Year 2000 Decennial Census, USA

Award Finalist, North America 2001

SYSTEM OVERVIEW

The Lockheed Martin Mission Systems Data Capture System 2000 (DCS 2000) utilized barcode, digital imaging, optical recognition and Staffware workflow technologies to check-in and extract the data from 151.4 million forms. This is the equivalent of 1.5 billion 8.5” x 11” pages processed for the United States 2000 decennial census. Seventy percent of the data was extracted automatically with an accuracy of over 99 percent. This feat was accomplished at four data capture centers starting March 6, 2000 and ending August 3, 2000. All of the US Bureau of Census' intermediate processing milestones were met ahead of schedule. However, the sheer size and daily production of more than 1.2 million forms make this the largest workflow and automated data capture system in the world. The system has provided a model for other similar systems in other countries for a faster and more accurate census.

The requirement for a census was specified in the United States Constitution adopted in 1787 to deal with the problem of apportioning representatives and, at that time, for assessing taxes. The decennial census has become the nation’s most comprehensive and expensive statistical data-gathering program. Census results are used during the ensuing decade to address countless public and private data needs. Most importantly, census results are used to reapportion seats in the House of Representatives and redraw congressional, state, and municipal legislative district boundaries. Thus, a key ingredient of an accurate census is not just counting everyone—a formidable challenge in itself—but counting persons in their proper locations. As a result of the 1990 census, eight states gained and 13 states lost at least one congressional seat.

Public decision-makers also use census data to help meet numerous program policy needs. For example, census results are used to help guide the distribution of billions of dollars in Federal aid. After the 1990 census over 100 Federal programs providing grants at the state and local levels use data from or based on the census in formulas that ultimately obligated an estimated $116 billion in fiscal 1991.

Businesses make extensive use of census data for various marketing purposes, including selecting business sites, forecasting demand, allocating advertising and managing sales forces. Census data provide the foundation for volumes of academic and policy research and also assist individuals doing historic research by providing, for example, time-series data and genealogical information.

Since 1970, the Bureau of the Census (BOC) has used essentially the same census methodology—the Bureau develops an address list of the nation's housing units, mails census forms to the vast majority of households, and asks them to mail back the completed forms. Temporary census-takers, known as enumerators, are then hired by the hundreds of thousands to gather the requested information from each non-responding household. The data entered on the forms is then converted to computer process-enabled data for compiling the census reports.

The 1990 decennial census low mail back response rate contributed to making it one of the most controversial censuses in the nation's history. In 1990, 63 percent response was achieved as compared to a predicted 70 percent (1980 was 75 percent). The undercount of millions of persons, 2.1 percent of the enumerated population of 248.7 million, or approximately 5.3 million persons was another contributing factor. Of the many major thrusts identified to correct the 1990 deficiencies, the Bureau resolved to provide a user friendly form to improve response rates and capture the data from returned forms more quickly to better inform the field operations to improve their enumeration efforts.

Since the 1960s, a specially developed Film Optical Sensing Device for Input to Computers (FOSDIC) has been used to convert the census form from paper to microfilm and to display the form to a computer operator for computer input. In 1990, this technology was enhanced to sense marks on the microfilm image and eliminate the manual input process for marks. However, the form required the use of computer registration marks and the form was compacted to reduce the number of paper pages to be processed. If the form could be simplified, the time needed to understand and complete the form could be reduced leading to higher response rates. This led the Bureau to seek a form data capture solution that utilized digital imaging, optical recognition and workflow technologies that did not require a specialized form and could capture the data submitted on the form quicker.

