Using Software to Collect Data Electronically for the Economic Census
Amy E. Anderson, M. Diane Harley
U.S. Census Bureau

Abstract
The economic census is a complex data collection process that encompasses over 500 different forms. In 2002 the Surveyor software application was built to handle all forms and accommodate companies with anywhere from 1 to 2,000 establishments. The U.S. Census Bureau has been working for several years to re-engineer the 2002 software version in preparation for the upcoming 2007 Economic Census.

This paper describes the data collection software and changes that are being made for the next economic census. It will discuss the Census Bureau’s decision to use a software application to collect data for an economic census and highlight research that has been conducted in collaboration with the redesign effort. This research has been a key resource for design decisions made for the new version of the software, as well as further illuminating the economic census response process.

Keywords: software, usability, economic census, establishment surveys

1. Introduction
Electronic reporting methods for establishment surveys have generally focused on Web data collection. While this currently remains the most common form of electronic data collection, it is not the only form in existence. There may be characteristics about the survey being collected and/or that survey’s respondents that make data collection over the Web a less than ideal option. At the U.S. Census Bureau, the economic census is one such example. For many companies the economic census is a complex survey. For a single-location company, completing the survey could involve a half-hour of digging through records. For a diverse company with thousands of locations, completing the survey could involve weeks of coordinating a data collection effort between multiple systems and people in different offices or locations.

2. Background

2.1 History
Before the creation of a software package to collect data electronically, some respondents worked with the Census Bureau to transmit data on various electronic media, such as magnetic tape or through electronic data interchange (EDI). These alternative electronic reporting modes were used almost exclusively by the largest companies, and typically involved arrangements made with the Census Bureau prior to data collection (McDonald, 1996). Although these alternative methods met large companies’ needs for electronic reporting, they required intense efforts on the Census Bureau’s part. Moving respondents to an electronic mode that would seamlessly work with the Census Bureau’s systems became a priority.

The Census Bureau’s first experience with software for establishment data collection was in 1993 for the Company Organization Survey (COS). At this time, Web surveys were in their infancy. The Census Bureau did not begin online data collection for establishment surveys until 1999. Online security was a main concern in the early days of Web reporting. Most computer browsers in the 1990s did not meet the encryption levels required by the Census Bureau. To enhance the encryption level on a respondent’s browser, the Census Bureau had to issue digital certificates, which were costly to the Census Bureau and required effort on the part of respondents to obtain and maintain. To meet the growing respondent demand for electronic reporting and to avoid the pitfalls associated with early Web reporting, the Census Bureau began developing software to collect several of its more complex surveys, including the economic census.

Most of the surveys collected using our software package, which is called Surveyor, are establishment based, rather than company based, surveys. Establishment based surveys are typically more burdensome than company based surveys due to the fact that multiple establishments within a single company could be included within a particular survey’s sample.

2.2 The Survey
The economic census is the largest economic data collection program in the United States. The census is conducted every five years and includes a large portion of the establishments in the U.S. Beginning in 2002, all respondents had the option to report data via paper or electronically through our Surveyor software. The data collected from the economic census are used as a main input into the calculation of the U.S. Gross Domestic Product (GDP). The data are also used widely by government officials, policy makers, academics, and by businesses themselves.
Companies can be assigned to any number of the more than 500 different form types available for the economic census.

Each form type corresponds to an industry or a closely related set of industries. The more diverse a company is, the more form types their locations may be assigned to.

2.3 The Software

The software used to collect the economic census has evolved over the last 10 years. Prior to the Surveyor software, electronic data collection during an economic census was limited to EDI, flat file reporting, and a software application created for the 1997 Economic Census that collected data only for a select number of retail industry forms.

The Surveyor software application was originally developed to collect data for the 2002 Economic Census. The software also supports electronic data collection from three other annual surveys: the Report of Organization, the Annual Survey of Manufactures, and the Survey of Industrial Research and Development. The first two surveys collect data at the establishment level, while the Survey of Industrial Research and Development collects data at the company level.

