Implementing the Data Documentation Initiative (DDI) for the Consumer Expenditure Surveys

Taylor J. Wilson
Bureau of Labor Statistics
Federal Committee on Statistical Methodology (FCSM)
Washington Statistical Society (WSS)
Washington, D.C.
14 September 2018

Roadmap

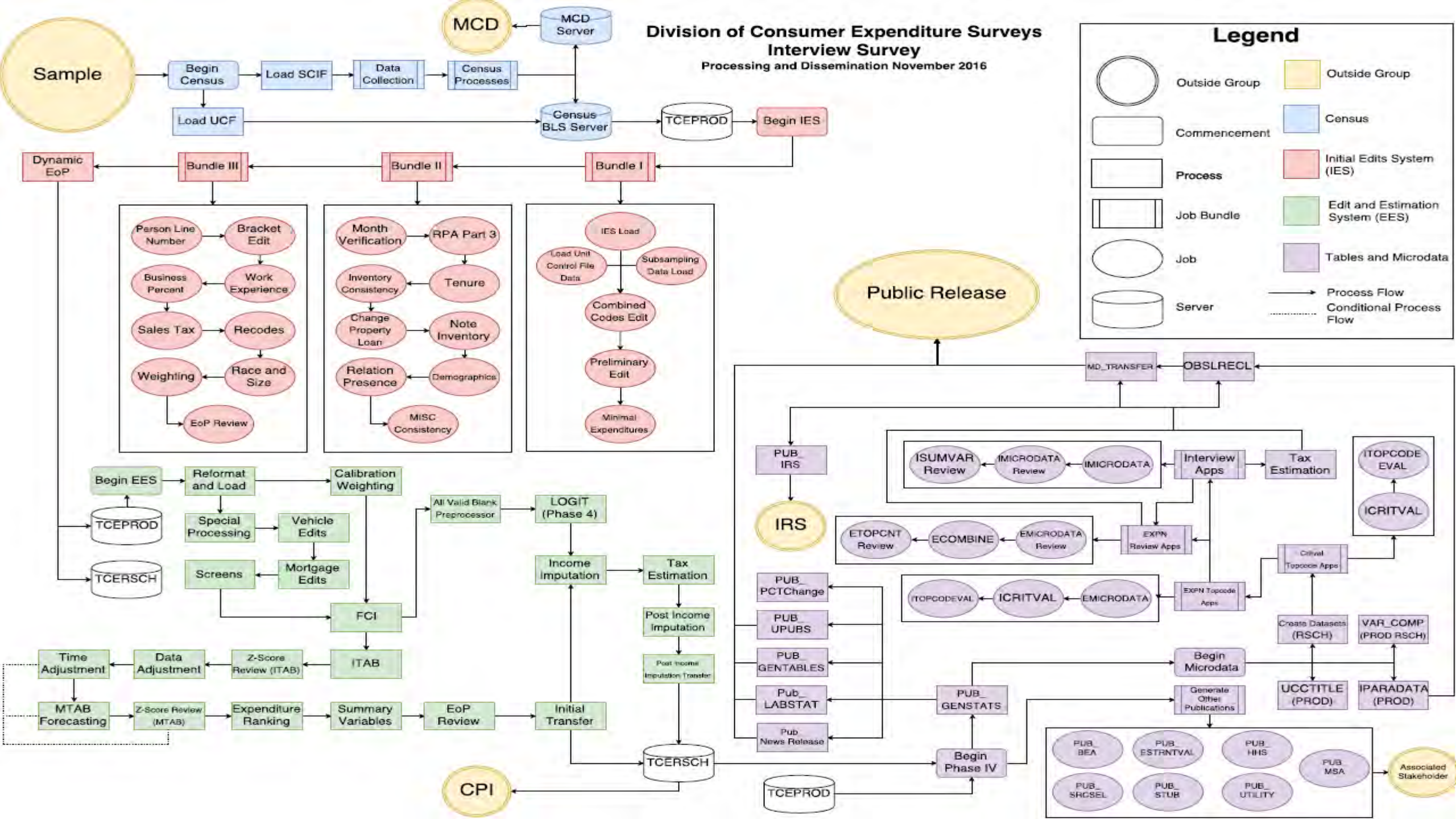
- Survey challenges
- DDI in the Consumer Expenditure Surveys
- Implementation to date
- Technical requirements
- Going forward