In March 1997 the Bureau awarded to Lockheed Martin Mission Systems the Data Capture System 2000 (DCS 2000) contract to develop, integrate, deploy, maintain, administer and decommission the system to capture and provide to the BOC the data from 130 million census forms in 150 days, an increase of 30 million forms from 1990. The major objectives of the system were:

- Check-in of mail returns within 48 hours of receipt, where mail receipts may be as high as 12 million per day, by recording the barcode viewable through the form envelope's window. This data was forwarded to the Bureau for their purposes of determining non-responding households.
- Capture 85 percent of the surname data from 7.5 million specific forms, or the equivalent of 85 million 8.5" x 11" pages, in a 22 day period ending on April 11, 2000. This data was forwarded to the Bureau to aid in field operations.
- Use digital imaging and recognition technologies to reliably extract the constrained hand-printed and marked contents of census forms and record those contents as ASCII data with 98 percent accuracy.
- Provide capability to manually review and correct data that cannot be recognized with confidence.
- Audit and control quality of the data capture.

The system processes a short form and a long form. The short form consists of a single sheet of paper measuring 11" x 25.5" that is trifolded with data entry possible on both sheet sides. The long form consists of 10 sheets measuring 11" x 17" that is folded and bound with two staples to form a booklet with data entry possible on both sheet sides of all sheets. The long form requests considerably more information; however, it was only requested of approximately one in six households.

The system implementation was verified through three demonstrations and a census dress rehearsal in a fifteen-month period. The demonstrations were scheduled about four months apart and each was characterized by increasing system functionality and integration, i.e., from technology proof of concepts and these technologies integrated into subsystems to a working system. The demonstrations led to a census dress rehearsal where a live census was performed for Columbia SC, Sacramento CA and the Menomonee Indian Reservation.

Following the implementation the system was to have been "locked down" and the system deployed. However, after the dress rehearsal, many new requirements were identified as critical to a successful census implementation. One fundamental change included redesigning the census form to handle six people in a household versus five, which reduces form tracking through the data capture process. Because the data capture period was fixed these requirements were implemented in parallel to deployment, taxing both requirements implementation and configuration controls. Requirements expansion was considered a major risk to a successful 2000 decennial census.

The system was deployed during a nine month period between May 1999 and January 2000 to four data capture centers (DCCs) in Baltimore, MD, Jeffersonville, IN, Phoenix, AZ and Pomona, CA. Deployment consisted of installation, checkout and certification, site acceptance test and operational test and dry-run. The DCCs average about 200,000 square feet with 800 tons of high volume air conditioning (HVAC) and require about 2000 people to operate two production shifts. During peak mail-back period over six million forms can be unloaded at a DCC from over 60 tractor-trailers. The system deployed to Jeffersonville was sized to handle 20 million forms and the systems at the other sites, 40 million forms.

The system consists of 15 commercial-off-the-shelf (COTS) products with capabilities consisting of mail check-in and sorting, paper to digital image conversion, data base management, workflow management, digital image processing, optical character and mark recognition, data review and correction, digital tape backup and recovery, system administration and some custom software. Maximum use of COTS products was required due to the short duration of the implementation period.

The products intended for form image capture, image processing, automated recognition and data correction were integrated through the Staffware workflow management product to create a forms processing cluster. Use of the Staffware workflow management software to integrate the COTS product reduced the amount of custom software development and through a unique generalized interface to the workflow COTS products, could be "plugged" into the solution allowing easy replacement of a product for a better performing one.

The cluster is capable of processing 36,000 short forms a day. The workflow management software tracks the progress of each form's images through the cluster. The workflow process consists of a series of sequential steps. However, while this simplicity is designed to contribute toward a high performance and a scalable system, the workflow system handles the complexities of rerouting a wide variety of exceptions that are equally important in achieving high performance.

Clusters are combined to form a data capture system based on each data capture center's form processing demands. The cluster as a replicable unit offered the ability to completely test the system's forms processing capability within a lab environment and to quickly combine clusters into a production system capable of processing over 400,000 forms a day.

