Are Generalized Systems the Way of the Future: A Case Study on the Standard Economic Processing System (StEPS)

Shirin A. Ahmed and Deborah L. Tasky, U.S. Bureau of the Census*
Shirin A. Ahmed, U. S. Bureau of the Census, Washington, DC 20233

Key Words: Economic surveys, generalized systems, standards, processing

Introduction

Over the last several years, the economic area at the U.S. Bureau of the Census has been on a mission to standardize, generalize, and streamline processes. One key initiative has been the Standard Economic Processing System, or StEPS, built for the 100+ surveys conducted within the economic area.

The vision for StEPS is to be an all-inclusive processing system – taking care of core survey needs from start to finish. As a processing system, StEPS is to be state-of-the-art in terms of its methodologies, analytic tools, and direction toward new ways of conducting business. To date, 90 surveys use StEPS. The remaining surveys – primarily the economic indicators – will migrate to StEPS by 2004.

The surveys designed for StEPS cover the areas of retail, wholesale, service industries, transportation, and manufacturing. These surveys collect a variety of economic data – from general output measures, such as sales, to detailed commodity information. The surveys vary in size. For example, the Annual Survey of Capital Expenditures has 60,000 respondents and collects 90 items, whereas some of the Current Industrial Reports have 24 respondents and collect as few as five items. Two other agencies have given serious consideration to StEPS. Statistics Canada is using StEPS to process product data from their Annual Survey of Manufacturers. The Energy Information Agency will pilot a survey on StEPS in Fall 2001.

Objectives Behind StEPS

The decision to develop StEPS came from senior management within the economic area, who viewed StEPS as meeting the following key objectives:

- Reducing resources devoted to processing economic surveys. Prior to StEPS, each subject area developed their own processing system, working closely with assigned programming staffs. Over time, 16 separate processing systems, plus variations, existed within the economic area. These systems required staffs for maintenance, enhancements, and conversion to new versions of software. Subject analysts, or survey statisticians, in these areas were focusing more on processing needs than on data or program needs.

- Eliminating redundant programs and code found across the legacy systems. A study in 1994[1] revealed these separate legacy systems performed similar functionality. This meant separate staffs were duplicating efforts. For example, if a new editing methodology had to be implemented, 16 separate groups would have to figure out how to develop, code, and test it for their systems.

- Sharing enhancements corporately. Prior to StEPS, subject areas with more resources had better systems – that is, systems with more functionality or that were more technically advanced. Enhancements were program-specific.

With a generalized processing system, other expectations could be met. First, there would be an infrastructure in place to do new surveys quickly. In the past, every time the economic area acquired a new survey, it meant building a new processing system, or a variation of one of the processing systems in place. Second, a generalized system could provide program flexibility. Given that surveys change content from one reference period to the next, a generalized system – driven by parameters – could handle these changes without modifications to actual code.

This brings us to the last benefit with a generalized system, which was eliminating the bureaucracy associated with maintaining systems. For changes – be it a new survey or items on a survey – the economic area had a laborious process of writing specifications, getting systems staff to code the changes, and then testing the changes. This was an iterative process.

---

*This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

With a generalized system, parameters would represent the specifications to the systems code. Users could modify parameters without requiring programmers to make the change. Since the generalized code was "proved-in," it would not need to be retested when parameters changed. Only the parameters would need to be reviewed. More important, analysts in the subject areas could make changes without relying on systems staff.

Components of StEPS

To meet these objectives, StEPS became a system consisting of *standard data set structures* that support all aspects of survey processing and *integrated modules* that provide program functionality. The standard data set structures let users keep on-line as many historic years as their survey requires.[2] The integrated modules fall into four major categories, described below.

Administrative Modules

Administrative modules let users modify StEPS to meet their survey requirements. Through interfaces, users enter or change parameters to customize functionality for a particular survey. In StEPS, the parameters serve as specifications to the system. For example, users indicate which survey they want to work on, the printer to be used, or the font size for their displays.

