Executing a Multi-Year Multi-Method Electronic Data Collection Re-engineering: Experiences from 2017 Economic Census Development and Pretesting

Amy E. Anderson Riemer
U.S. Census Bureau
Washington, DC 20233

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract
The 2017 Economic Census will be collected entirely online for the first time. In previous economic censuses, all respondents were offered the option of reporting via paper or electronically. Smaller companies could report on the Web starting in 2012, but larger companies were required to download a software application in order to enter and upload data.
For the 2012 Economic Census, there were over 600 questionnaire versions, tailored by industry. The software application included several features to assist business respondents with the management and collection of potentially thousands of pieces of data from larger companies. A key feature for larger businesses was the availability to upload pre-formatted spreadsheets containing their data.

In order to prepare for the change in platform for the 2017 Economic Census, a multi-year/multi-method research effort took place to identify key requirements, test prototypes, and evaluate early versions prior to implementation in the economic census. In addition to ensuring that key functionality from the software and legacy Web platform were transferred, researchers took the opportunity to identify and evaluate improvements. Researchers and survey managers also took advantage of annual establishment-level surveys, such as the Annual Survey of Manufactures, for early adoption and evaluation of some of these new features.

This paper will discuss the pretesting research methods used to transition these various response options into a single re-engineered Web system. These methods included requirements gathering, usability testing, respondent debriefings, paradata analysis, and behavior coding. The paper will also address the methodological and practical challenges faced in creating and conducting this research, including lessons learned.

Background
The Economic Census is a mandatory survey conducted every 5 years ending in 2 or 7. The economic census requests comprehensive accounting, payroll, and business activity information from all locations within a company. Nearly 4 million businesses of varying sizes and industry classification receive the survey. The data collected provides a measure of the U.S. economy and is used by a variety of stakeholders such as policy makers, trade and business associations, other federal agencies, and individual businesses.
Respondents have had the option to report electronically to the economic census since 1997. The survey was initially available within a software application that respondents could download from a CD. Starting in 2000, the data collection software was available to download from the Web. After hearing concerns from small companies

about the burden associated with downloading a software application, and receiving lower electronic response from this segment, a web application was built for the 2012 Economic Census for single location businesses. Multi-location businesses were still required to download a software application if they wanted to respond electronically.

The software application wasn't an ideal platform for larger companies either. During usability testing and debriefing visits with respondents to prepare for the 2012 Economic Census, researchers often heard concerns from larger companies about their ability to download software onto their systems due to increasing security restrictions. We would often hear that Information Technology (IT) staff would need to assist with downloading the software. There were also concerns about barriers for allowing others within the company to access the software, and the inability of respondents to work on the survey from a remote location, such as their homes. Some respondents, with help from their IT departments, were able to create workarounds.

In addition to concerns that respondents raised during research visits, internal staff were also reporting that they were receiving phone calls from companies that couldn't overcome IT hurdles in order to access the software. In these instances, Census Bureau staff would have to work with respondents on an alternative reporting arrangement that usually involved a spreadsheet submitted through a secure web site.

In order to address these concerns that were increasing in frequency, and to move the data collection tool into a more modern setting, planning began early to ensure that the 2017 Economic Census would be conducted completely on the Web. This secure online portal would allow for improved data quality and a reduction in data collection costs.

Overview of the Research Strategy

Moving all respondents to a Web only application that would support businesses of all sizes was going to be a massive undertaking that would require input from respondents and internal stakeholders from the beginning. We were fortunate to know far in advance of these plans. This allowed us the time to create a multi-year research plan that included a variety of quantitative and qualitative evaluation methods. It was necessary to break the research into smaller pieces that were managed by different researchers in order to accomplish all of the research goals. Table 1 shows an overview of the various methods that were a part of this research effort over the last five years.

