Applying Knowledge of Business Survey Response Processes to the Design of Data Collection Software at the U.S. Census Bureau

Amy E. Anderson
Rebecca L. Morrison
Charles F. Brady

U.S. Census Bureau[1]

Presenter:
Amy E. Anderson
amy.e.anderson@census.gov
U.S. Census Bureau
Room 3108-4 Mail Stop 6200
Washington, DC 20233-6200
Phone: 301-763-7544
Fax: 301-457-2304

Key Words: Establishment surveys, Electronic reporting, Software requirements, Response process

Abstract: The U.S. Census Bureau gathers establishment data electronically for several surveys using data collection software. This system, called Surveyor, was initially created for the 2002 Economic Census, which was the first economic census in which an electronic reporting option was offered to all eligible business respondents. Surveyor was later adapted to collect information from other surveys during non-census years. In order to re-engineer the existing Surveyor system for the 2007 Economic Census, software developers requested detailed requirements. The Census Bureau augmented its own "lessons learned" from the 2002 Economic Census experience by developing investigative partnerships with business respondents in order to identify software requirements that meet the needs and expectations of these respondent users.

The goal of this research was to delve deeper into the business survey response process to obtain detailed descriptions of respondents' data collection tasks (specifically for the economic census), such as the types of people involved and the location of data. After identifying these tasks, we translated them into requirements for the software. Our goal is that, by working closely with actual respondents, we will enhance the software to dovetail with the respondents' established data collection routines.

This paper will describe how we gathered information about the data collection tasks business respondents engage in and how we applied that knowledge to develop requirements for software enhancement. We will also discuss the challenges we faced in creating requirements that would work for the many different types of response processes used by businesses.

[1] This report is released to inform interested parties of research and to encourage discussion. The views expressed on methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

I. Introduction

In preparation for re-engineering the software, called Surveyor, that was used to collect the 2002 Economic Census, researchers at the Census Bureau conducted a year-long study with business respondents in an effort to uncover detailed user requirements. These requirements will be used by the Census Bureau and its software contractor to make respondent-centered design modifications in preparation for the 2007 Economic Census. The goals for the redesign of the Surveyor software are to enhance the user-friendliness of the instrument and to provide functions that dovetail with users' response processes.

In order to gather user requirements, researchers conducted four rounds of testing with business respondents recruited for voluntary participation. Each round of testing investigated different topics related to software features and included exercises and low-fidelity prototypes to facilitate discussions. Our paper will describe the methodology that we used to gather information from our respondents, selected findings, and how those findings were translated into user-requirements for the software.

II. Background

A. The Economic Census

The economic census is collected every 5 years from all non-employer establishments operating in the United States. The economic census collects detailed establishment level data including such items as annual payroll, total employment, detailed and total sales/shipments/receipts/revenue, class of customer, expenses, inventories, cost of materials/parts/supplies, and information about newly closed, sold, or opened establishments. These data collected are used by the Bureau of Economic Analysis to estimate the Gross Domestic Product. The data are also used for various purposes by government officials, politicians, academics, the publics, and even the companies that the data is collected from.

B. The Software

Surveyor is a software system that was developed initially to collect data for the 2002 Economic Census. Subsequent to the 2002 Economic Census, the software has been used to collect data from the annual Report of Organization, the Annual Survey of Manufactures, and the annual Research and Development Survey. Like the economic census, these surveys collect data at the establishment level, rather than at the overall company level. The software is built to accommodate electronic reporting from single-establishment companies as well as very large companies with thousands of establishments. Nevertheless, because the economic census is a complex data collection, the features of the software are relatively sophisticated.

An executable file containing the software can be downloaded onto the respondents' own computer either from a CD-Rom or from the Internet. Respondents are invited to go onto the Internet upon receiving the paper forms or the survey notice letter to download the software. Arrangements can be made for respondents to receive the software on CD-Rom. Along with the executable file, respondents also download files in extensible markup language (.xml) format.