In addition, users set up dictionaries for their surveys. Dictionaries define the collected and derived items. With other modules – such as the *Survey Specifications* module – users define through parameters the edit and imputation methods, and where and when to execute these methods.

Post-collection Modules

The post-collection modules are the heart of StEPS. Functionally users have the capability to do the following operations: editing, imputation, interactive data review and correction, data query, estimation, analysis with tools, disclosure, variance estimation, seasonal adjustment, and benchmarking. Regarding seasonal adjustment and benchmarking activities, a sub-process within StEPS handles the storage and manipulation of macro estimates to accommodate these functions. That sub-process is referred to as the Time Series Analytic Repository, or TSAR.

Collection Modules

The actual collection activities for mailout, batch data keying, and electronic reporting use Computer-assisted Survey Information Collection (or CASIC) technologies, that are outside the StEPS environment. Users, however, need to manage information required for these external systems.

Users for example, can create survey-specific information to display on a mailout label, noting that the mail label is generalized across all survey forms within the economic area. Also, users can interact with external CASIC technologies. For instance, within StEPS, users can run processes to create files for mailout or to apply batch updates from separate data capture activities.

Linkages to External Systems

Other than CASIC technologies, the major areas external to StEPS are sample selection, data dissemination, maintenance of the business register, and the processing infrastructure that supports census and census-related programs within the economic area. Most economic surveys use the business register as the source for their survey frame, as well as changes in their frame (i.e., births, deletes, and reactivations). The changes that affect data collection and processing are carried to StEPS via a standardized batch update program. (This is the same batch update program for carrying back to StEPS the output from data capture. Hence, the name is the StEPS standard data output [SDO] format.)

Process for Creating the Initial Version of StEPS

The StEPS Team

A team of programmers, subject area analysts (or survey statisticians), and mathematical statisticians created the initial version of StEPS (that is, StEPS 1.0). The programmers and analysts worked full time on the project, and were organizationally formed as a unit in the Economic Planning and Coordination Division of the economic area. Initially, the mathematical statisticians worked part time on StEPS, and resided in their home area.

Developing the System

Where possible, the team adopted a life-cycle approach for systems development. They spent one-year gathering requirements and designing the system, followed by a second year of full program coding and testing. The team, which started in Spring 1995, migrated three annual surveys to StEPS by December 1997. These annual surveys served as pilots, using StEPS to process their production work for the 1997 reference year. Full scale migration of other surveys started after the pilot phase.

The requirements process required different approaches

depending on the functionality under development. For the interactive routines and data analysis tools, the team relied on a group of "advisory consultants" for requirements. The advisory consultants comprised analysts from the subject areas. For these types of applications, prototyping with feedback from the advisory consultants proved most effective.

For other processes – such as estimation – the team reviewed existing documentation of current survey practices, organized methodologies, then presented it to experts for review. Quite often within a survey area, the person that was an expert in imputation, was not the expert in estimation. So dealing with each process within StEPS often required consulting with many different users from one survey area. On topics where there was extensive discussion of requirements, and some compromise among users, the team would prepare decision documents to represent the formal agreement. As an instance, the team issued a decision document defining the imputation flags and the codes for imputation actions. All decision documents received wide-scale review within the economic area, and were fine-tuned accordingly.

In creating StEPS, the team employed many practices to speed up the design and development. They established standards for how programs were to be coded. Doing this allowed team members to proceed independently, but ensured that component parts would fit together upon completion. Programmers would hold a walk-through of code to ensure design standards and requirements were met.

Where possible, the team incorporated good design techniques from the existing legacy systems. Even though the economic area had separate legacy systems, some of these systems had generalized components.

The team embraced best practices from these systems. For example, they used, as a starting point, the basic design of the "data review and correction" module from work done on the processing system for the Current Industrial Reports (CIR) system. This system processed 75 of the surveys now on StEPS. For imputation, the team took the design from the Generalized Annual Survey Processing (GASP) system. GASP had processed eight of the surveys now on StEPS. Then there was the system for the Farm, Ranch and Irrigation Survey (FRIS) which provided powerful concepts for organizing survey information into dictionaries and categories for editing. Of significant value, the FRIS system highlighted the potential use of SAS® for survey processing.