Table 1. Research Timeline including Instrument Milestones

	2014	2015	2016	2017	2018
Instrument Milestones	2013 ASM/COS – Web Instrument for Single Location; Software for Multi-locations	2014 ASM/COS – Web Instrument for Single Location; Software for Multi-locations	2015 ASM/COS – **Revised** Single Location Web Instrument; Software for Multi-locations	2016 ASM/COS – **Revised** Web Instrument for Single and Multi-Location	2017 Economic Census – **All Web**
Single location	1. Task Analysis 2. Paper Prototype Testing 3. Requirements Gathering - Internal	Usability Testing (2015 ASM/COS)	1. Respondent Debriefings (2015 ASM/COS) 2. Paradata Analysis (2015 ASM/COS) 2. Usability Testing (2016 ASM/COS)	1. Two Rounds Usability Testing (2017 Economic Census) 2. Respondent Debriefings (2016 ASM/COS) 3. Paradata Analysis (2016 ASM/COS)	1. Paradata Analysis (2016 ASM/COS) 2. Paradata Analysis (2017 Economic Census) 3. Respondent Debriefings (2017 Economic Census)
Multi-location	1. Task Analysis & Respondent Debriefing 2. Paper Prototype Testing 3. Requirements Gathering - Internal	1. Two Rounds High-Fidelity Prototype Testing 2. Respondent Debriefings (2014 ASM/COS)	1. Usability Testing (2016 ASM/COS) 2. Respondent Debriefings (2015 COS/ASM		
NAPCS testing	Task Analysis	Prototype Testing	Prototype Testing – Focus on Search Function	Prototype Testing – Focus on Spreadsheet for Multi-Location	Pilot Test – Focus on Write-in Formatting & Respondent Debriefings

Research was initially split by single and multi-location businesses. Part of the reason for the split was because single location companies already had a Web based instrument and their transition to a new Web design would be different then multi-location companies. The other reason was because small and large companies tend to have different response processes when gathering and responding to an economic census. Separate instructions and features have historically been created for multi-location respondents over the years. Table 1 shows that in 2016 research efforts converged to meet the requirement of building one instrument for all locations.

The research for both single and multi-location companies followed a similar path. They both started with debriefing interviews and task analyses to evaluate how respondents were using their respective instruments and to learn about their survey response process. After researchers had a deeper understanding of the survey response process, they worked closely with internal stakeholders to discuss design ideas and translate those into prototypes for testing. These prototypes often evolved over time and went from paper versions or simplified HTML mock-ups developed by researcher or internal stakeholders to high-functioning prototypes developed by programmers.

Prior to launching a new design, researchers would conduct one or two rounds of usability testing. After the release of a newly design instrument, researchers would follow-up soon after data collection to conduct respondent debriefings and analyze Web paradata in order to further evaluate the instrument and provide supplementary feedback for improvements.

In addition to designing and evaluating the new data collection instrument, we created a separate research effort to identify the most effective layout for the new NAPCS product collection. Other research projects were also underway to evaluate new and revised content for the Economic Census.

The Annual Survey of Manufactures and the Company Organization Survey

The Annual Survey of Manufactures (ASM) and the Company Organization Survey (COS) are annual surveys that are collected in the years between the economic censuses. These related surveys ask a subset of economic census inquiries from fewer businesses and use the same data collection platform. In comparison to the sample of 4 million establishments during and Economic Census, the ASM collects data from approximately 50,000 establishments and COS collects data from approximately 47,000 establishments. Because of the relationship between the surveys, changes to data collection instruments are often incorporated into the ASM or COS prior to a wider release for the economic census. There are several advantages to this strategy. If issues arise, it is contained to a smaller number of respondents then during a larger survey collection. Issues can often be addressed during non-census years leading to a better experience for respondents during a census collection.

Having the ASM and COS available during non-census years allowed researchers the ability to conduct continual testing and debriefings with respondents prior to the launch of the newly designed Economic Census Web site.

Requirements Analysis – Internal Stakeholders

The first step towards transitioning from a software based application to a web based application was to analyze and gather requirements for the new data collection platform. To achieve this, researchers met with internal stakeholders and respondents during 2014. The main goal for this initial step was to identify what features within the software application should be maintained, and which should be modified for the Web.

Researchers met with internal stakeholders within the Census Bureau. These included analysts and staff who worked closely with developing the software application and those that worked closely with the respondents during the past economic census. In addition to asking about which features should be maintained, we also asked internal stakeholders for overall ideas for improvements to the data collection tool.