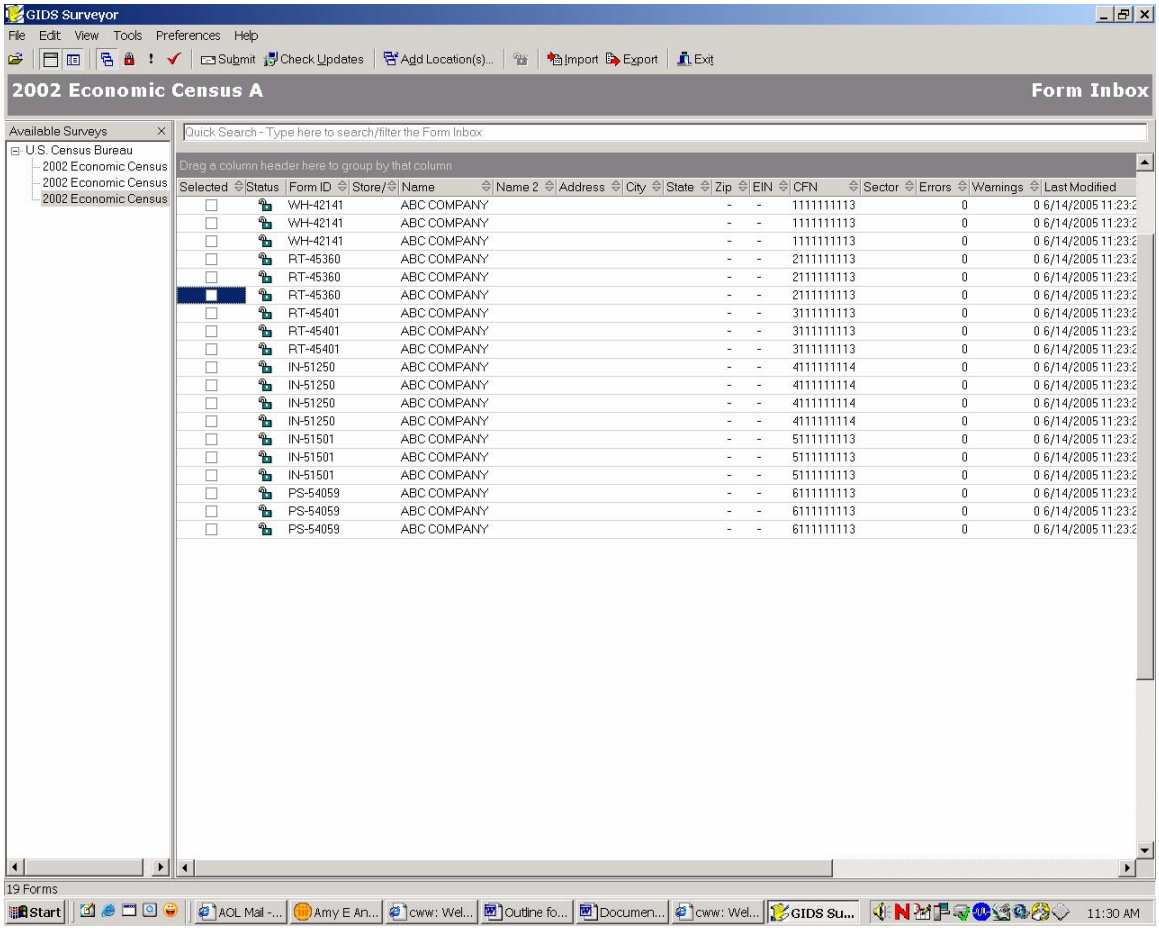
These files contain company specific information that is usually printed directly onto paper forms in the mailing label (i.e., location name, address, Employer Identification Number, etc.).

There are two main components of the software: the Form Inbox and the forms themselves.

1) The Form Inbox
The Form Inbox (see Figure 1) is the main menu of the software. Respondents use the inbox to select the forms they want to work with. The inbox displays establishment information that can be sorted, filtered, or grouped. The number of errors and warnings for each establishment is also displayed after data are entered.

Respondents use the Form Inbox to add new establishments, check for data or software updates, and to submit data.

