A Decade in Review:

The evolution of online data collection of establishment surveys at the U.S. Census Bureau

FCSM 2024

Temika Holland, U.S. Census Bureau

Melissa Cidade, Bureau of Justice Statistics

Rebecca Keegan, U.S. Census Bureau

A case study: Economic Census

- The Economic Census is a **mandatory survey** conducted by the Census Bureau **every five years**. The survey collects data electronically from over **4 million businesses** (including large, medium and small companies representing all U.S. and Island area locations and industries) on a range of operational and performance questions. Data from the survey are used as the official five-year measure of American **business and the economy**.

United States® **Census** Bureau

Let's journey through time….

- **2012 Economic Census:** continued migration from paper forms to electronic reporting
  - Single-units have the ability to use online reporting
  - Multi-units reported electronically using Surveyor
    - Downloadable application

- **2017 Economic Census:** online reporting
  - Online reporting for multi- and single-unit companies
    - Only paper forms for Island Areas single units (Puerto Rico, U.S. Virgin Islands, Guam, American Samoa, and Northern Marianas)

2019: Record keeping study

- Objectives
  - **Definitions**: How do businesses define themselves relative to the Census Bureau definitions?
  - **Accessibility**: How accessible are key data points at varying business units?
  - **Burden**: How resource intensive is gathering data at these varying business units?

- What did we learn?
  - **Mismatch in North American Industry Classification System (NAICS) classification**
  - Organizational structure of business impacts reporting
  - Some company level data are inaccessible

North American Industry Classification System (NAICS)

Primary Business or Activity (PBA)

Historically, the Census Bureau has assigned an establishment's NAICS code before sending the corresponding Economic Census form

- NAICS write-in responses:
  - Write-ins = 500K in 2017,
  - Referral rates = 15% in 2017

- Misalignment of the NAICS could lead to survey questions or pre-listed response options that are not applicable to the respondent
  - Prelisting for North American Product Classification System (NAPCS) product/service
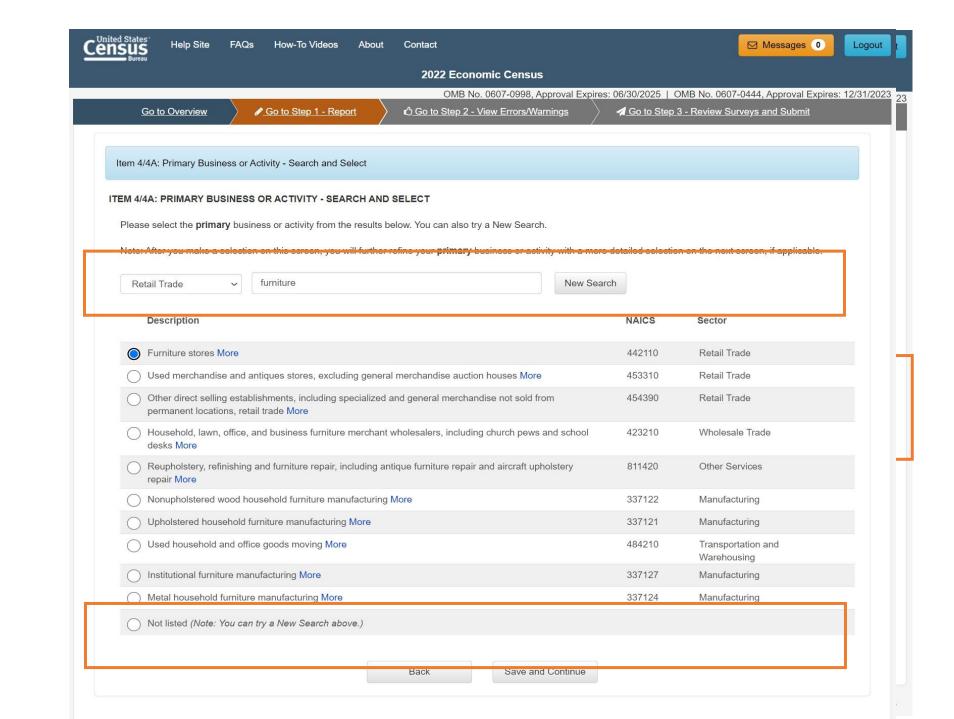
2021: Introducing BEACON!