THE KEY MOTIVATIONS BEHIND INSTALLING STAFFWARE'S WORKFLOW SYSTEM

The key motivations behind installing this workflow system were:

- The requirement to track the progress of forms through digital image capture, image quality assessment, optical character and mark recognition, contextual analysis, image backup and data review and correction.
- The need to integrate multiple best of breed COTS products with minimal custom software development during an implementation period of short duration.
- To quickly improve the system through alternative products by establishing the ability, through the use of a standard workflow to COTS product interface, to easily replace a product.
- To have a system that was stable and robust and had the benefit of extended testing through its COTS product implementation in various applications.
- Flexibility to easily add or modify processing steps to meet late breaking requirements.
- Ease of administration
- Cope with the rerouting required for handling a wide variety of exceptions.

THE OVERALL BUSINESS INNOVATION

The Bureau has a long history of technological and business innovation, having introduced punch cards and electric tabulating equipment in 1890. In the 1940s the universal automatic computer (UNIVAC) was designed and built specifically for the Census Bureau. As a result of the growing workload there was a need to capture data more rapidly and a need to improve response rates. This was done using a respondent friendly form; the Bureau continued this trend by embracing the move to digital imaging, optical recognition and workflow technologies.

The Bureau understood that when applying new technologies with a fixed schedule that a solution provider would have to be flexible and creative in accepting requirements late in the system implementation phase and that their proposed system must accommodate change easily. The Bureau made this clear in their data capture system request for proposal.

Lockheed Martin Mission Systems was selected to develop, integrate, deploy, maintain, administer and decommission the system. Reasons for the selection were best of breed COTS products, a proposed architecture that was scalable and where change could be easily implemented and significant experience with a similar system for the United States Internal Revenue Service workflow software had been custom developed.

In order to save costs and cut implementation cycle times, COTS products were utilized in major components, with the Staffware workflow package acting as COTS integrator. This approach was based on the understanding gained on the IRS project.

Development was staged to prove the system at each stage as it was gradually scaled up to become the world's largest ever workflow and imaging system. The deployment of a system this large and the ability to accept late requirements was a business innovation in its own right.

At the proposal stage Lockheed Martin had developed a level A demonstration of the technology that used a single scanner to process 50 forms. A level B demonstration was prepared three months after the contract award. This was for a 'cluster,' a repeatable unit, but with just one scanner, more equipment and more output functionality. Three months later, the Level C demonstration was conducted with a full forms processing cluster with three scanners and more functionality that scanned 2,000 forms. It also included site functionality.

By March 1998 a full dress rehearsal was conducted with the equipment installed at the Bureau processing site in Jeffersonville, Indiana, with two full clusters and the site infrastructure. It was part of a larger dress rehearsal that the Bureau conducted that involved mailing out forms to three cities, processing the returns and carrying out an enumeration exercise to follow up those households that had not responded.

The Bureau users gave a lot of input to the system, which resulted in changes to its 'look and feel.' However, significant requirement changes were also made that challenged the flexibility of the architecture and the process controls for controlling these changes. The following are a few of those changes.

Form Redesign

Following the Dress Rehearsal, the Bureau redesigned the forms for the 2000 census allowing the full data input of a sixth person's data, thereby eliminating a significant amount of follow-up. This increased the physical size of the short form as well as added pages to the long form. The scalability of the system permitted it to respond quickly to the change in workload presented by the increase in form size and pages.

Two passes

At the end of 1999, dry runs were conducted with the first 100 data correction personnel hired in order to test speed and accuracy. The system had been modeled to expect 8,000 keystrokes per hour (kps). This was reduced from 12,000 kps because rejects from optical character recognition are difficult to interpret. Tests showed that the keyers were keying only 4,000 characters per hour, and that the ability of the Bureau to meet their intermediate processing milestones was in jeopardy.

At this point in the deployment, it was not possible to scale the solution because the facilities were at capacity. The only alternative was to process data critical to meeting the intermediate milestones first and the remaining data on a second pass of the images. The first pass would scan all the forms and process all the data that was required to produce the deliverable to the President by the end of the year. This represents all the data on the short forms and the same questions on the long forms.