Figure 1. Form Inbox Screen Shot from Surveyor



Two other functions available through the Form Inbox allow respondents to create and accept spreadsheets: exporting and importing. The exporting functionality can be used to create a spreadsheet containing company specific data. The exported spreadsheet contains this data along with any other data that the respondent may have key-entered into any of the forms prior to exporting.

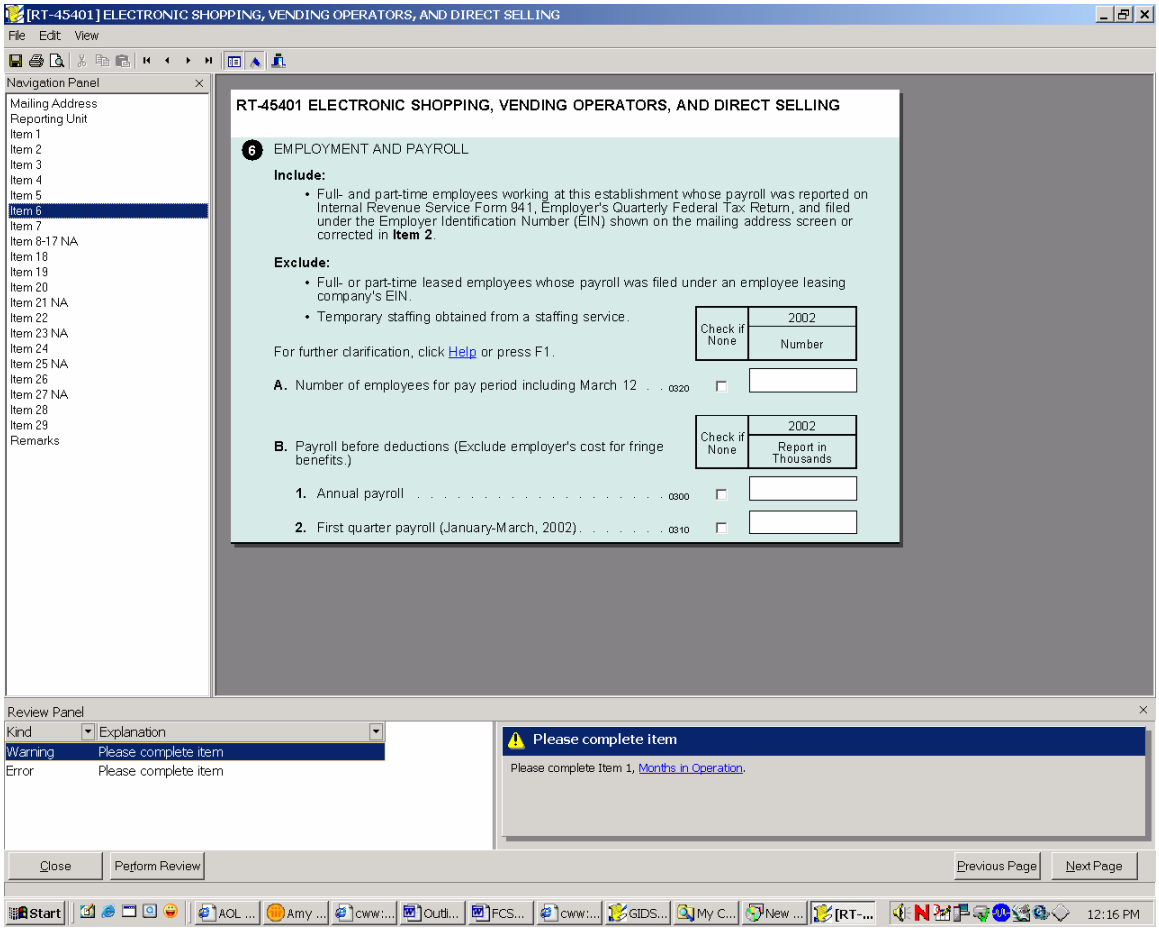
Respondents can import data from a spreadsheet into the software. They can import this information using an exported spreadsheet, or using a spreadsheet that has been "mapped" to the software. Through mapping, respondents take an existing spreadsheet and match columns of data to the comparable fields within the software. Respondents must also use the mapping process to format a blank spreadsheet for exporting.