Survey Challenges

Consumer Expenditure Surveys

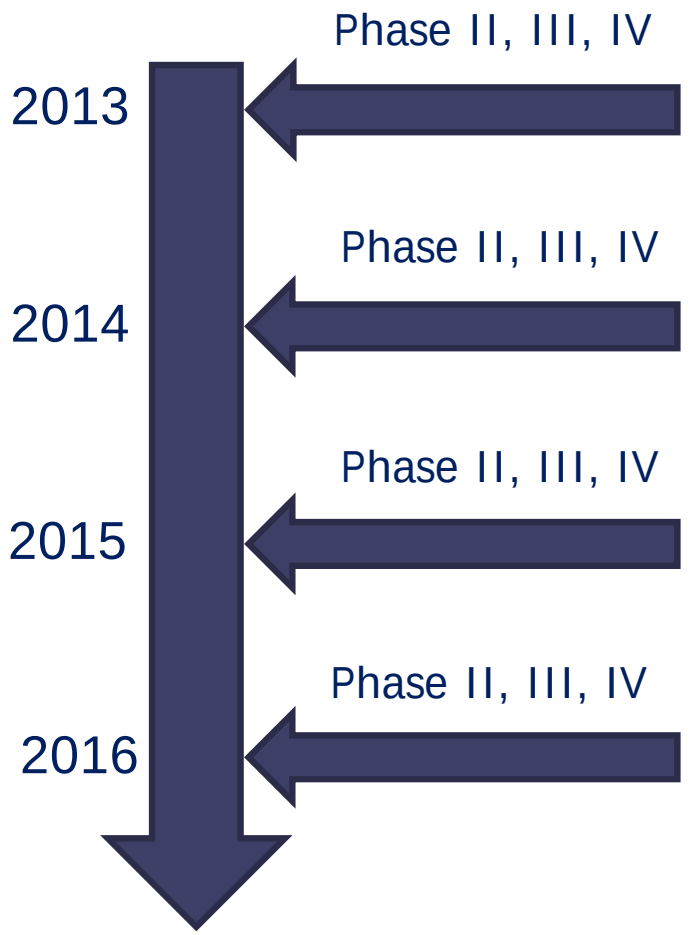
- Complex surveys: massive amounts of data
- Data stored on servers going back to 1981
- Interview Survey: over 60 datasets; 1,800 variables
- Diary Survey: 10 datasets; about 1,600 expenditure items
- Over 70 data processing jobs or edits

Division of Consumer Expenditure Surveys
Interview Survey
Processing and Dissemination November 2016



One Survey, Two Dimensions

- Most end users are interested in changes across time.
- Changes to a variable through the processing phases help users understand how the data are generated.



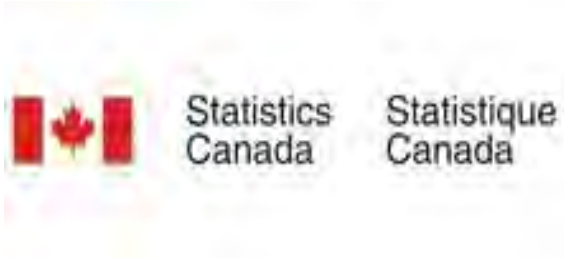
DDI in the Consumer Expenditure Surveys

Data Documentation Initiative (DDI)

- International metadata standard useful for describing social science datasets.

- Tool box framework allows for an ad hoc solution to the interoperability problem between survey programs.
  - "Enter once, use often."
  - Effective Data Sharing
  - Standardized Terminology

Who uses DDI?


Statistics Canada
Statistique Canada


Australian Bureau of Statistics


Statistics Netherlands


eurostat


Statistics
New Zealand


BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

Problem: Current Metadata Structure

- Fragmentation among the subsystems
- Metadata exists in different folders and forms, for different surveys, subsystems, and years
- Many different points of manual input and alteration
- There is a need across all subsystems of CE to develop a standardized method for documenting and linking metadata

Problem: Two Surveys Same Concept

- Diary and Interview both capture expenditures
  - Some expenditures are unique to interview
  - Some expenditures are unique to diary
  - Some expenditures are shared between interview and diary.
- Data are integrated in the table format at the end of processing.
  - Creating perfect interoperability between the surveys would be ideal.
  - Document conceptual differences between survey items.

Ideal Solution: DDI for the Entire Process

- DDI is implemented with the Colectica System
- Each step is cohesively described from survey questionnaire to public-use microdata (PUMD)
- This involves documenting the processing steps within the DDI framework across both surveys
- Documenting processing is useful internally but not externally to better understand what we do and how we do it
  - Continuity of operations (COOP)

Interim Solution: Endpoints of the Survey Lifecycle

- Documenting the end state of the data across time using Colectica
  - Solves persistent user issues
  - Highlights discontinuities and adds context
- Documenting the beginning of the data at collection
  - Questions and responses
- Link these two things at a cursory level
  - Tells users what survey questions produce what PUMD variables
  - How that has changed over time

Implementation to Date

Prototype Goals

- Keyword **search** for variables
- Display **lifecycle** of a variable, from question text through Microdata, UCCs*
- **Query** relationships of variables across CE subsystems*
- Display **changes in a variable across subsystems and/or time** (e.g. change in response options to a question)*
- Display/link to **reference documents** (e.g. Information Booklet)*

Questionnaire Development

- **Question bank** that allows us to retrieve questions or compare questions across years

- **Integrated mapping** of questionnaire metadata
  - Consolidate multiple documents in different locations

- **Streamline changes** for each revision out of word/pdf/csv
  - Make easier to navigate: current full specifications ~ 4,000 pages
  - Iteratively making design changes is difficult

View of Prototype

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.