Requirements Analysis – Respondent Task Analysis

In addition to meeting with internal stakeholders, researchers met with single and multi-location business respondents and conducted detailed task analyses in 2014. A task analysis is an early step in the user-centered design process that involves observing users in action to understand how they perform tasks in order to meet their goals (usability.gov). In order to facilitate a task analysis, respondents were asked to provide a detailed description of the process they used when responding to the 2012 Economic Census (Tuttle, 2014).

Conducting real-time observation with businesses is very burdensome on the respondent and can be difficult for a survey like the economic census when it typically takes many hours across days or weeks to complete. In larger companies, the response process often involves multiple staff accessing multiple internal data sources to gather information (Snijkers, 2013).

The goal for the multi-location task analysis was to identify features from the software that respondents did and did not find useful and to solicit recommendations for improvements for the future system. Respondents were also asked to identify useful features from other electronic applications or websites that they felt could be pertinent for the economic census Web application.

Although single location companies had a Web instrument available to them for the 2012 Economic Census, it was expected that their Web design would be altered in order to create one application for all businesses. The goals for the task analysis for these smaller companies was to learn about their record-keeping practices and how electronic reporting could help or hinder response on the Web. It was expected that their response process would not be as complex as multi-location companies.

Prototyping

Feedback from respondents, internal staff, and user experience experts were translated into detailed requirements for the Web instrument and used to create various prototypes for testing. Prototyping (usability.gov) is the process of drafting versions of the final product in order to explore ideas and show the intent behind features or the overall design concept prior to programming the final instrument. Prototypes can range from paper drawings (low-fidelity) to semi-functioning instruments to a fully functioning site (high-fidelity).

A major benefit to prototyping is that early ideas can be quickly tested to see if they are successful before developers spend time on expensive programming. Prototypes are used early in the design process and are often tested iteratively as new features are developed or proposed features are redesigned.

Single location prototype testing was less involved because their instrument was already on the web. However, plans were in place for incorporating a more response-driven design. Low-fidelity paper prototypes were developed and tested with single location respondents to evaluate the best way to display this design. These prototypes were developed with support from researchers and other internal staff to visually display creative solutions that were being discussed. Feedback from respondents during prior debriefings and the task analysis were also used during prototype development. Where possible, programmers were shown the prototypes prior to testing to ensure the proposed designs were feasible.

Early testing for multi-location businesses involved the use of both low-fidelity and high-fidelity prototypes across several rounds of testing. Initial visits with respondents involved paper screen mock-ups that displayed new screen designs. After obtaining initial feedback with paper prototypes, a low-functioning prototype was programmed in HTML which allowed for navigation, branching, and the ability to input selected data. Subsequently, programmers proposed developing a Web application that would be built within the Microsoft SharePoint environment as an option. Programmers developed a semi-functioning prototype in the SharePoint environment as a partial "proof-of-concept." Researchers took this prototype to respondents for their reaction and input. After decisions were made to go in another direction based on many factors, programmers built a high-fidelity prototype for testing in another platform that was again tested with respondents.

Usability Testing

Usability testing is a method for evaluating a product, in this case a Web site, to identify issues. Web-site usability is about the ease at which a user can achieve their goals on a site. The goals in designing a usable web-site include making it learnable, efficient, and satisfying while avoiding as many errors as possible. Usability testing can be conducted in a lab or in the field.

The usability testing that was conducted for the economic census was handled in the field. It is challenging to schedule meetings with business respondents that are outside of their place of business. Going into the field was less burdensome on the business respondents and allowed researchers to observe how the respondent would interact with the survey in their own setting (Nichols, 2017).

Once programmers, researchers, and stakeholders were comfortable with the design direction based off of prototype feedback, plans were made for launching the new design for both the single and multi-location instruments. In 2015 usability testing was conducted on the new response driven design of the single-location Web-site in preparation for the 2015 ASM/COS.

Plans were then made to launch the new web design for all locations during the 2016 ASM/COS. In 2016 two rounds of usability testing were conducted prior to mail-out. The version tested during the first round had the overall look and feel of the final survey, but several features weren't fully functioning. A fully functioning version was available for the subsequent round of usability testing.