2) The Forms

The forms display screen (see Figure 2) is made up of three different parts: the navigation panel, the form display, and the review panel. The navigation panel is displayed on the left and allows respondents to move from item to item. Respondents may also navigate through these pages using the Next and Previous Page buttons located at the bottom right-hand side of the screen, using Page Up and Page Down keys on their keyboard, accessing Next and Previous Page options under the View menu, or using navigation buttons displayed on the toolbar. The form display portion of the screen displays the selected questionnaire item.

The review panel is located at the bottom of the screen and is divided into two sections. The section on the left displays the kind of edit (error vs. warning) and a short explanation. The respondent may sort their edit messages by kind or explanation. The section at the bottom right side of the screen displays the detailed description for the edit message highlighted on the left side. The messages contain hyperlinks that respondents can use to navigate to the item associated with the edit. When the respondent performs a review of the data, a flag icon appears next to the "edit failing" item. The respondent can then click on the flag to receive a detailed description.

Figure 2. Form Display Screen Shot from Surveyor



C. The Response Process in Establishment Surveys

Traditional models of the survey response process developed before the advent of electronic data collection were, naturally, predicated on the assumption of interviewer-administered or self-administered paper questionnaires. Additionally, to accommodate the establishment survey setting, Tourangeau's (1984) original response process model, which contained four cognitive steps -- comprehension, retrieval, judgment, and communication -- required some adaptations. Most notably, Sudman *et al* (2000) proposed a hybrid model incorporating steps occurring at the organizational level, building upon prior research by Edwards and Cantor (1991) and Tourangeau. Sudman et al's eight-step response process for establishment surveys included the following steps:

1. Encoding of information in company records
2. Selection and identification of the repondent(s)
3. Assessment of priorities
4. Comprehension of the data request
5. Retrieval of relevant information from existing company records
6. Judgment of the adequacy of the response
7. Communication of the response

8. Release of the data.

Electronic data collection adds factors that must be taken into account, vis-à-vis the response process, especially the need for a user-centered design, or "usability", of the interface with the features of electronic instruments. Usability's "goals are to understand users, develop good user-oriented design principles, apply them to the design of the human-computer interface, and then to test the product to make sure the interface is likable." (Marquis et al, 1998) Nielsen (1993) listed the attributes of user-centered design as: learnability, efficiency of use, memorability, few and noncatastrophic errors, and subjective satisfaction.

Understanding the response process and the data gathering steps undertaken by respondents helps survey organizations to design paper or electronic instruments that fit the respondent's data gathering process.

IV. Methodology

From August 2004 to July 2005 we gathered detailed user-requirements from respondents using user-centered design testing methods. We divided our research into four rounds lasting approximately three months each. Our respondents were also divided into two types of panels: longitudinal and rotating.

A. User-Centered Design Testing Methods

User-centered design testing is done with various methods. For the purposes of this paper, two are of particular interest – task analysis and prototypes. Task analysis should be done as early as possible in the testing process, as it provides early input into how the system should be designed (Nielsen, 1993). The benefit of engaging in this method of research is that it provides information about the tasks respondents need to perform (Usabilitynet.org, 2005). Typically, as part of a task analysis, a researcher meets with respondents to discuss their methods for completing specific jobs. While asking questions about what users expect from an electronic system, and what they think they will need in order to complete tasks electronically, the researcher can probe about how and why certain actions are taken by the respondent.

Prototypes are used to "save on the time and cost to develop something that can be tested with real users." (Nielsen, 1993) Prototypes could be as simple as a paper mock-up (also known as "low-fidelity" prototoypes) or as complex as a functional computer interface. Either way, a prototype moves beyond task analysis in that it is a more concrete visual representation of how the interface will look and/or work.