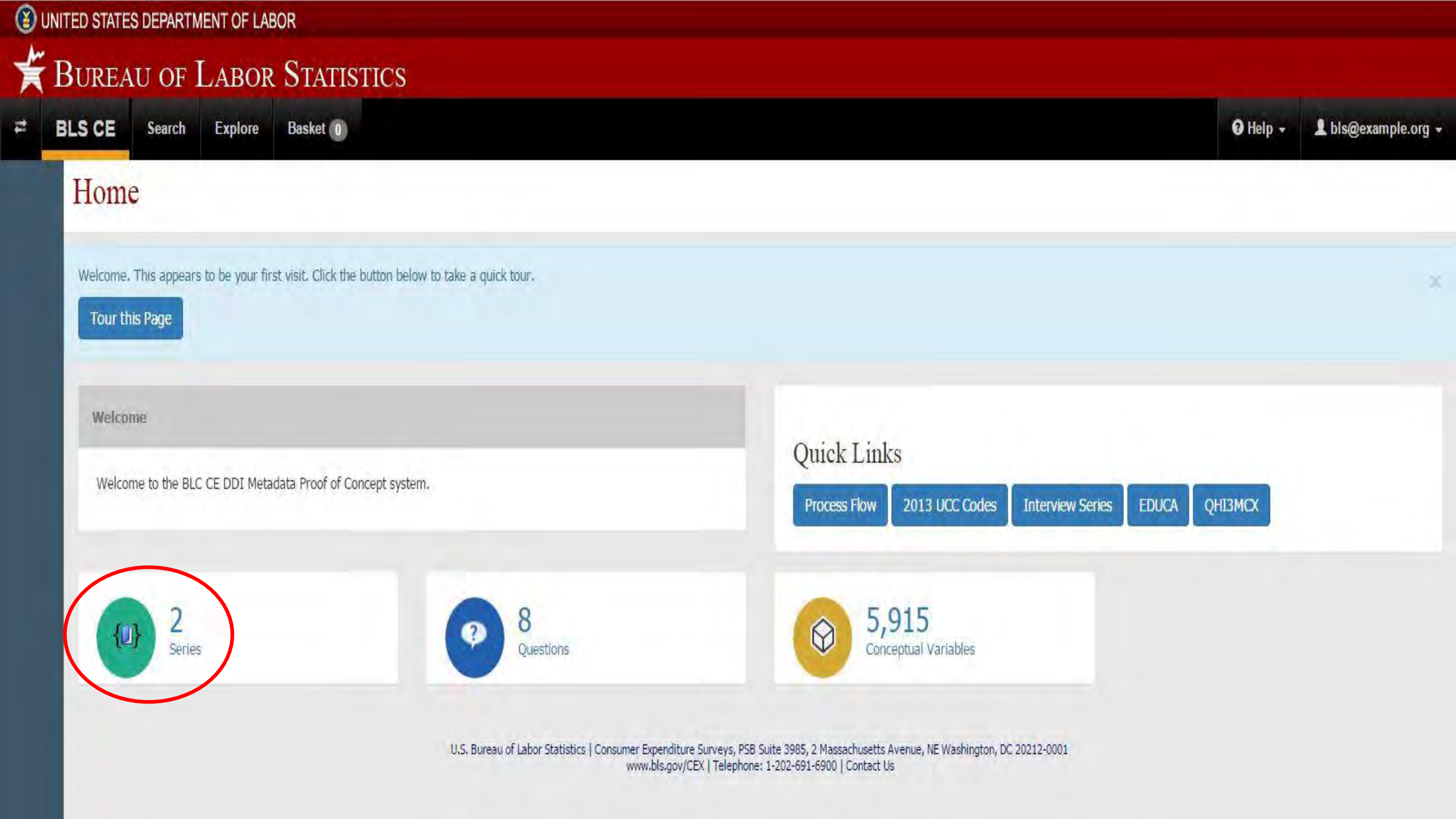
Quick Links

Process Flow | 2013 UCC Codes | Interview Series | EDUCA | QHI3MCX

2
Series

8
Questions

5,915
Conceptual Variables


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
U.S. Bureau of Labor Statistics | Consumer Expenditure Surveys, PSB Suite 3985, 2 Massachusetts Avenue, NE Washington, DC 20212-0001
www.bls.gov/CEX | Telephone: 1-202-691-6900 | Contact Us

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

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

Statistics | Code Comparison | Correspondence Tree

Appears Within

- Interview
  - 2013
  - 2012

QHI3MCX — Health insurance payments, reference period minus current month

QHI3MCX — Health insurance payments, reference period minus current month

2013 EES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

ihb13 - QHI3MCX — Amount paid for health insurance premiums during the reference period

2013 IES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

QHI3MCX — Health insurance payments, reference period minus current month

2012 IES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

ihb12 - QHI3MCX — Amount paid for health insurance premiums during the reference period

2012 EES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

Statistics | Code Comparison | Correspondence Tree

Appears Within

- Interview
  - 2013
  - 2012

QHI3MCX — Health insurance payments, reference period minus current month

QHI3MCX — Health insurance payments, reference period minus current month

2013 EES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

ihb13 - QHI3MCX — Amount paid for health insurance premiums during the reference period

2013 IES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

QHI3MCX — Health insurance payments, reference period minus current month

2012 IES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

ihb12 - QHI3MCX — Amount paid for health insurance premiums during the reference period