The remainder of the questions on the long forms could be processed later from the digital image captured during the first pass and stored to a large storage device. This change was accomplished 60 days prior to the start of the data capture. The first pass of the data was successful and transition plans were executed to process the remaining data in a second pass. The ability to reconfigure the system to accommodate the two-pass variation was greatly facilitated by the Staffware workflow management component.

THE OVERALL TECHNOLOGICAL INNOVATION

Although digital imaging, optical recognition and workflow technologies are a technical innovation for the Bureau, these technologies have been effectively proven in other areas. However, the sheer size and daily production of more than 1.2 million forms make this the largest workflow and automated data capture system in the world and a technological innovation. This is the equivalent of 1.5 billion 8.5" x 11" pages processed for the United States 2000 decennial census. This aside, other more subtle technical innovations where workflow improvements were made that played a role in the success of this application.

Generic application interface

Staffware is used as the integrating architecture for the COTS products that are utilized in the forms processing cluster. A generic application interface 'wrapper' was used to integrate the workflow with the COTS products. The products vary from being just toolkits and libraries that can be compiled into the application to standalone applications.

The application interface can support both the synchronous and asynchronous applications. When the interface sends data to a synchronous application to work on, the executable in the interface waits for the application to return it. As soon as the application returns it, Staffware knows the task is finished.

In the case of asynchronous applications, the interface and the application are two separate executables that pass information to each other. However, there is no direct communication link when it has finished. The interface therefore polls the application periodically to ask if it has finished.

From the application side, the interface can be viewed differently, but from the workflow side all the applications look the same. This significantly reduced both the development time for interfacing the products to the workflow and reduced the support time required during operation of the system. As well as controlling the applications with simple functions like stopping, starting and pausing, it also allowed additions to the system to monitor the applications.

Windows NT

DCS 2000 is the first large image-processing project completed by Lockheed Martin Mission Systems that does not use the UNIX operating system. Windows NT was chosen because there was a greater selection of digital imaging and optical recognition COTS products. While the communication to the workflow was all done through Staffware, the communication to the interface layer services was done using Windows NT's built in administrative capabilities. The interface layer in many applications logged errors and other kinds of information in Windows NT event logs. Administrators can easily view these remotely.

System administration

Because of the sheer size of the system, software and system design were geared towards simplifying system administration. A COTS product originally part of the solution was replaced because it was difficult to administer remotely.

The workflow is superficially straightforward, consisting of a series of sequential steps with most processes passing their results to the next, with little forking of workflow cases. However, underneath it is complicated because of the rerouting that is required to handle the wide range of different exceptions.

The functionality available in Staffware is used in a complex and innovative way in order to help administrators to manage the effective handling of exception cases. The software was developed with the approach that if something fails it should do so as elegantly as possibly and it should be as easy as possible for an administrator to recover the process.

One of the key workflow innovations was to create a' fail' queue. If the unexpected occurs in any processing step in the workflow that causes it to alert an administrator, the system will fail the batch and put it into a fail queue. This might be caused, for instance, by losing a connection to the database or to an image server.

The administrators at each site will monitor the fail queues and the service that has failed the batch will log an error. It will also put a comment or error message on the workflow case, so that when the administrators are looking at the Staffware interface they can usually detect immediately what went wrong and investigate further. Once the administrators have resolved the error they can move the batch and send it back to the queue it came from or any other queue to help resolve the problem.

The system administrators find the information they are getting from the built-in Windows NT functions extremely helpful when they are trying to identify problems with the system. There is a list of each service that is running on the cluster and a count of how many form batches are being processed through that service. It gives an indication of whether the process is initialized and running correctly, if it has failed or if it is down.

One feature has proved helpful in a lot of cases, particularly with asynchronous applications. If the secondary application fails for some reason, the first application will still be waiting and would not necessarily know that the other application has failed. A monitoring function keeps track of how long those applications have been waiting and alerts an administrator if something is taking too long to process.