As a researcher begins to understand how respondents typically work their way through specific tasks, and how they expect the instrument to work and/or look, she can better piece together how to fit the instrument to the tasks the respondent must complete. To the extent possible, it is best to design the instrument to meet

the users' expectations of how a questionnaire operates, as well as how an electronic interface operates (Dillman *et al*, 1999). When users face an instrument that behaves differently than expected, they will be led down "unproductive paths though the underlying data structure, creating frustration and workarounds to minimize use of the tool." (Murphy *et al*, 2000) Determining what the user is expecting the interface to accomplish, and then designing the interface in such a way so that those expectations are met, serves to reduce frustration and burden.

B. The Panels

For our research, respondents were selected for one of two panels: rotating or longitudinal. The longitudinal panel was comprised of approximately 20 companies that we visited once every round. This panel was formed to give periodic feedback about the software and prototypes used to represent proposed software modifications.

The rotating panel was comprised of new companies not visited during previous rounds. These respondents offered a fresh perspective compared to the longitudinal panel members who had been working with the software over time.

Panel members were selected to ensure a mix of companies in both service and manufacturing industries. We mainly targeted large companies for our longitudinal and rotating panels, but did include three medium sized companies in our longitudinal panel, and several others in our rotating panel.

C. The Rounds

Different topics were discussed during each round. Dividing up the topics into rounds allowed us the ability to discuss topics in-depth and to test prototypes created based on feedback from previous rounds of testing.

During an initial visit with a company, either longitudinal or rotating, we conducted a task analysis to identify how the data are gathered within the company for the purpose of completing the economic census. The main topics we investigated during the task analysis included:

- the types of people who were involved, what departments they represented, and what types of information they provided, where data were kept,
- who had access to the data,
- whether the respondents' technical support staff got involved; and
- how they moved data from their own records and systems into the software.

During the "initiation" visit with longitudinal panel members, the task analysis included several in-depth probes related to the topics listed above. The goal was to obtain detailed information. Rotating panel members were asked similar

questions with less probing. The goal with rotating panel members was to get an overview of their reporting process and then move on to the core interview topics.

V. Findings and User-Requirement Recommendations

Learning about the companies response process and how it can impact reporting electronically turned out to be the most valuable finding from our research. The next step was to translate these diverse response process tasks into user requirements that will be valuable for most of the respondents using the software.

A. Data Gathering by Respondents

Data gathering for surveys such as the economic census typically involves collaboration among various personnel within a company to pull together all of the necessary information requested. Respondents to the economic census typically fall within one of three categories: gatherers, compilers, coordinators, or some combination of the three.

Gatherers are those respondents who have complete access to the data that is requested on the forms. They receive the forms, complete the forms, and submit the forms. These respondents may have a superior review the data before it is released to the Census Bureau.

Coordinators and compilers are those who are responsible for the completion of the forms, but do not have direct access to all the necessary data. They are often found in companies where the requested data are decentralized. The distinction between a compiler and coordinator was first described by Nichols et al (1999). "Coordinators" coordinate data collection by sending forms and instructions out to other data providers within the company. Compilers, on the other hand, are respondents who request data from others who have direct access. Both coordinators and compilers must go through the task of identifying what data are necessary along with which "local data provider" (LDP) within the company that can supply this information. LDPs associated with coordinators typically complete the actual forms. They may then submit the form directly to the Census Bureau or they may return completed forms to the coordinator. In contrast, compliers obtain information from LDPs and enter this data onto the proper forms themselves.

Although these roles were defined within the confines of paper-based reporting, there are analogous distinctions when moving to an electronic mode. We found that many respondents in large and medium sized companies often choose to gather data via spreadsheet and then import it into the software. Respondents who must gather data for their spreadsheets from other LDPs within the company can still be categorized as coordinators or compilers. Coordinators will send the final spreadsheet (or a copy of the final spreadsheet) to their LDPs along with instructions on how to format the data on the spreadsheet. The LDPs must complete the spreadsheet and send it back to the coordinator. Compilers, on the other hand, take the data that is provided to them by their LDPs, oftentimes

provided on a spreadsheet in the LDPs format, and fit it into a final spreadsheet. Compilers may have to re-format data provided electronically by LDPs to match the format required by the software. Those respondents that choose to forego spreadsheets altogether and key enter data provided by an LDP into the actual electronic instrument are also considered compilers.