2012 EES Subsystem - EIHB.QHI3MCX — Health insurance payments, reference period minus current month

Series

Topic

Conceptual Variables

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

Tour this Page

Search

Sort by: Relevance

Syntax

Item types: Conceptual Variables

Results **1** to **20** of **5,915** (0.043 seconds)

PERSLT18

# of CU Members less than 18 AGE < 18

C_AGE3

# of children whose age is 12-17 CU_CODE=3 and 12 <= AGE < 18

C_AGE2

# of children whose age is 6-11 CU_CODE=3 and 6 <= AGE < 12

UNITED STATES DEPARTMENT OF LABOR

BUREAU OF LABOR STATISTICS

BLS CE | Search | Explore | Basket 0 | Help | bls@example.org

Series

Topic

Conceptual Variables

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

Tour this Page

Search

Syntax

Sort by: Relevance

Item types: Conceptual Variables

Results **1** to **20** of **5,915** (0.043 seconds)

PERSLT18

# of CU Members less than 18 AGE < 18

C_AGE3

# of children whose age is 12-17 CU_CODE=3 and 12 <= AGE < 18

C_AGE2

# of children whose age is 6-11 CU_CODE=3 and 6 <= AGE < 12

Series

Topic

Conceptual Variables

education

Syntax

Sort by: Relevance

Item types: Conceptual Variables
Query: education

Results **1** to **20** of **21** (0.047 seconds)