The theory of “elegant failure” also applies to form processing clusters. If a cluster should fail then the system is designed so that this does not affect other clusters. The system architecture maintains as much independence as possible, while at the same time being as tightly integrated as possible. Each individual application can survive without another one. Along the same lines, the interface between the cluster and the site is minimized, so if the site services fail the clusters can continue to operate and vice versa.

Quality assurance

In a data capture system of this size and providing the data for a mission critical application such as a census, it is imperative that the quality of the data being produced is continuously assessed. The data capture system continuously monitors the quality of the data produced by both the automated recognition and the manual review and correction functions.

The workflow in the forms processing cluster implements and controls the continuous monitoring. After data has been produced by either the automated or manual functions, this output stream is sampled and the digital image that was the source of the data presented to data correction personnel without their knowledge that they are performing a quality assessment function.

The original output is compared with the output from the quality assessment step to determine whether the original output was correct. If they match then the original output is considered correct. If they do not match then the process is repeated and a match attempted again with the original or the output from the first quality assessment step. If it matches the original then it is considered correct. If it matches the output from the first quality assessment then the original output is considered incorrect. After a significant sample is obtained then the overall quality of the system can be derived. The system as of this reporting has produced an overall accuracy in excess of 98 percent with 151 million forms processed.

THE SYSTEM USERS AND WHAT THEIR JOBS NOW ENTAIL

In 1990, the form marks were automatically recognized using a specially developed capability called Film Optical Sensing Device for Input to Computers (FOSDIC). The marks were recognized from the forms microfilm using a light sensing technique. Write-in fields were keyed by data input personnel from the microfilm. The forms were processed at seven processing centers.

Since the implementation of the DCS 2000, the 2000 Census operators are checking in and sorting mail using mail sorting equipment, scanning paper using high-speed optical scanners, reviewing the image quality on a personal computer display when it doesn't pass automated assessment criteria, reviewing and correcting the 30 percent of data fields that are not recognized automatically using the digital image of the field and finally using barcode wands and database technology, verifying that the Bureau acknowledged the receipt of a form's data.

THE BIGGEST HURDLES OVERCOME

Late Requirements Changes

As discussed in section 4, significant requirement changes occurred late in the deployment and final test phases. This challenge was successfully met as a result of a system architecture that was scalable and flexible that included a workflow management function that could be easily customized to accept processing flow changes.

COTS Product Substitution

An early problem was that Lockheed Martin was forced to re-evaluate its original choice of the critical image-processing package. It had been selected because it provided a lot of the functionality needed, including an interface to the Kodak scanners and full image processing capability. However, in early demonstrations it was unstable.

Just as serious for a project of this scale, it did not quite fit the overall architecture. It had its own workflow engine that controlled its functionality with which the workflow manager had to interface. After a re-evaluation, it was decided to replace the package with custom code. Lockheed Martin used a toolkit from TMS that provided some basic capability but built the software to its own design using its own code. The flexibility of the Staffware package proved to be significant, because it made the change very easy. Each of the workflow functions performed by the package was transferred to Staffware individually and the new software could interface with it at a low level.

There were also some initial performance problems. The image server and its redundant array of inexpensive disk (RAID) fault tolerant storage came from different suppliers causing problems when communicating with each other. After trying several different combinations, an integrated product was sourced from Dell, who built both.

Keying

Lockheed Martin experimented with character keying. This involves keying only individual characters to which the software has ascribed low confidence. However, the project transferred to field keying, where the whole field is re-keyed if any character within it is low confidence.

Although it may be faster for some keyers, character keying is not quite as accurate as field keying. Whereas it easy to know that something is wrong in a field, it is extremely difficult to determine exactly which character is incorrect. The potential for additional speed did not justify the accuracy lost.