Some respondents perform both gatherer and coordinator/compiler roles. These respondents have access to a subset of the data requested, but must go to LDPs for the remainder. Most of the respondents in our study acted in multiple roles, as gathers and coordinator/compilers.

B. Spreadsheets

The one common factor between all of the companies, regardless of how they gathered the data, was the use of spreadsheets. All of our respondents were using them. Whether it was one single respondent at his/her desk who had access to everything (a gatherer), or a coordinator/compiler dealing with multiple respondents who had to balance several spreadsheets filled with data.

Many respondents stated that they wanted to "get out of the software as quickly as possible" and into a spreadsheet environment. Respondents were interested in putting their data into spreadsheets that met our format, rather than just passing along spreadsheets in whatever format suited them. They wanted to use the software as a tool for creating a spreadsheet formatted to meet our specifications.

Spreadsheets were used for the following tasks:

- Consolidate data: None of the respondents that we worked with were able to pull the entire contents of the Economic Census out of one internal database. Even if we were dealing with a gatherer, a respondent who had access to all the requested data, the data were dispersed. Payroll data were housed in a different system than financial data. Sometimes location information (current addresses, or information about newly closed or opened establishments) could be found in the payroll or financial data systems, or in an entirely different system. Respondents would gather this data in pieces and then use cut, paste, or copy options to move them into one consolidated spreadsheet that would eventually get imported into the software when it was complete.

- Gather data: Coordinators who needed to gather data from LDP's would use spreadsheets as the tool for gathering the data. The respondents in our study noted two different methods for gathering data using spreadsheets. Some respondents would send out the complete spreadsheet with instructions for their LDP's to complete the parts of the spreadsheet that were applicable to them. The coordinator would then piece together all of the different spreadsheets from their LDPs. The second way that coordinators used spreadsheets to gather data was to cut apart the spreadsheets first and send the LDP's only what applied to them. After receiving the spreadsheets back, the

coordinator would then use cut, copy, and paste to move the information from the individual spreadsheets into one master spreadsheet.

- Compare data: One of the initial tasks that respondents often do is to compare our list of their establishments to their list of establishments. The main purpose of this task is to identify establishments that have newly closed or opened. Both compilers and coordinators were using spreadsheets to perform this task. We often saw respondents put the two lists of establishments side-by-side in a spreadsheet and then compare the lists using a function in Microsoft Excel. For any establishments that did not match, the respondent would then have to decide if the mis-match was typographical or if a location had opened or closed.

- Verify data: Prior to submitting data, many respondents compare location totals to aggregate company totals available in balance sheets and income statements. Respondents use summation functions available within the spreadsheet to derive these totals for comparison.