EDUCACQ

Education this quarter

EDUCACQ

Education this quarter

EDUCA2

Education of spouse

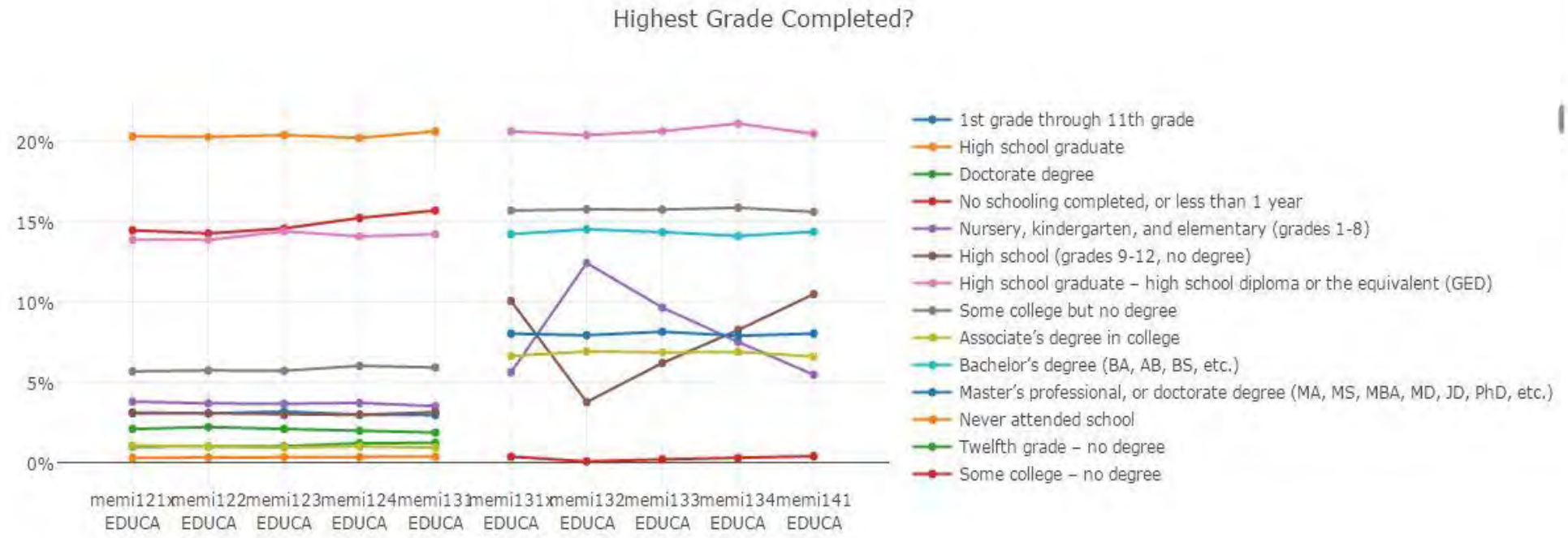
EDUCA2

Education of spouse

Appears Within

- Interview
  - 2012
  - 2013

Statistics | Code Comparison | Correspondence Tree

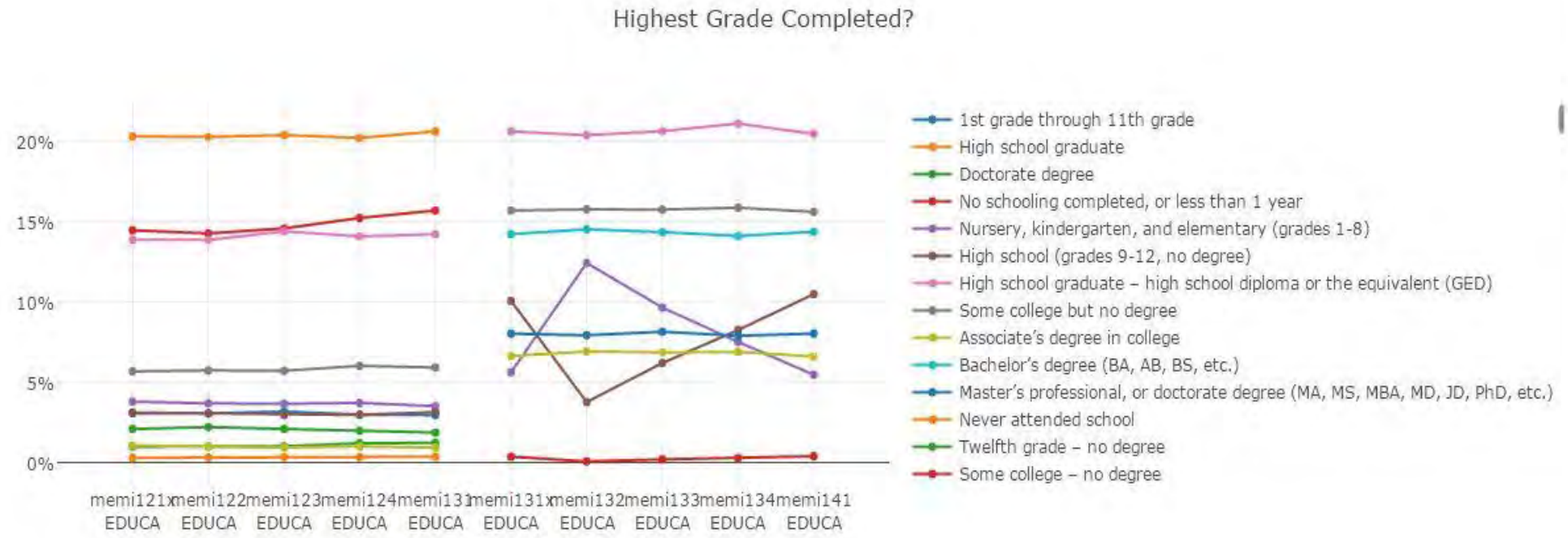


Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
memi121x	EDUCA	17090	0							
memi122	EDUCA	16950	0							

Appears Within

- Interview
  - 2012
  - 2013

Statistics | Code Comparison | Correspondence Tree



Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
memi121x	EDUCA	17090	0							
memi122	EDUCA	16950	0							

Conceptual Variable

Name EDUCA

Label Highest Grade Completed?

Statistics | Code Comparison | Correspondence Tree

	2012 IES Subsystem	2012 EES Subsystem	memi121x	memi122	memi123	memi124	memi131	2013 IES Subsystem	2013 EES Subsystem	memi131x	memi132
Never attended	00	00									
1st grade through 11th grade	01-11	01-11	11	11	11	11	11				
12th grade, no diploma	38	38									
High school graduate	39	39	39	39	39	39	39				
Some college, no degree	40	40									
Associate degree, occupational/vocational	41	41									
Associate degree, academic	42	42									
Bachelor's degree	43	43									
Master's degree	44	44									
Professional school degree	45	45									
Doctorate degree	46	46	46	46	46	46	46				
No schooling completed, or less than 1 year								1	1	1	1

Conceptual Variable

Name EDUCA

Label Highest Grade Completed?

Statistics | Code Comparison | Correspondence Tree

	2012 IES Subsystem	2012 EES Subsystem	memi121x	memi122	memi123	memi124	memi131	2013 IES Subsystem	2013 EES Subsystem	memi131x	memi132
Never attended	00	00									
1st grade through 11th grade	01-11	01-11	11	11	11	11	11				
12th grade, no diploma	38	38									
High school graduate	39	39	39	39	39	39	39				
Some college, no degree	40	40									
Associate degree, occupational/vocational	41	41									
Associate degree, academic	42	42									
Bachelor's degree	43	43									
Master's degree	44	44									
Professional school degree	45	45									
Doctorate degree	46	46	46	46	46	46	46				
No schooling completed, or less than 1 year								1	1	1	1

Conceptual Variable

Name EDUCA

Label Highest Grade Completed?