The redesigned Surveyor software includes five different screens that respondents can toggle through using tabs at the top of the user-interface. The following sections provide a description of each screen within the software as well as a description of a key function within the software that allows users to export spreadsheets from the software and import spreadsheets into the software.

2.3.1 Welcome Screen

When respondents first open the software, they are taken to the Welcome screen. This screen contains basic information about completing the economic census and using the software.

2.3.2 Inbox Screen (Appendix A)

The Inbox screen displays a listing of all the establishments/forms assigned to a company. This screen contains basic information about each establishment. The Inbox screen also contains information about the number of edits or warnings that each locations has, as well as the status of each form (i.e., in progress, submitted, not started). Respondents can use sorting and filtering tools built into this screen to search for particular locations.

2.3.3 Form Screen (Appendix B)

The contents of the form are displayed in this screen. In general, the electronic version of each form closely resembles the corresponding paper form. Past research has found that many establishment survey respondents who choose to report electronically still use the paper forms to help gather data (Morrison, 2005).

Edits are built into the software to check for missing data, miscalculated data, or invalid data relationships. These edit checks fall into one of two categories: error or warning. Errors must be resolved in order for a respondent to successfully submit their data electronically. Data associated with an warning can be submitted without being resolved. Respondents can view the details of their errors or warnings for each form/location at the bottom of the screen.

2.3.4 Workbook Screen (Appendix C)

The Workbook screen displays survey data in a spreadsheet layout. This alternative view of the data gives respondents the ability to input and view their information in a different format. When information in this screen is updated, its corresponding fields on the Form screen are also updated. Using the tabs at the top of the screen, respondents can ‘toggle’ seamlessly between each screen in order to input or view their data.

Rows correspond to each individual location, and columns correspond to each data field in the form. Errors and warnings are also displayed at the bottom of the screen.

2.3.5 Errors and Warnings Screen (Appendix D)

This screen displays all of the errors and warnings that respondents have incurred for all locations. Respondents can sort the information on this page in several ways, including by location or by type of edit. This screen enables respondents to choose the most effective way to deal with potential problems with their data. Selecting an edit on this screen will take respondents back to the corresponding item within the Form screen.

2.3.6 Exporting and Importing

A key functionality built into Surveyor is the ability of the software to generate a spreadsheet (exporting) and accept a spreadsheet back into the software (importing). Each exported and imported spreadsheet contains the establishments for only one specific form type. Diverse companies that are assigned to multiple form types may need to coordinate their data gathering effort among several spreadsheets.

Analogous to the Workbook screen, exported spreadsheets contain one row per location within the selected form type. The columns correspond to each of the data fields in the form. Any data, such as pre-listed address information, that exists within the software during exporting will be included on the spreadsheet. Respondents can choose to import the software-generated spreadsheet, or they may choose to import their own spreadsheet, as long as it meets the software’s importing rules.
The exporting and importing feature is extremely popular with respondents from large companies. These respondents have the ability to merge the data from their internal systems with the exported spreadsheet. This can be done in a variety of ways depending on the type of systems within the company and the sophistication of the respondent in charge of completing the spreadsheet.

2.4 The Research

Several research projects were conducted in preparation for re-engineering Surveyor for the 2007 Economic Census.

2.4.1 Spring/Summer 2003
During data collection for the 2002 Economic Census, observational visits were conducted with several respondents who were reporting electronically with the original version of Surveyor. The purpose of this research was to evaluate the respondent burden incurred during an economic census as well as to gather respondent evaluations of the response task (Hak, 2003).

2.4.2 Fall 2003
Debriefing interviews were conducted with respondents who completed the 2002 Economic Census electronically. The purpose of these interviews was to learn about respondents’ experiences with the economic census and the software.