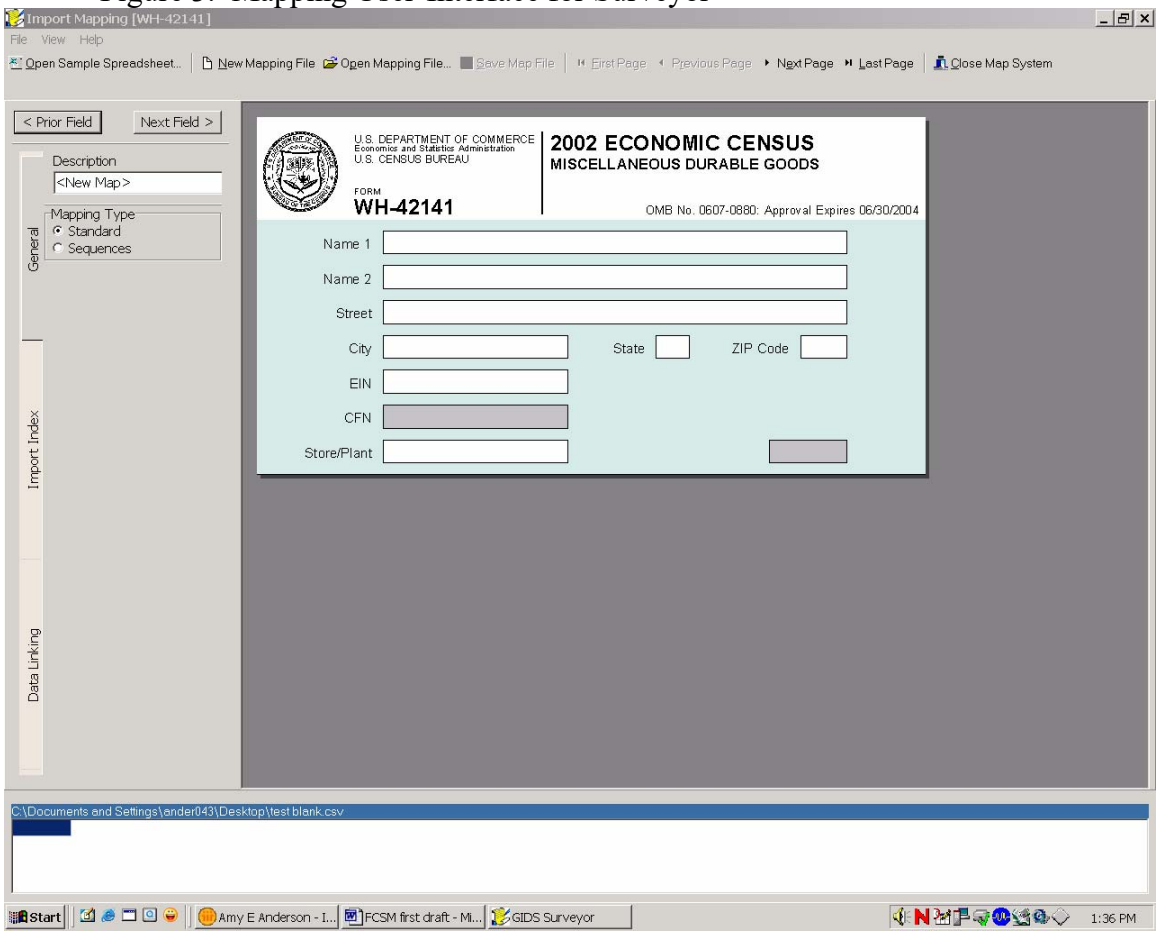
- Correct data: After respondents have imported their spreadsheet into the software, they are given a listing of establishments that have data with edit failures or warnings. Some respondents would make the necessary corrections to the data within the software while other respondents would return to their spreadsheet, make the corrections, and re-import the corrected information. During the study respondents noted that they would like to see more information about their edit failures and warnings given in a list format so that they could print the list and make quick corrections to the spreadsheet. Respondents noted that that it added extra burden to go through the software to make corrections.

C. Creating User Requirements

Based on our findings about company response processes, we began identifying user requirements aimed at assisting respondents with creating spreadsheets for gathering data. It was very clear from our respondents that using spreadsheets was how they preferred to gather the data. It was also clear that gathering data using spreadsheets was a task that fit with the response processes of both gatherers and coordinators. One of the criticisms of the software is that it was built mainly for the respondents who fit the gatherer profile.

To create a spreadsheet using the current version of the software, respondents must first map; or link; the fields on the survey form to the columns on a blank spreadsheet. The mapping user-interface can be seen in Figure 3. In the screen shot example shown in Figure 3, the fields shown would be Name 1, Name 2, Street, City, State, ZIP Code, EIN, CFN, and Store/Plant. Once mapping has been completed, respondents can then export pre-existing information from the software into a spreadsheet. On the spreadsheet each location is listed as a row, and each field mapped by the respondent was a column.

Figure 3. Mapping User-Interface for Surveyor



Respondents in our research noted that they did not always know that they had to perform the tasks of mapping and exporting in order to create a spreadsheet from the software. Both parts of the process were confusing to respondents. Respondents were confused by the user-interface for the mapping process. They were not sure what steps they needed to perform to get the process started and how to interpret the terminology used within the software. Respondents thought that the mapping process was very tedious and burdensome. Respondents wanted the software to automatically create spreadsheets without them having to go through the process of identifying what fields/columns should be part of the spreadsheet. Respondents wanted to see all fields and then have the option of deleting fields/columns once outside of the software and inside of the spreadsheet.