Computing Environment

The StEPS software is written using SAS products. Since SAS runs on many platforms, SAS fits in with Bureau objectives to move to open systems. For the economic surveys, StEPS is configured for the Unix operating system (on Compaq Alpha machines). Users access StEPS via a graphical (X-windows) communications emulation package loaded onto their microcomputers. For one survey, the Survey of Construction, the variances are calculated using the Bureau's VPLX software.

Managing StEPS Ongoing

Once we started migrating surveys to StEPS, the complexion of the project changed for many reasons. First, the roles and responsibilities among survey participants required redefinition. StEPS affected the way we would conduct surveys. Second, as an economic-wide project, we needed to carry out enhancements to StEPS in a fair manner. Finally, as a product that was maturing, StEPS required more formal processes for software release.

Changing Roles and Responsibilities

As the migration of surveys to StEPS started in force, three key areas underwent changes in their roles and responsibilities, as follows:

- Subject specific programmers who supported the legacy systems now had the job of moving their surveys into StEPS. They become responsible for the StEPS code for their surveys, any linkages to outside information sources, and any development of customized code written in conjunction with StEPS. While StEPS provides most survey functionality, there is still a need for some customization. In circumstances where survey areas use DocuPrint technology to reproduce form images for mailout, as an example, applications are customized with the only standard components from StEPS being the programs for selection of ID to mail and those to retrieve label information.

- Original programmers from the StEPS team now resided back in the systems division of the economic area. This became necessary for knowledge transfer and buy-in. The team's original group of programmers now expanded to a net gain of two programmers. The team maintains and enhances the StEPS code, assists with survey migration and training, and communicates on technical issues with outside organizations interested in StEPS.

- Processors who supported legacy systems now were responsible for scheduling migration activities, developing requirements and user documentation, and conducting training. The one survey analyst from the team works on this staff, and serves as a valuable resource for knowledge transfer. The processors monitor and manage the ongoing StEPS Change Control Process.

StEPS Change Control Process

Now in its third year, the purpose of the StEPS Change Control Process is to manage enhancements. At its core is a User Review Board (URB) comprising project managers across the economic area. The URB members meet monthly to prioritize changes.

Users of StEPS submit changes or problems through Bureau-supported Remedy® software. (This software is used for IT troubleshooting and help desk support at the Bureau.) StEPS users comprise analysts in the subject areas, the mathematical statisticians – who have responsibility for developing parameters for modules such as imputation and estimation, the subject programmers, and the processors themselves. As Remedy tickets are received for new enhancements, processors research the feasibility of the request, the number of surveys that benefit from the change, and the resources required to implement the change. Changes are reviewed at StEPS User Group meetings then submitted to the URB for their decisions.

StEPS Version Control

Version control is tied to the StEPS Change Control Process. With 90 surveys on StEPS, and the migration of economic indicators starting this year, our goals are to stabilize the code for production, and issue new software releases less periodically. We have been in a mode of constant change with StEPS, issuing new code almost weekly. In some situations, our releases have not been documented enough for sufficient testing.

We are now in the process of trying to develop a formal release process from the test machine to production. Analysts use the test machine to verify parameters and test new enhancements to code. The formal release system will indicate the enhancements to be reviewed, document the enhancements, and outline the testing plan for approval. In conjunction with this test plan we are slowly controlling the release of new software – with a short term goal of monthly releases.

Performance Measures

In evaluating StEPS, the authors looked at several measures. There are measures related directly to the primary objectives behind StEPS. Referred to as management objectives, they cover the categories of organizational effectiveness; adaptability and flexibility; elimination of redundancies; and costs. Then there are measures related to the users' perspective on StEPS, which we refer to as user dimensions. Both sets of measures give us a starting point for continuously evaluating and improving StEPS.