Statistics | Code Comparison | Correspondence Tree

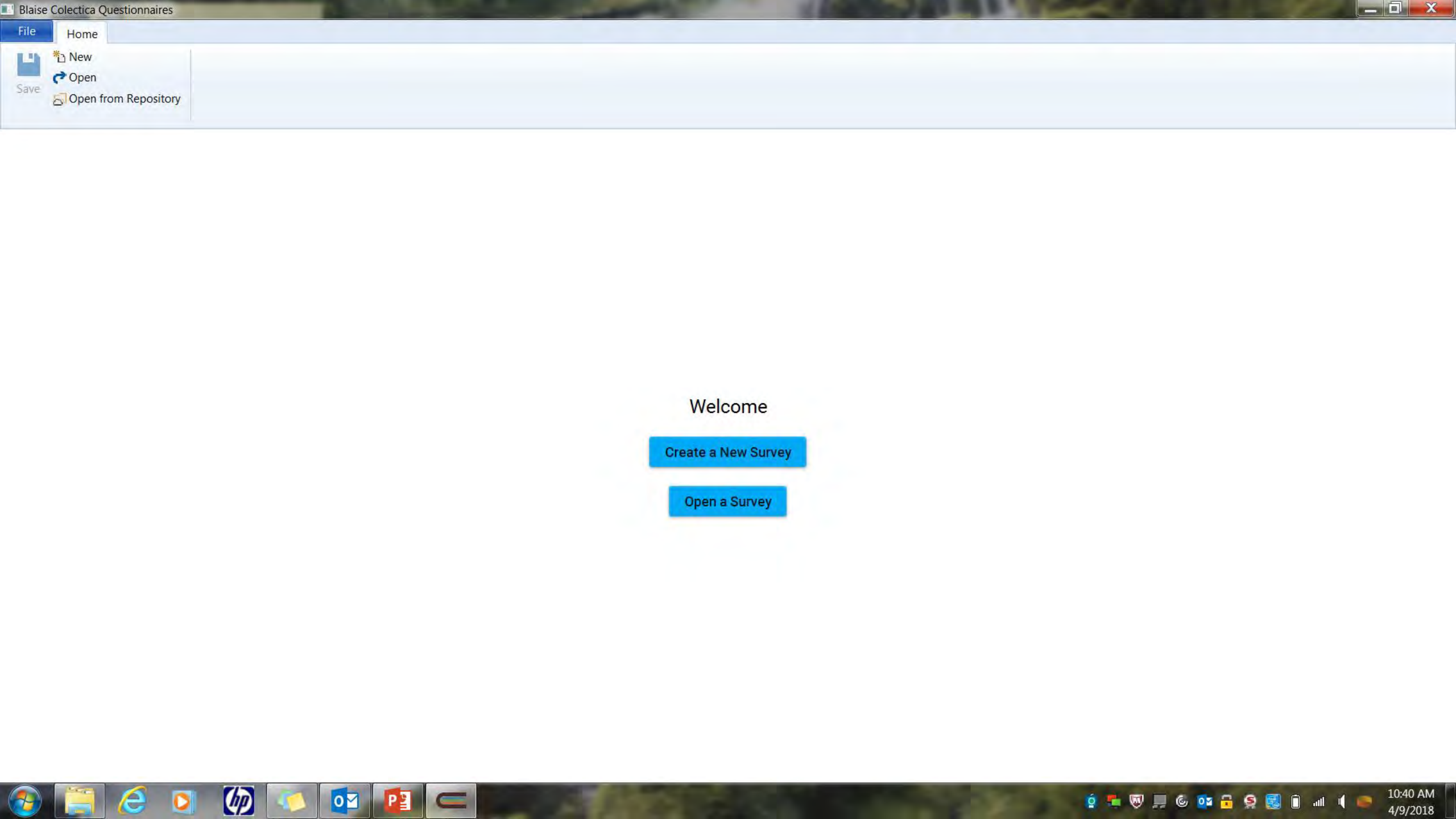
	2012 IES Subsystem	2012 EES Subsystem	memi121x	memi122	memi123	memi124	memi131	2013 IES Subsystem	2013 EES Subsystem	memi131x	memi132
Never attended	00	00									
1st grade through 11th grade	01-11	01-11	11	11	11	11	11				
12th grade, no diploma	38	38									
High school graduate	39	39	39	39	39	39	39				
Some college, no degree	40	40									
Associate degree, occupational/vocational	41	41									
Associate degree, academic	42	42									
Bachelor's degree	43	43									
Master's degree	44	44									
Professional school degree	45	45									
Doctorate degree	46	46	46	46	46	46	46				
No schooling completed, or less than 1 year								1	1	1	1