The term 'export' did not make sense to many of the respondents. It was not clear to respondents that they needed to use this option in order to have a spreadsheet created from the software.

Based on these results, we have developed many user-requirements aimed at simplifying the functionality for creating spreadsheets along with providing labels

that users readily identify and understand. We have recommended that the exporting process be named "Creating a Spreadsheet" and that it include two different options for respondents. One option will allow respondents to customize a spreadsheet. The customization option would be similar to the current mapping functionality, but with a re-designed user-interface that points users to the tasks that respondents need to perform using language and labels that are easy to understand. The second option is a default spreadsheet option. This option is in response to the respondents' request for having the software quickly generate a spreadsheet that they can customize (by deleting fields/columns) on their own outside of the software.

D. Using Prototypes, Examples, and Exercises

To test these options, we created prototypes, examples, and exercises. We developed a prototype of the screens that respondents would see when "Creating a Spreadsheet". The prototype enabled respondents to see and react to our proposal. The prototype was created in Microsoft PowerPoint. This allowed us to create a prototype quickly without having to involve the time and expense of using the software programmer. In order to gather more information about respondents request for default spreadsheets that the software would create automatically, we created example spreadsheets and exercises. The exercises were created to mimic many of the tasks that respondents had told us they typically perform with spreadsheets. Rather than talking abstractly, they were able to focus and "play" with an actual spreadsheet. We found that respondents were able to be more articulate about their spreadsheet needs and wants when faced with the example spreadsheet.

V. Conclusion

Learning how respondents conducted the task of responding to the economic census revealed the complexity and inter-connectedness of activities that survey software would need to support. Knowing details about how our respondents had used previous versions of the software, as well as how the spreadsheets were used in conjunction with the software, helped us to identify user-requirements to improve the user-interface and functionality of Surveyor. Prototypes, exercises, and examples helped the respondents to visualize the concepts that we were proposing and to articulate their recommendations for improving upon our ideas. Providing a user-friendly software for the many types of users is a major goal of re-engineering Surveyor. Through our research, we feel confident that the changes that we have proposed will increase the usability of Surveyor for most types of companies.

VI. References

Dillman D.A., Tortora, R.D., and Bowker, D. (1999). "Principles for Constructing Web Surveys," SESRC Technical Report 98-50, Pullman, Washington.

Edwards, W.S. and Cantor, D. (1991). "Toward a Response Model in Establishment Surveys." In Measurement Errors in Surveys, P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds). New York: Wiley.

Marquis, K., Nichols, E., and Tedesco, H. (1998). "Human-Computer Interfact Usability in a Survey Organization: Getting Started at the Census Bureau," Working Papers in Survey Methodology. Washington, DC: U.S. Census Bureau.

Murphy, E., Marquis, K., Hoffman, R., Saner, L., Tedesco, H., Harris, C., and Roske-Hofstrand, R. (2000). "Improving Electronic Data Collection and Dissemination Through Usability Testing," Working Papers in Survey Methodology. Washington, DC: U.S. Census Bureau.

Nichols, E., Willimack, D., and Sudman, S. (1999). "Who are the Reporters: A Study of Government Data Providers in Large, Multi-Unit Companies." Paper presented at the Joint Statistical Meetings of the American Statistical Association, Baltimore, MD.

Nielsen, J. (1993). Usability Engineering. San Diego: Morgan Kaufmann.

Sudman, S., Willimack, D.K., Nichols, E., and Mesenbourg, T.L. (2000). "Exploratory Research at the U.S. Census Bureau on the Survey Response Process in Large Companies." Paper prepared for presentation at the Second International Conference on Establishment Surveys, Buffalo, NY.

Tourangeau, R. (1984). "Cognitive Sciences and Survey Methods." In Cognitive Aspects of Survey Methodology, T.B. Jabine, M.L. Straf, J.M. Tanur, and R. Tourangeau (eds). Washington, DC: National Academy Press.

Usabilitynet.org. (2005). "Task Analysis," http://www.usabilitynet.org/tools/taskanalysis.htm. Accessed 29 June 2005.