Organizational Effectiveness

The most significant measure of StEPS as a successful integrated processing system is the fact that 90 surveys use the system, and that other agencies are interested in the product. This measure is significant given the technical issues that had to be overcome and, more important, the cultural issues.

The risk with any generalized system is its acceptance by users. Our user community initially viewed StEPS as one-size-fits-all, and having less functionality than their customized systems. (In fact, users not familiar with StEPS still think this.) In truth, StEPS has met each survey's requirements, though at times it may have required rethinking, on the part of the user, in the way that StEPS handled the functionality. As surveys migrated to StEPS, what was not in StEPS was added to StEPS. One program manager – interviewed for the Lessons Learned section that follows – noted that StEPS provided far more functionality than his legacy systems.

Adaptability and Flexibility

Nearly all surveys on StEPS have the capability to add or change inquiries on their survey forms without major (or minor) system rewrites. This was put to the test last year when several inquiries related to e-commerce and detailed merchandise lines were added to the Annual Retail Trade Survey. These inquiries were added without changing or retesting code. Take another case with the Current Industrial Reports. Commodity inquiries are added and deleted every reference period without any code modifications.

Elimination of Redundancies and Sharing of Enhancements

Without a doubt, StEPS has eliminated programming redundancies associated with the legacy processing systems. For any specific methodology or procedure, only one set of code exists. Additionally, StEPS means sharing enhancements corporately. Several situations point this out. The CIR program wanted to view a variable for an item across all cases within StEPS. This

"item by ID" routine is very popular with the annual areas collecting service statistics.

In another situation, users in the annual service statistics area requested routines to view company information across surveys. This feature is not only used by other areas, but by staff managing the economic areas' Customer Relationship Management program, an outreach program for data providers.

Staffing and Costs

In terms of staffing and costs, we viewed StEPS in two distinct time periods. There was the initial team that developed StEPS 1.0 – used for three pilot surveys. Then there is the structure today to migrate surveys and continue enhancements, referred to as StEPS ongoing.

- *For StEPS 1.0*, these ten persons served as full time members of the StEPS Team: one team leader, three processors (or former analysts), four programmers, and two contractors that assisted with code development and documentation. In addition, four mathematical statisticians worked part time on the team. We estimate these costs at $1.5 million annually for three years.

- *For ongoing StEPS*, a total cost comparison between StEPS and legacy systems is difficult because of the following:

  – While the economic area has migrated more than 90 surveys to StEPS, the fact is only half of the legacy systems are eliminated. The other half will be eliminated when economic indicators go to StEPS.

    The indicators consume significant resources. For example, four staffs of subject programmers exist. Two of those staffs (about 25 programmers) maintain eight economic indicators yet the other two staffs (also about 25 programmers) maintain about 95 annual, quarterly and monthly surveys.

  – StEPS did not automate all tasks associated with processing current surveys – for example, secondary activities related to updating samples, or activities related to data publication and dissemination, all of which are customized applications. (These are on the list for future enhancements to StEPS.)

  – Lastly, the actual migration of a survey into StEPS requires more resources than the status quo. A survey's first year on StEPS requires set-up, historic data conversion, and some enhancements to StEPS to handle unique requirements of the survey. For example, the grid used to collect industry information on the Annual Capital Expenditures Survey (ACES) required introduction of a "rostering" concept within StEPS. This functionality will be used for the Medical Expenditures Panel Survey (MEPS) when it migrates to StEPS in two years.

  With this said, we did look at surveys in their second year of StEPS, and *programming and processor resources* did decrease. For example, resources for the Current Industrial Reports, Plant Capacity, Annual Capital Expenditures Survey and Pollution Abatement programs decreased by five full time persons. Work associated with the Annual Service Survey, the Annual Retail Trade Survey and the Annual Trade Survey decreased by 1.5 full time persons.

User Dimensions

In April 2001, we canvassed the 29 members of the StEPS User Group and asked them to complete a StEPS User Satisfaction Survey.[3] This survey asked users to assess StEPS on several dimensions, such as ease of use, functionality, response time, recovery time, data review tools, documentation, processing options, and processing time. We received 21 responses.

