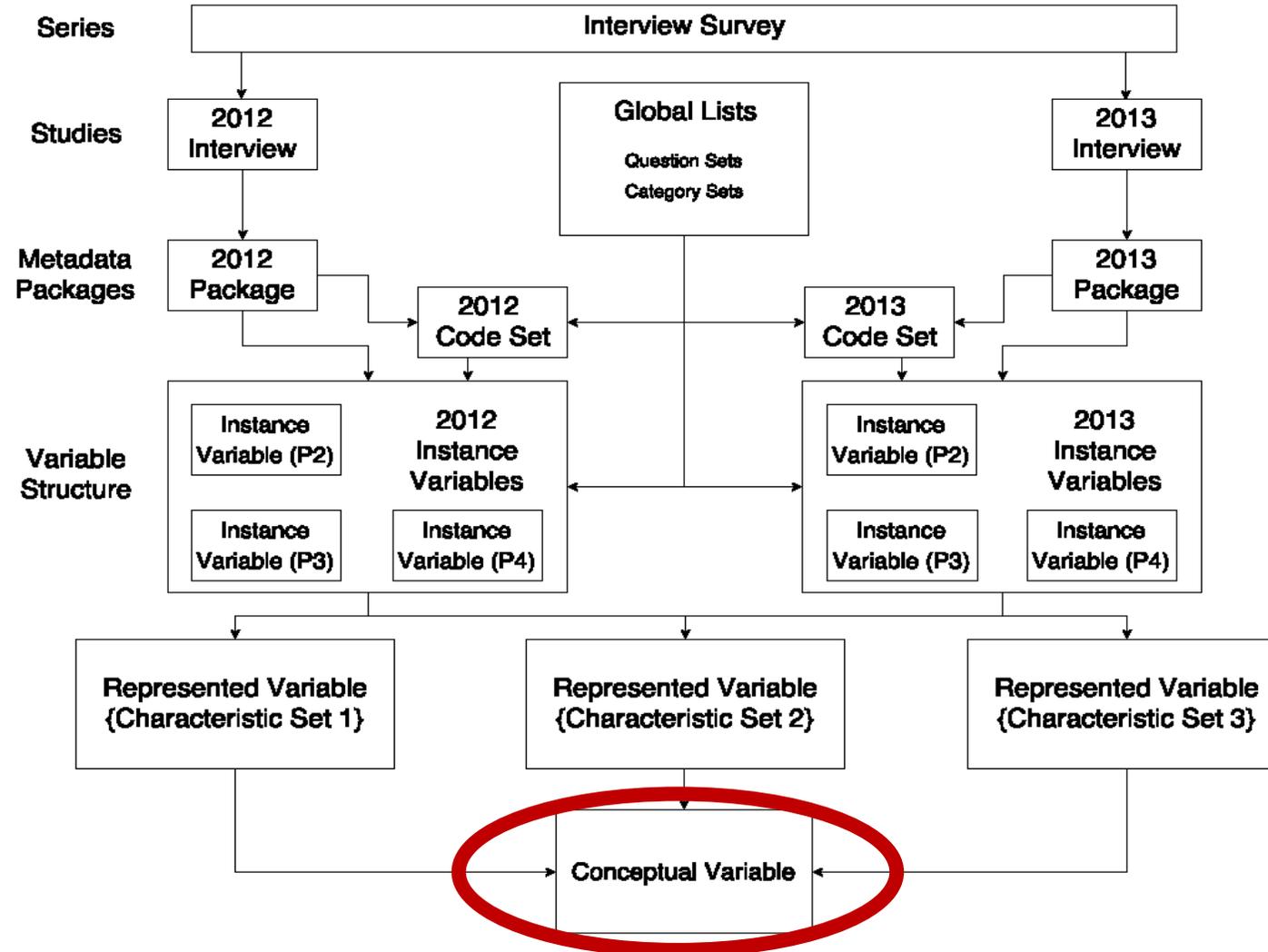


# Extra slides