Business Establishment Automated Classification of NAICS (BEACON)

- Utilize a "Search" functionality with **machine learning** that enables determination of the "correct" questionnaire path, in real time (i.e., "dynamically"), based on respondents' self-identification of their Primary Business or Activity (PBA)

- Results are based on *relevance scores,* ranging from 0 to 100
  - BEACON presents the NAICS codes sorted from highest to lowest score; Scores are not presented to respondents

Introducing BEACON!

United States Census Bureau

Machine Learning and survey design

- How do we communicate the use of this new feature effectively to respondents?
  - What level of detail are needed?
  - Instructional text and Help content
  - Display features
  - Error messaging

- User experience
  - Meeting the needs, wants and expectations of respondents

Pre-test evaluation:
Iterative Usability Testing

Method Details:

- Multiple rounds of testing
  - Round 1 – N = 29
  - Round 2 – N = 24
  - Round 3 – N = 27
  - Round 4 – N = 43

- Started with task-based interviewing
  - Low fidelity mock-ups

- Heat maps and subjective satisfaction ratings

Advantages:

- Large volume of feedback over multiple modalities

- Flexibility in iterations of instrument design

- Ability to verify performance improvements

- Refining communications for the new feature (e.g., instructions, labels, etc.)

Pre-test evaluation:
Pilot Study

Method Details:

- 2021 Industry Classification Report Field Test ("Refile" Pretest)
  - Approx 37,000 establishments
  - 20,000 write in responses for BEACON
  - Inclusion of usability web probes

- 43 participant debriefing interviews interviews (respondents and non-respondents)

Advantages:

- Induces field conditions for response

- Enormous amounts of paradata

- Contextual data from mix methods design

Production/Post Survey evaluation:
Respondent Debriefing Interviews

Method Details:

- Rolling interviews over the length of the field period:
  - 24 single-location businesses
  - 23 multi-location businesses
  - 9 "alternate reporting units" (complex businesses)

- Pulled response data to inform interviewing

Advantages:

- Post-completion evaluation of instrument performance

- Additional feedback for future iterations

- Contextualized response data

Production/Post Survey evaluation:
Analyst Focus Groups

Method Details:

- Two 1-hour long focus group sessions with analysts

- Moderator protocol guided conversation

- Challenges and successes of the instrument

Advantages:

- Respondent issues through expert filters

- Specific vignettes from the field

- Brainstorming next iterations

The impact: 2022 Economic Census

Utilized 526,000 times during the 2022 Economic Census by single unit establishments

- 82% searches resulted in selections from BEACON machine learning results

Lower referral rates and fewer write-ins in 2022

NAICS

NAICS Referral rates
- 15% in 2017
- 5.3% in 2022

NAICS Write-ins
- 500K in 2017
- 166K in 2022

NAPCS

NAPCS Referral rates
- 34% in 2017
- 7% in 2022

NAPCS Write-ins
- 1M in 2017
- 219K in 2022

Beyond the Economic Census:
Machine Learning at Census

- **Commodity Flow Survey**
  - Categorization of products
  - Pilot test results: Ability to confidently classify 95% of shipments
- **Annual Capital Expenditures Survey**
  - Collects information about expenditures at the company level
  - Increase of write-ins with a prediction of "Structures", or "Equipment" being moved to the appropriate category
    - Early results showed a 78% increase

Looking ahead…

- Expansion of Machine Learning – pending research
  - Expanding the data dictionary
  - Spanish language inclusion
  - Predictive text/autocomplete searches

- Spreadsheet reporting
  - Online, embedded spreadsheets
  - Offline, downloadable templates
  - Flexible reporting, unstructured data

- System to system response

Thank you!

Temika.holland@census.gov