Table 1 (next page) presents the average response and counts for each individual response. The possible values are 1-strongly agree, 2-somewhat agree, 3-ambivalent, 4-somewhat disagree, 5-strongly disagree, and 0-don't know.

Two eye-opening outcomes are the average scores associated with documentation and time freed up with StEPS.

A follow-up to this survey, in which 15 responded, indicated that most users did not know StEPS documentation was accessible via the StEPS Intranet site. Others focused on the fact that there was inadequate documentation for Imputation and Estimation modules, for which user manual chapters have not yet been written. In terms of time freed up, most users could not determine whether StEPS had resulted in decreased processing time for their surveys.

Lessons Learned

In compiling lessons learned on StEPS, we interviewed program managers within the economic area, plus added a few of our own, as managers of the StEPS

initiative.

Table 1: StEPS User Satisfaction Survey

Question Text	Avg	Answer Equaled:					
		0	1	2	3	4	5
Easy to use	2.3	0	7	8	1	2	3
Navigation simple	2.4	0	5	9	2	4	1
Response time acceptable	2.8	2	1	7	5	6	0
Easy to manage/track my survey	2.2	3	3	10	4	1	0
Recovery time acceptable	2.5	2	2	9	4	4	0
Confidence in data security/restore	2.3	2	3	9	5	2	0
Useful data review tools	2.1	2	6	8	3	1	1
Screens readable and simple to use	2.5	0	5	8	2	5	1
Easy to manipulate data	2.1	4	2	4	4	4	3
Documentation complete	4.0	0	1	3	1	7	9
Less processing time	3.0	13	1	2	2	2	1
Time freed up	3.8	5	1	2	2	6	5
Set own parameters	2.2	2	8	6	1	2	2
More options	2.5	3	5	5	3	4	1

Project Management

StEPS is a directorate project which requires more formal and structured management than our historic treatment of software development projects. Coordinating work across subject and functional areas is a challenge within StEPS, not only in the migration of surveys but also in its ongoing maintenance.

Program managers suggested more detailed, and frequently updated, time schedules with realistic dates. Activities need to clarify roles and responsibilities. Where possible, program managers wanted to understand the communication strategy, and whom to talk to on various issues. Also, they suggested someone to track issues until they are either resolved or turn into enhancements (and formally entered through the StEPS Change Control Process).

Part of project management is configuration management and to this extent we need to research best practices of leading software firms to improve our version release process. This is critical for the economic indicators.

As with any project, there needs to be more periodic assessments of dates, shared experiences and project scope. With StEPS, "scope creep" ended up pushing back survey migration dates for some areas. Finally, we need to periodically document lessons learned, and regroup where necessary.

Documentation and Training

All users – survey statisticians, mathematical programmers, processors, and subject area programmers – had a learning curve with StEPS. Documentation did not exist that helped users visualize how component parts fit together, and only grasped the total system functionality through trial and error. While excellent user documentation exists, keeping it up-to-date and covering areas more complex in nature (such as estimation) has been a challenge.

Regarding training, some program managers thought the basic click-and-point SAS ASSIST class (the only non-StEPS required training) was inadequate. Users with good SAS skills – and these are users across all job series – are in a better position to understand and maximize the capability that StEPS has to offer. While we have conducted StEPS training for each user group over the last three years, program managers suggested it be done just prior to production.

Testing

As we move to migrating the indicator surveys, parallel testing becomes critical. Too, our experience has shown that problems found with StEPS are often the result of a parameter problem and not a software bug. Finally, we have encountered numerous problems in moving to new versions of SAS soon after it is released by SAS Institute. There is sentiment from developers that they are helping SAS Institute find bugs in their software.

Hardware Environment

The hardware environment may be a source of frustration when using generalized software. Generalized systems do not perform as efficiently as customized systems. The hardware has to be adjusted to account for the performance – especially with the larger I/O requirements. On StEPS, performance has been both a perceived and real problem. Perceived in the sense that most surveys – based on diagnosing

existing run times – show excellent speed. Real in the sense that the imputation module, using warm deck parameters, is taking too long for the Annual Retail Trade Survey. (Work is focused on improving this.)