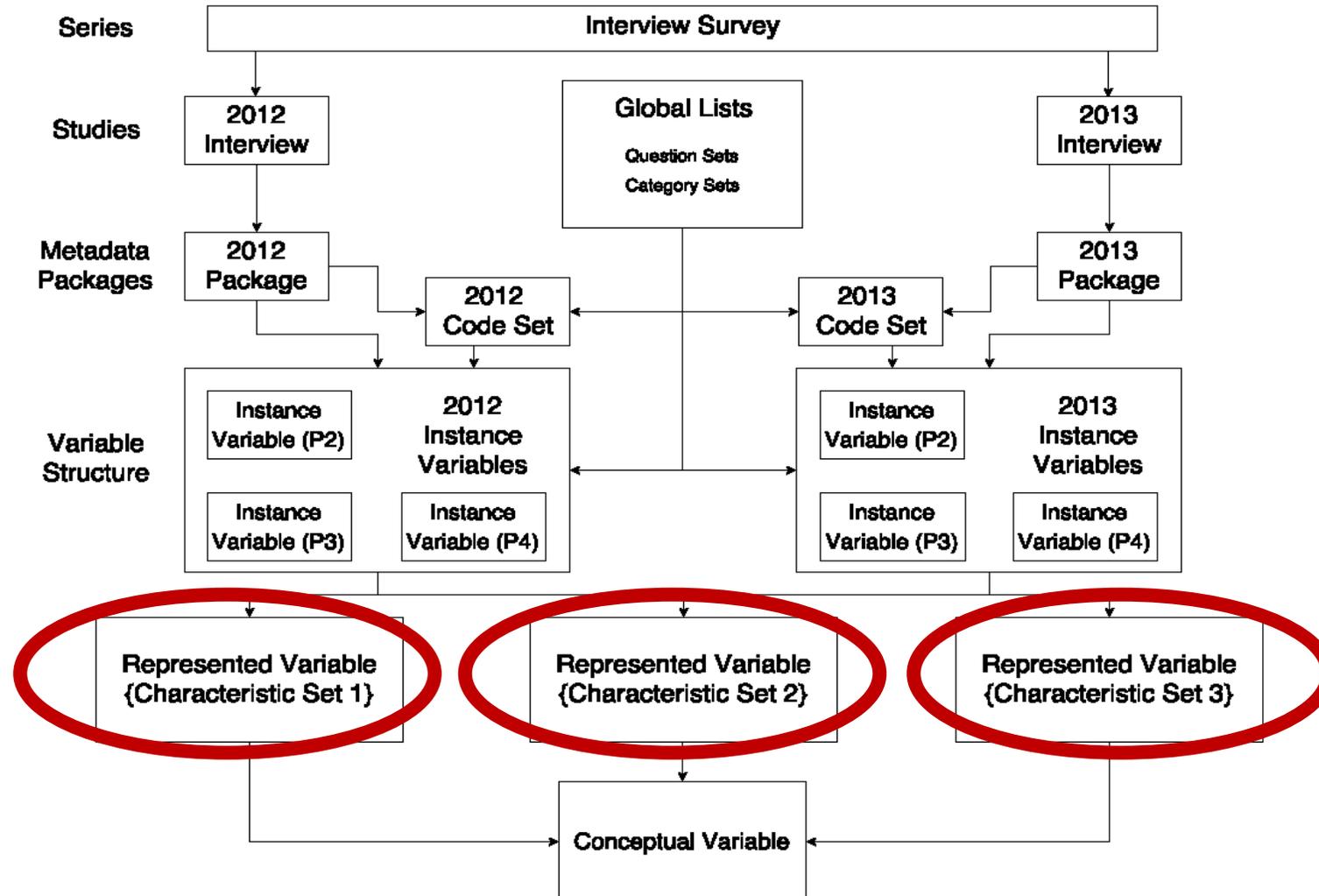
# DDI Structure for Consumer Expenditure Quarterly