2.4.3 Fall 2004 through Fall 2005
Based on the findings and recommendations from prior research on Surveyor, and on internal experiences with the original Surveyor software, re-engineering of the software was proposed. To accomplish this, software developers needed detailed user requirements to guide changes. To gather these detailed user requirements, a year-long research project was developed which included user-centered design testing methods such as task analysis and use of prototypes. The research resulted in over 75 company visits and numerous user requirements.

2.4.4 Summer 2006 through Fall 2006
Surveyor programmers created a partially functioning electronic prototype based on the detailed user requirements that had been gathered during prior research. Two rounds of usability testing were conducted using this prototype.

3. Research Highlights

Several internal and external papers have highlighted the major findings from the research on Surveyor (Anderson, 2005; Morrison, 2005). This paper will highlight the most significant findings from the user-requirement and usability testing research.

Task analysis information enabled us to better understand how the response process is impacted, and sometimes modified, when respondents deal with an electronic data collection instrument.

Respondents who report data for the Economic Census typically do not have access to all of the requested information. To complete the data request, respondents often work with others within their companies. Most of the modifications made to the data gathering process for the electronic mode deal with the use of spreadsheets instead of paper forms as a tool for gathering data. Typically, respondents are either using spreadsheets generated by the software, or they are creating their own spreadsheets, in order to pull together information from different sources.

In addition to using spreadsheets as a tool for gathering information, respondents use spreadsheets in other capacities:

- **Comparing data** - One of the first tasks that multi-establishment companies do after receiving the economic census is to reconcile the list of locations that the Census Bureau has on record for a company with the actual locations that are active within the company. Respondents use spreadsheets to complete this reconciliation. Locations that do not reconcile usually indicate a newly opened or closed location. Some reconciling mismatches are due to typographical differences or incorrect address information.

- **Consolidating data** – Respondents for the economic census typically need to obtain data in pieces from different internal computer systems or from co-workers. Once these pieces are compiled, respondents use spreadsheets to consolidate the information into one “master” spreadsheet.

- **Verifying data** – Upon completion of the forms, respondents may choose to make comparisons between the data gathered and known aggregate totals in order to verify that the information being reported is reasonably accurate. Mathematical tools built into spreadsheet programs help respondents accomplish this.

- **Correcting data** – After Surveyor has identified errors and warnings, respondents will either make the correction within the software or return to their original spreadsheet. Once the spreadsheet has been updated with the correct information, respondents will then re-import the data into the software and keep this updated spreadsheet for their internal records.

4. Major Software Changes

4.1 Additional Screens Developed

During the research to gather user requirement, respondents from large companies pointed out several
deficiencies within the software. Based on these findings, new screens were added to the existing user interface. Once developed, these screens were then tested using a prototype during the usability testing phase of our research. The usability testing phase gave researchers the opportunity to determine if the newly incorporated screens were adding value to the software as a whole. It also allowed researchers to identify additional changes needed for the new and existing screens.

4.1.1 Welcome Screen
Surveyor contained an initial pop-up screen with introductory information about the survey as well as the software. During user requirements gathering research, respondents made several suggestions about including a screen upon initially running the software that contained basic information about the features within Surveyor. In essence, these respondents were describing what was already built into the software in that initial pop-up screen. During further investigation, respondents stated that they either forgot that such a screen existed, or they closed it so quickly that they did not realize that it existed.

To make this information more visible to respondents, developers placed this information in a separate screen instead of in a pop-up window. Although this wasn’t incorporated in time to be usability tested, we expect this new design to answer the needs of respondents.

4.1.2 Workbook Screen (Appendix C)
The respondents from large- and medium-sized companies that took advantage of the importing feature were disappointed that the software did not include a visual overview of the data. This overview is important for respondents because it allows them to confirm that the importing process successfully placed their data in the right locations and in the right format. Respondents who imported their data were somewhat wary of the process and usually felt the need to verify that the data had successfully imported. Rather than going through the burdensome task of viewing every piece of imported data, respondents would open the forms for randomly selected locations. In addition to the problems with verifying imported data, respondents were also interested in having a screen that would permit faster data entry.