Although the Economic Census is similar to the ASM/COS, there are several necessary changes that need to be made to the layout and functionality of the web site for an economic census. The economic census design was also updated to include improvement that weren't able to be included with the 2016 ASM/COS. Because the instrument had significant changes, two rounds of usability testing were conducted in 2017 before the survey was released.

Respondent Debriefings

Respondent debriefings are interviews that are conducted after respondents complete a survey to evaluate either survey content or the data collection tool. During this research, respondent debriefings were continually conducted for both single and multi-location companies after the release of a newly designed web instrument. Respondent debriefings were also conducted early in the process as a tool for requirements gathering. If evaluations of the instrument were unfavorable during respondent debriefings, researchers would include that with other information gathered during the task analysis to inform prototype designs.

Research will not conclude with the mail-out of the 2017 Economic Census. Respondent debriefings are planned during 2018 to evaluate the instrument and provide feedback for continual improvement.

Paradata Analysis

Paradata from both the Web application and software application were analyzed throughout the research process and provided additional insight and direction for usability testing and respondent debriefings as well as design decisions. Paradata from the 2015 ASM/COS was analyzed to isolate which screens and questions within the Web-site were troublesome for the single location respondents. The paradata also identified where there were indicators of burden and what features or functions were being most utilized in the Web application. Researchers reviewed software application paradata from the 2015 ASM to identify how often the various screens within the software application were visited and which features were used or under-utilized by multi-location respondents.

We are currently analyzing 2016 ASM/COS paradata and will be looking to dig deeper into how respondents are using key features within the Web application. We will be combining the paradata with response data in order to identify any potential characteristics (e.g., size of company, industry classification) associated with the degree to which Web application features were used or not used.

We are also preparing for analyzing paradata as responses are arriving from the 2017 Economic Census. The plan is to analyze paradata as it is being captured to help guide where to focus topics of inquiry for respondent debriefings. We also hope to use this information to help target which respondents to debrief.

NAPCS Testing

The 2017 Economic Census will include product categories from the North American Product Classification System (NAPCS). This will allow for the collection of additional detail on products (goods and services) economy-wide (Moore, 2018). The collection of this additional information provided challenges for instrument design. Each business location will be provided with a set of products that would be listed based on their industry classification. In addition to those listed, respondents will be able to write-in products not already listed.

A separate, but related, research project was created in order to develop and test alternative design strategies for the NAPCS item. This research began with a task analysis to understand how respondents record product information in their records and how they might translate that onto an economic census. Several rounds of prototype testing occurred with each focusing on different aspects of the questions (e.g., overall layout, write-in layout, a proposed product search function). Feedback from respondents on the important write-in section of the question was inconclusive during prototype testing. Because of this, a pilot was conducted that included a split-sample design showing two different ways of collecting write-in data. As part of the pilot, respondent debriefings were conducted to further evaluate the designs.

Challenges

Given such a large research effort, it is no surprise that we faced several challenges along the way. One major challenge that we faced was dealing with the logistics of having multiple stakeholders. There are hundreds of staff that work full or part time on the various survey lifecycle pieces on the economic census. Many of them have a stake in the data collection effort. During the research several teams were created to help manage the development of the instrument and survey content. Where possible, researchers became a part of these teams in order to provide

expert guidance and to keep members updated on research plans, progress, and to solicit feedback. In addition, there were other stakeholders that were not part of these teams, but had an interest in the outcome of the research. Managing the communication amongst all of these stakeholder was sometimes challenging. There were times when certain teams or stakeholders did not receive timely updates or have the opportunity to provide input. Because this project spanned many years, there was movement on and off teams that was also a challenge. Teams were being formed and re-shaped as managers tried to meet the needs of the work.

Research plans were also constantly in flux. We knew the overall goals and which methods we wanted to utilize and in which order, but schedules were constantly being adjusted because we were dependent on results from previous rounds. We were also adapting to the schedules and requests of our programmers. During prototype development, it was difficult to predict the amount of time that would be necessary or the amount of input that stakeholders would provide. There were rounds where prototype development would go quickly and the design would have a lot of support. There were other rounds where the prototype discussion lagged and development was delayed. Some of these delays were a result of stakeholders and respondents not wanting to give up certain features that were present in the software.