## Iteration I



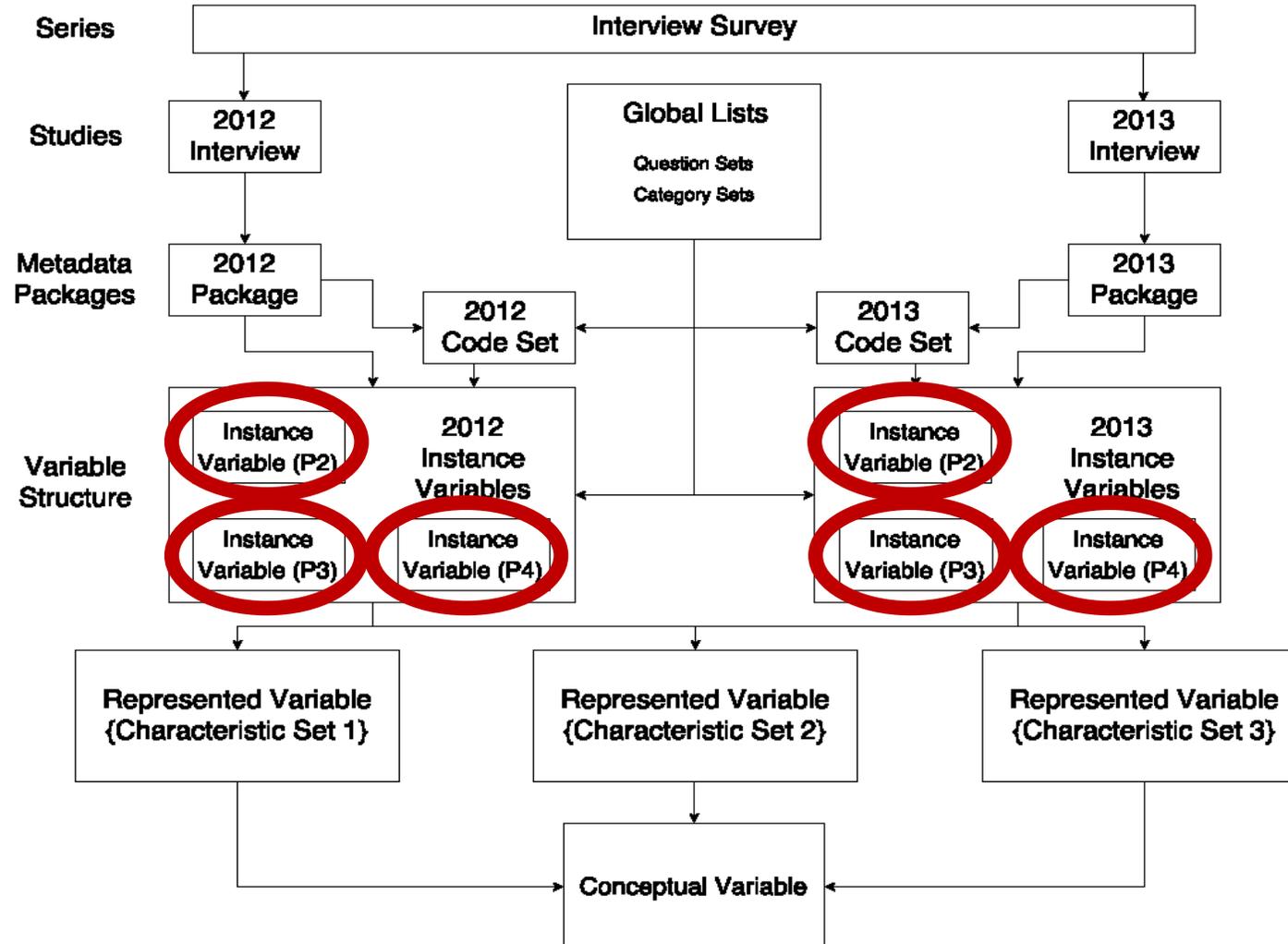
# DDI Structure for Consumer Expenditure Quarterly

## Iteration I



# DDI Structure for Consumer Expenditure Quarterly

## Iteration I



# CE DDI/Colectica Experience

## Iteration 1

- Explored use of a few basic DDI constructs to describe the CE Interview Survey metadata. [Colectica Designer]
- Colectica implements DDI

## Iteration 2

- Proof-of-concept prototype [Colectica Portal and Repository]
- Eight survey questions across 2 years of the Interview Survey.
  - Establish and view the linkage of a survey question to its end as a variable in the public-use Microdata file
  - Identify changes to a variable: for e.g., a change in the response options to the EDUCA (member education attainment) variable

# What's Next?

## Iteration 3

- Will focus solely on Microdata and Tables (**Ph4**)
  - Inputting metadata from **1996 to 2017**
- Continue development work with Colectica
- Scale up prototype, including **Diary**
- Incorporate the use of **Colectica Questionnaire**
- Create one portal for **the public** and another for **interoffice use**

# Potential Hurdles

## Iteration 3

- Requires IT support
- Setting up a server
- Colectica programming language (C-sharp, .NET)
- Equipment requests
- Limited funding



# Cover page

## Household Operations: Cover Page

### A. Section Instructions:

This section is a laundry list section that follows the laundry list standards.

Design will be a table/grid format

Repeat OPSCSC through OPSOTH until OPSCSC = 99.

After OPSOTH, the FR is expected to continue reading the list of expenditures from the cluster screener (OPSCSC)

### B. Section Format:

Table with a maximum of 14 rows

### C. Don't knows and refusals:

1. Do not allow Don't Know or Refusal on:  
OPSCSC, OPSIDE, OPSECE, OPSCOM, OPSSEC
2. Do not display Don't Know or Refusal options on screen
3. Where Don't Know and Refusal are allowed, use keyboard shortcuts:

Don't know = CTRL + D

Refusal = CTRL + R

Bookmark=CTRL + B



# Cover page (continued)

## D. Fills: 1.OPSCOM, OPSOUN and OPSOTH:

OPSCSC =	Category Name	OPSIEX Less than	OPSIEX Greater than
1	Gardening or lawn care services?	15	600
2	Housekeeping services?	20	1260
3	Termite or pest control services?	16	742

**E. Fills:** For OPSCOM, OPSOUN and OPSOTH: fill in...

**F. Outlets eligibility:** Outlets questions will be assigned based on a rotation matrix with POPS categories by PSU by quarter. The following categories are outlets eligible:

**G. Item description:** For OPSIDE: The text field should be a blank, non-required field to allow entry of the item description by the FR. After the FR enters a list number the instrument should place the cursor in the blank item description field. FRs can choose to hit enter to skip this. In case the item description is needed for a subsequent fill, the item code title will be used.