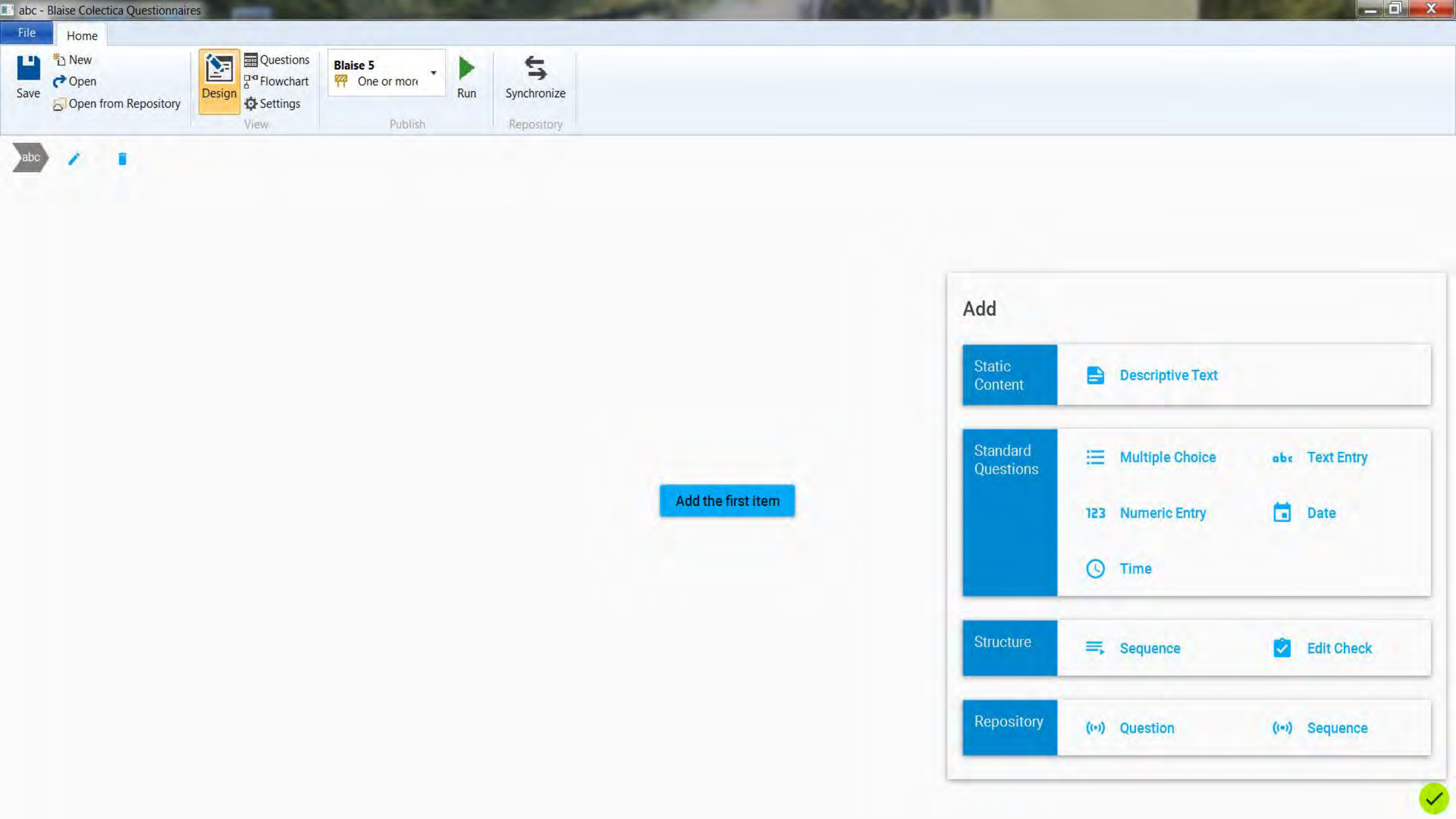
high school diploma or the equivalent (GED)								
Some college but no degree						5	5	5
Associate's degree in college						6	6	6
Bachelor's degree (BA, AB, BS, etc.)						7	7	7
Master's professional, or doctorate degree (MA, MS, MBA, MD, JD, PhD, etc.)						8	8	8
Never attended school	00	00	00	00	00			
Twelfth grade – no degree	38	38	38	38	38			
Some college – no degree	40	40	40	40	40			
Associate's degree (occupational/vocational)	41	41	41	41	41			
Associate's degree (academic)	42	42	42	42	42			
Bachelor's degree	43	43	43	43	43			
Master's degree	44	44	44	44	44			
Professional degree	45	45	45	45	45			