Not Standardizing the Naming Convention

StEPS allows for accessing data across surveys. Ideally, it would have been nice to have standardized the naming convention of data items across all survey programs to better support data sharing. Even though StEPS standardized the common variables associated with respondent characteristics – such as mode of collection technology – users had complete freedom on how to name the data for their program area..

New Areas of Development

Over the next three years, the emphasis for StEPS will be on the following:

- Front end collection instruments to cover activities for computer-assisted data entry (CADE), computer-assisted telephone interviewing (CATI), and call scheduling. We require this functionality for the economic indicators. Existing off-the-shelf software, evaluated last year, does not meet our requirements for real-time updates to the home data sets (database) structures. Currently, this functionality resides with the indicator programs.

- Expansion of macro analytic capability to accommodate drill-down among levels of summary data and to view detailed cases comprising a cell; enabling graphical data analysis in a more structured environment to save analytic time in set-up; and on-line table review systems for insertion of suppression codes, publication footnotes, and locking data for dissemination. On the data dissemination front, we will develop standard file formats for outputting data in several media.

- In terms of new methodology, work will progress on the delete-a-group-jackknife method for variances, the link-relative estimator, and hot deck imputation.

- With the 2002 Economic Census, the economic area heralds in a modern business register. Current economic surveys will be tied to the new register through a new common identification structure. StEPS will be enhanced to build on-line linkages to the business register to handle organizational changes that emanate in new identification assignments.

End Notes

[1] To better manage scarce resources, in 1994 the economic area conducted a planning exercise to assess resources devoted to processing. The goal was to determine critical "must" activities over the next several years.

[2] We are working on procedures to store off-line historic data for archiving. To date, all survey areas have kept on-line their required historic information.

[3] Deborah Chew and Ronald Farrar, U.S. Bureau of the Census, conducted the StEPS User Satisfaction Survey and the follow-up survey. The StEPS User Group represented 29 users, of which 22 were analysts.

Bibliography

Ahmed, Shirin A.(1995), "Background Paper About the Teams in the Economic Directorate," unpublished report, Washington, D.C., U.S. Bureau of the Census.

Ahmed, Shirin A. and Deborah L. Tasky (2000), "An Overview of the Standard Economic Processing System (StEPS)," *Proceedings of the Second International Conference on Establishment Surveys*, pp 633-642.

Chew, Deborah, (2001), "Follow-up to the StEPS User Satisfaction Survey," unpublished report, Washington, D.C., U.S. Bureau of the Census.

Farrar, Ronald, (2001), "Summary of StEPS User Satisfaction Survey," unpublished report, Washington, D.C., U.S. Bureau of the Census.

Monahan, James L. (1994), "The FRIS Processing System," unpublished report, Washington, D.C., U.S. Bureau of the Census.

Sigman, Richard S. (1997), "How Should We Proceed to Develop Generalized Software for Survey Processing Operations Such as Editing, Imputation, Estimation, etc.?," *Proceedings from the Meeting of the Census Advisory Committee of Professional Associations.*

Sigman, Richard S. (2000), "Estimation and Variance Estimation in a Standardized Economic Processing System," *Proceedings of the Second International Conference on Establishment Surveys,* pp. 677-686.

Tasky, Deborah L. (1999), "Basic Functions within StEPS," unpublished report, Washington, D.C., U.S. Bureau of the Census.

Tasky, D., A. Linonis, S. Ankers, D. Hallam, L. Altmayer, and D. Chew (1999), "Get in Step with StEPS: Standardized Economic Processing System," *Proceedings of the North East SAS Users Group.*

Tasky, Deborah L. (1998), "StEPS General

Imputation Development Activities," unpublished report, Washington, D.C., U.S. Bureau of the Census.

Tasky, Deborah L. (1997), "Brief Description of StEPS Files," unpublished report, Washington, D.C., U.S. Bureau of the Census