Paper handling

One of the big problems in the early demonstrations was the paper handling. As paper moves through the system, it is in a batch if it has been mailed in, or a box if it has come from enumerators. Each box is processed as a batch. If a form in a batch was mis-fed so that it couldn't be automatically processed, the paper form had to be physically pulled at a later step so that it could go back for re-scanning or to be fully keyed from paper.

The system told the user which batch the form was in and where in the batch the form could be found. However, counting the sheets was very difficult and too complicated for users to do correctly each time. It was decided to completely change the way this task was performed.

After the clusters have processed them, every form is scanned with a handheld barcode scanner at site level and compared to a data receipt confirmation file provided by the Bureau. This verifies that each sheet is ready to be destroyed and identifies those that the system wants reprocessed.

The scanner actually gives a green light if it is ready to be destroyed and a red light if it needs to be pulled out for reprocessing. If reprocessing is required, it goes to a second stage where a different person rescans the bar code to find out why it has to be pulled so they can direct it either to rescanning or to be keyed from paper. This is a more labor-intensive process but has proved to ensure that every form received is fully processed.

Communications

The Bureau contracted with TRW to operate three of the four data capture centers. Due to the very aggressive schedule and the speed with which system deployment was to be executed, communications were always a challenge. Consequently, an integrated product team structure was utilized, which facilitated communications and demonstrated both Lockheed Martin's and TRW's commitment to the success of the Census.

THE NEW SYSTEM CONFIGURATION

The four data capture centers are approximately 250,000 square feet with 4,000 amp electrical service and 900 tons high volume air conditioning a located in Baltimore, Maryland, Jeffersonville Indiana, Phoenix Arizona and Pomona California. Each has a staff of approximately 2,000 people operating on two shifts, seven days a week. Jeffersonville is designed to handle 30 million forms the other three sites 40 million.

The architecture is split into site infrastructure and a number of identical forms processing clusters consisting of three high-speed scanners and supporting hardware and software. Windows NT was used as the operating system throughout the entire system. The network is 100Mbps switched Ethernet using the TCP/IP protocol.

Site hardware

Site	Sorters	Clusters	Scanners	Servers	Workstations
Baltimore	9	15	45	578	580
Jeffersonville	6	10	30	391	381
Phoenix	7	15	45	576	566
Pomona	8	14	42	545	547
Total	30	54	162	2,090	2,074

Each site has two Oracle databases, one for status and a management database with statistics on the system, such as operator keying rates, accuracy rates, etc. Docutronix is used for mail sorting and bar code reading. Data backup was performed using Legato and StorageTek tape backup and Tivoli software used for network and system management.

Each cluster has Staffware as the workflow manager to control every step. It was selected because it was a mature product that offered all the functionality needed at a cost-effective price.

Each cluster has three Kodak High Speed Scanners with Kofax scanner controllers. The scanners were the most mature and robust products available and Lockheed Martin had used them successfully on previous projects.

TMS Sequoia toolkit was used to build the image processing and image quality assessment functions to Lockheed Martin's design. Optimum Solutions Fast Accurate Questionnaire Scanning System

(FAQSS) was used for optical mark recognition, as it was the most accurate available at the time.

RecoStar from CGK in Germany was used for optical character recognition (OCR), as it was the most accurate and offered contextual functionality. It was used in conjunction with custom code for contextual analysis written by Lockheed Martin to improve accuracy.

Data correction software from Captiva was used for both keying from image and keying from paper. It offered a lot of flexibility for different ways of keying that affect performance and user-friendliness.

Each cluster has an Oracle database to record the file location of each form. This is held as an image file on a RAID server with each page of the form in the same file. The data processed from it is held as an adjacent text file. Lockheed Martin had tried storing the images in a database, but it became so large it was unmanageable.

COST SAVINGS, INCREASED REVENUES, AND PRODUCTIVITY IMPROVEMENTS

Due to the length of time between the decennial censuses, it is difficult to make direct comparisons. Better comparisons can be made between the decennial census plan and its subsequent performance.