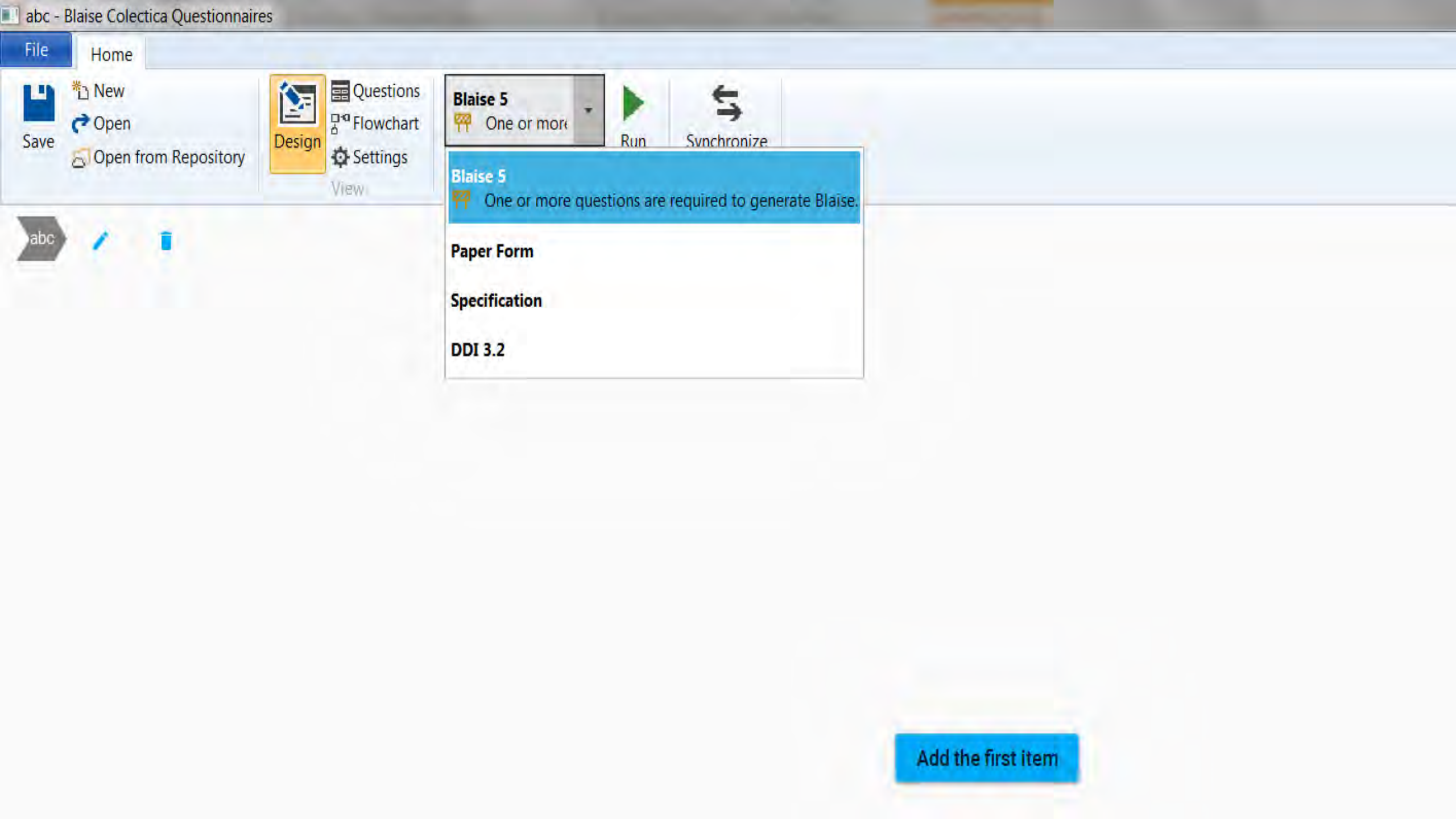
Professional school degree	45	45										
Doctorate degree	46	46	46	46	46	46	46					
No schooling completed, or less than 1 year								1	1	1	1	1
Nursery, kindergarten, and elementary (grades 1-8)								2	2	2	2	2
High school (grades 9-12), no degree								3	3			
High school graduate - high school diploma or equivalent (GED)								4	4			
Some college, but no degree								5	5			
Associate's degree in college								6	6			
Bachelor's degree (BA, AB, BS, etc.)								7	7			
Master's, professional, or doctorate degree (MA, MS, MBA, MD, JD, PhD, etc.)								8	8			
High school (grades 9-12, no degree)										3	3	3
High school graduate – high school diploma or the equivalent (GED)										4	4	4

Professional school degree	45	45										
Doctorate degree	46	46	46	46	46	46	46					
No schooling completed, or less than 1 year								1	1	1	1	1
Nursery, kindergarten, and elementary (grades 1-8)								2	2	2	2	2
High school (grades 9-12), no degree								3	3			
High school graduate - high school diploma or equivalent (GED)								4	4			
Some college, but no degree								5	5			
Associate's degree in college								6	6			
Bachelor's degree (BA, AB, BS, etc.)								7	7			
Master's, professional, or doctorate degree (MA, MS, MBA, MD, JD, PhD, etc.)								8	8			
High school (grades 9-12, no degree)										3	3	3
High school graduate – high school diploma or the equivalent (GED)										4	4	4

View of Questionnaire


Blaise Colectica Questionnaires
File
Home
New
Open
Open from Repository
Save
Welcome
Create a New Survey
Open a Survey
10:40 AM
4/9/2018


abc - Blaise Colectica Questionnaires
File
Home
Save
New
Open
Open from Repository
Design
Questions
Flowchart
Settings
View
Blaise 5
One or more
Publish
Run
Synchronize
Repository
abc
Add the first item
Add
Static Content
Descriptive Text
Standard Questions
Multiple Choice
Text Entry
Numeric Entry
Date
Time
Structure
Sequence
Edit Check
Repository
Question
Sequence


abc - Blaise Colectica Questionnaires
File
Home
Save
New
Open
Open from Repository
Design
Questions
Flowchart
Settings
View
Blaise 5
One or more
Run
Synchronize
Blaise 5
One or more questions are required to generate Blaise.
Paper Form
Specification
DDI 3.2
abc
Add the first item

File | Home

Save | New | Open | Open from Repository

Design | Questions | Flowchart | Settings — View

Blaise 5 | Run — Publish

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

Technical Requirements

Servers and C#

- SQL Server software necessary to handle Colectica materials
  - PostgreSQL
- C# Programming using a Software Developer's Kit (SDK)
  - We outsourced this to the Colectica developers

Going Forward

Next Steps: Short Term

- Internal server being set up by BLS for us to use.
  - Must meet the requirements of the Colectica software to be useful.
  - Prototype was built on Colectica local servers.
- Team is placing as much within-scope metadata (1996-present) as possible into machine readable formats.

Next Steps: Long Term

- The processing systems are being redesigned!
  - Once this process is complete the new system will be documented by our system.
- Roll out a public-facing portal to make the user experience with the data more efficient and effective.

Contact Information

CE_DDI@bls.gov

Taylor J. Wilson
(202) 691-6550
Wilson.Taylor@bls.gov