Our respondents are not one size fits all. As mentioned earlier, we knew a priori that the response process amongst companies of different sizes can be different. There was a strong need to develop a 'one size fits all' design. For years, the economic census had been supporting two electronic designs and it was logical to desire one tool for all respondents. Trying to develop a tool that met all of the needs was a challenge, and in the end two paths were developed within the same instrument that allowed for some customization for small companies, but maintained many shared features for all.

Other challenges that we faced included finding resources, typically financial, in order to conduct necessary research in person with businesses. Research schedules sometimes conflicted with the typical government budget schedule and we sometimes found ourselves dealing with a continuing resolution situation that was a challenge for the managers who were responsible for finding the necessary funding for travel. At times, we were able to find less costly local solutions.

Another challenge that we faced was fitting our work into the typical production schedule of the ASM and COS. These schedule often dictated, and sometimes conflicted, with the research schedule. The production schedule also meant that at times stakeholders needed to focus on production and couldn't provide as much focus on the research.

Managing and coordinating the research across the various different efforts was a challenge. There were separate threads of research that needed to come together into one instrument at the end. Different researchers were assigned to focus on these separate threads in order for them to become experts and not be overwhelmed. It was challenging to keep researchers informed of all these different pieces early on. It was critical that they stayed informed about the issues and plans for each related research effort. One way that we managed this was to hold weekly status update meetings for everyone involved. Where possible, researchers would work on or provide some form or support for each other's projects.

Lessons Learned

Overall, the research was successful. There were lessons that we learned that were positive and negative along the way.

Prototyping allowed programmers and researchers the ability to assess early design ideas before spending resources on full development. This was especially useful when programmers were debating about whether the Microsoft SharePoint platform could be used to collect the economic census. An early semi-function prototype was built in SharePoint and tested with respondents while staff were still debating whether it could work with internal systems and support all of the requirements necessary for the secure electronic collection of an economic census. Soon after testing it was decided that this was not the ideal platform for respondents based on a number of reasons from

respondents, internal staff, and systems. The use of an early prototypes allowed us to assess SharePoint's viability without a significant amount of time and financial investment.

Using prototypes also encouraged stakeholders working closely with researchers the ability to be creative. Researchers were able to discuss a variety of design ideas with respondents without much investment in resources. Researchers and stakeholders learned a lot from the trial and error of evaluating a variety of designs while working towards the final design.

Knowing far in advance that the economic census was moving to an all Web collection allowed researchers and programmers time to discuss and test a variety of ideas, including the Microsoft SharePoint idea that wasn't implemented.

Efforts started out separately for small companies and those with multiple locations and eventually merged as the design for those two respondent populations came together. In the beginning, one researcher was in charge of the smaller company's web application design and another was in charge of the multiple location design. This allowed each researcher to focus on their respondent population and become an expert and advocate on what was needed for that particular type of respondent. Additional researchers supported both small and multiple location company research over the course of time. Weekly meetings between all researchers helped keep everyone in the loop and allowed for an easier adjustment when the design merged into one instrument for all respondents.

Using multiple methods over the last four years allowed us to incrementally identify what worked best for respondents and has resulted in a final design that respondents should find usable.

References

Moore, K. & Samples, W. (2018). An Overview of the Improved 2017 Economic Census. 2018 FCSM.

Nichols, B., Olmsted Hawala, E., Holland, T., and Anderson Riemer, A.: Best Practices of Usability Testing Online Questionnaires at the Census Bureau: How Rigorous and Repeatable Testing Can Improve Online Questionnaire Design. In: 2nd International Conference on Questionnaire Design, Development, Evaluation and Testing. Miami, Florida November 9-16, 2016.

Snijkers, G. Haraldsen, G. Jones, J. and Willimack, D.: Designing and Conducting Business Surveys. Wiley, New York (2013)

Tuttle and Beck (2014). Findings and Recommendations from Debriefings of Surveyor Users for the 2012 Economic Census. Internal Report

Usability.Gov Prototyping, https://www.usability.gov/how-to-and-tools/methods/prototyping.html, last accessed 2018/05/15.