The decennial census has been extremely successful. As discussed earlier, the data capture participated in several majors thrusts for 2000 decennial census: the capture of census data faster within accuracy objectives using a respondent friendly form designed to increase response rates. The planned 2000 decennial response of 61 percent was exceeded at 65 percent. All intermediate processing milestones were met early with data that exceeded accuracy objectives of 98 percent. Certainly significant savings from these improvements were experienced throughout the Bureau; however, the direct cost savings are difficult to quantify. Certainly, the elimination of the data review and correction workload through the use of the digital imaging, optical recognition and workflow technologies produced a savings in manual labor that was at least $50 million.

During the implementation of the data capture system several productivity and cost savings were realized. A cost/benefit analysis showed that replacing the imaging software saved the project $3 million. The toolkit used was cheaper than the packaged software licenses, support costs were considerably reduced and fewer servers were needed. The vendor had recently rewritten the product, so the project was effectively a beta site, so considerable time was saved in testing.

Because the system was designed to be easy to administer, Lockheed Martin estimates that it would otherwise have required twice as many system administration staff. This saved four people per site over two shifts, which has saved $1 million, plus the cost of training them.

Lockheed Martin estimates that the use of commercial off-the-shelf packages, compared to writing the software from scratch as they did with the IRS, saved $1 million. The generic application interface has saved development time through reuse and reduced administration. The company has estimated this savings at $200,000.

COMPETITIVE ADVANTAGES GAINED

Whereas workflow and automated processing are in widespread use, the unique advantage of the DCS 2000 system is driven by the enormous scale of the program it had to support, which involved processing up to 17.5 million pages per day. The size and short time-scale forced the system to be made very robust, with features like failure queues and administrators moving batches, as there was little time to recover from problems.

Lockheed Martin has already won the United Kingdom contract and is investigating other opportunities. Other countries have visited the four US sites and have seen the scale of the processing. Many companies around the world stand to get more accurate and faster censuses as a result of this project.

IMMEDIATE AND LONG-TERM PLANS TO SUSTAIN COMPETITIVE ADVANTAGE

Lockheed Martin is already developing the system for the United Kingdom Census, which is using the same system as utilized in the United States. It is also proposing several other census processing systems on different scales, although none are on the same scale as the United States Decennial Census.

The constraint is not only the population to be counted, but also the speed in which the US decennial census needs to be completed. Other censuses have longer schedules enabling processing to be carried out over several more months.

Clearly, Lockheed Martin Mission Systems is considering improvements to the system. These improvements can be grouped into two categories: (1) extend the census form data capture to downstream census data processing and (2) continue to reduce errors and the number of required operators by reducing reject rates from the automated capture.

Census data is post-processed using statistical analysis and coding. The workflow system will permit the easy extension of the system for these additional processing steps. Different paths a form can take when it is coded complicate the workflow process. An additional complication is created when forms may travel through different operators; however, the Staffware workflow system permits and manages these complicated processes.

Coding can also improve accuracy and reduce reject rates. Even if something is low confidence from the automated capture, the automatic coding takes into account fuzzy logic and other techniques to avoid sending it to a keyer. These types of techniques require a complicated workflow configuration, because of the different ways of avoiding manual processing. If it still cannot be coded automatically, it may be sent for character keying, which should be faster than field keying. It can then be sent back to the automatic coder.

If it still does not code, to improve performance, just the numeric fields can be sent to a keyer. This allows the keyer to stay on the numeric keypad, rather than switching continuously to the alphanumeric portion of the keyboard. That hand movement actually slows down the keyer significantly.

After numerical keying the image may then be sent it to an automatic coder. Then it may go for alpha keying and then go back to an automatic coder. If all those steps fail, the image has to go back for manual coding.

Because keying is so labor intensive, automation can save hundreds of keyers at each site, resulting in a smaller site and smaller workload. Workflow innovations can reduce a lot of the risk and costs. In particular, merging field keying and character keying into the same system could generate major benefits.