The spreadsheet format adopted for the Workbook screen solves both problems. It is a quick ‘view’ of the data in a format that also allows for input. The reaction to this screen during usability testing was overwhelmingly positive. Large companies that use the importing feature felt that this screen adequately answered their need for an overview of the data.

In addition to ‘eyeballing’ their data to ensure that it imported correctly, respondents would be able to address data flagged by errors or warning more effectively than in the previous economic census. Respondents had only a limited amount of time to dedicate to the review of errors and warnings. Data in error got top priority and were always addressed. Respondents would focus on the data associated with a warning only when extra time was available. Respondents from large companies stated during the research that they would often have to leave warnings within their data because they did not have enough time to go through the software and open each individual form to investigate warnings, which was the only way of viewing errors and warnings in the prior version of the software.

Respondents from smaller companies also found the Workbook screen useful. This surprised researchers since most of the focus had been on improving the response process for large companies. The Workbook screen had been built with them in mind. Smaller companies felt that they could derive the benefits of a spreadsheet from the Workbook screen without having to work with the exporting and importing features within the software. The major attraction to this screen was the ability for respondents to quickly enter data. Smaller companies did note that they would need to work with the Form screen initially to understand the data that was being requested before shifting their data entry to the Workbook screen.

4.1.3 Errors and Warnings Screen (Appendix D)
Respondents are notified on Surveyor’s Inbox screen of the number of errors and warnings associated with a location. Before Surveyor was redesigned, to view the details of these errors and warnings, respondents had to open each location’s form in the Form screen or hunt within the Workbook screen. The task was increasingly burdensome, even prohibitive, for companies with hundreds, or even thousands of locations.

To aid respondents in working with their errors and warnings, researchers recommended the addition of a screen that would give a global view of all edit check failures for all locations. Respondents were very happy with this new screen when presented with it during usability testing.

4.2 Exporting and Importing Overhauled
The exporting and importing feature was the most contentious of the features discussed during user requirements gathering. Respondents were very forthcoming about the struggles they faced with not only the functionality of this feature, but with the layout and design of the user interfaces associated with it. Respondents recounted several examples from the past where they ran into problems or confusion related to the export or import functions.
In the 2002 version of the software, respondents began the exporting process by “mapping” or linking each field on a particular form to a column on a spreadsheet. This mapping process was done using an interface built into the software. Once the spreadsheet was completely mapped, respondents could then export it. The exported spreadsheet for the 2002 Economic Census contained a row for each location and a column (denoted by a title) associated with each field on the form. Columns or rows within the spreadsheet could not be deleted or moved from their original location or else the subsequent importing process would fail.

During the 2003 debriefing visits and the user requirements gathering research, respondents pointed out several problems they encountered with these features. Respondents’ difficulties started at the very beginning of the process in the mapping user interface. The interface was difficult to understand and navigate. Several respondents stated that they became so exasperated with the mapping interface that they abandoned the exporting and importing functionality completely. Those that used the interface spent long periods of time completing the tedious mapping process. And since this process was necessary for each form type a respondent was assigned to, diverse companies assigned to multiple form types faced additional burden.

Once exported, respondents ran into problems with the restrictions placed on the spreadsheet. Several did not become aware of the restrictions on moving or deleting rows and columns until their attempt to import failed. At this time, respondents would have to go back to the spreadsheet and try to return it to the original format. Sometimes this task was so difficult that respondents started over with a clean exported spreadsheet.

Respondents that worked with the exported spreadsheets were confused about the formatting rules for entering data. Some questions, such as multiple choice questions or questions asking for a date, required a different format than items asking for a numerical figure. These formatting instructions were not easily accessible to respondents and often required some searching.

The revised Surveyor software includes many enhancements to overcome the pitfalls previously experienced during exporting and importing. Improvements include:

- Elimination of mapping – After choosing a form type, the software will export a pre-mapped spreadsheet.
- Spreadsheet customizing – A planned overhaul of the importing feature will allow respondents the ability to move and/or delete columns and rows within the exported spreadsheet.
- More informative spreadsheets – The spreadsheet will now contain multiple worksheets. The first worksheet will contain welcome information and instructions related to working with the exported spreadsheet. Data entry worksheets will contain instructions within the second and third rows that give more descriptive information and relevant formatting instructions.

4.3 Redesigned Inbox Screen (Appendix A)

In addition to adding new screens to the user interface, modifications were made to the existing Inbox screen. During the user requirements gathering research, we collected information from respondents about their interpretation and use of each piece of information on the Inbox screen. We discussed the benefits and potential drawbacks of each piece of information. Respondents highlighted what was of little value to them, as well as items on the screen that were problematic. Respondents also pointed out the pieces of information that were the most important for them and how those pieces of information should be presented to be more helpful.

The redesigned Inbox screen includes an updated layout of the information presented. Problematic and unused information has been eliminated and important information, such as error and warning counts, have been moved to a more prominent position on the screen.

A new feature added to this screen permits respondents to create subfolders, similar to Windows Explorer. This feature will allow respondents with multiple locations to view those locations by whatever sub-division they prefer (business unit, region, type of operation, etc.). Generally, the redesigned information in the Inbox screen, and the added folder feature were positively received by respondents during usability testing.

5. Discussion

The use of software to collect a massive survey such as the economic census has proven over the years to be an effective means of data collection for many respondents. However, with any mode of data collection, there are pitfalls. Some of these challenges include transitioning smaller or single location companies from paper to electronic, convincing very diverse companies that they may reduce their burden by reporting electronically rather than on paper, and dealing with companies that run into Information Technology (IT) obstacles when downloading the software.

Uptake of electronic reporting for small companies and single location companies was significantly lower than larger companies during the 2002 Economic Census. Downloading software is an added burden to companies...
that choose to report electronically. For larger companies, this initial added burden will turn into less overall burden in the end because of the features built into the software, such as exporting and importing. Smaller companies may feel that this initial added burden is not worth sacrificing what little time that they have to complete the survey. Allowing Web reporting for these smaller companies could help to reduce the overall burden of electronic reporting for the economic census and may entice smaller companies to choose this as a mode of reporting. The Census Bureau is considering this option for future censuses.

Companies that are diverse by industry or by internal structure have often been reluctant to move to electronic reporting. Many of these respondents feel that the extra time required to set up the software and coordinate the reporting is not cost effective. These companies feel that working with paper is still the quickest and easiest way to gather the data. A challenge for the Census Bureau is to learn more about the response process within these diverse companies in order to discover ways to fit the software into that response process, or to identify added features that could be incorporated into future versions of the software to attract these companies to electronic reporting.

Companies across the country are locking down their computer networks and creating stringent computer-use rules for their employees to protect their companies and their systems from costly viruses and bugs that can be transmitted through software or the Internet. Because of this, downloading our software has become an issue for many respondents. Several respondents have to work with an IT staff in order to gain approval and access to the software. This extra response process step is heavily dependent upon the company’s IT resources, and is completely out of the control of the Census Bureau. One of the steps that the Census Bureau will be taking for the 2007 Economic Census is to work with the largest companies prior to the mail-out of the survey to identify those that will have to work with an IT staff. Every effort will be made to alert respondents to this issue and to encourage them to start working with their IT staff early, thereby avoiding problems with missed deadlines.

The landscape of electronic reporting for the economic census is going to be one that continually changes over time to adapt to advancing technology for electronic reporting, as well as the advancing technology within the companies themselves. As company’s computer systems change and become more sophisticated over time, the Census Bureau will need to rise to the challenge of understanding the response process within companies and adapting our electronic reporting mode, or modes, to that response process in order to maintain our position as the leading source of quality data for the United States economy.

6. References


### Appendix C – Workbook Screen

![Workbook Screen](image)

### Appendix D – Errors and Warnings Screen

![Errors and Warnings Screen](image)