Converting from Paper to an Automated Mode for Collecting Group Quarters Data for Demographic Surveys

Sherry E. Thorpe
U.S. Census Bureau, Demographic Statistical Methods Division
Washington, DC 20233

Key words: Frame Development, Group Quarters, Non-Sampling Error, Automated Instrument, Demographic Surveys, Sample Redesign

I. Introduction

Following each decennial census, both the U. S. Census Bureau and agencies sponsoring the major demographic surveys, redesign their sample surveys. The sample redesign allows the demographic area of the Census Bureau to obtain sample to:

- continue fielding the demographic surveys in the next decade
- replace the out-of-date list of housing units used for sampling and data collection
- account for changes in population and demographics during the decade
- accommodate defined survey objectives
- introduce new methodologies

The introduction of new methodology improves the accuracy of survey data. Sample redesign can enhance the incorporation of new methods for collecting data for surveys. During the ten–year period between 1990 and 2000, for example, new and improved listing methods, such as automating paper listings, were developed.

As part of the 2000 Sample Redesign, the paper listing operation for housing units and group quarters (GQs) was converted to an automated mode. The Demographic Area Address Listing (DAAL) operation was designed to combine two listing instruments, the Automated Listing and Mapping Instrument (ALMI) and the GQ Automated Instrument for Listing (GAIL) with a management system to list and update housing units and GQs in both the GQ and Area Frames. By automating the listing procedures for GQ listings currently done using paper listing sheets, the Census Bureau could reduce non-sampling errors while reducing total listing time.

II. Frame Development

For sampling purposes, the United States is divided into blocks. A block is a geographic area bounded on all sides by visible features such as roads, railroad tracks, or rivers; or by invisible features, such as county boundaries, city limits, or property lines. A block is the smallest geographic entity for which census information is collected. Each block is assigned to one of four frames: Area Frame, Unit Frame, Group Quarters (GQ) Frame, or Permit Frame. Every housing unit and GQ in the United States would be included in one of these four frames. Each frame is defined below:

Area Frame contains
- blocks that are in areas that do not issue permits for constructing new housing units; or
- more than 4% of the addresses in the block do not have a complete address; that is, a house number and a street name;

Unit Frame contains
- blocks that are in areas that issue permits for constructing new housing units; and
- at least 96% of the addresses in the block have a complete address;

GQ Frame
- this frame incorporates all the GQs within the blocks of the Unit Frame;

Permit Frame
- this frame includes all permits issued by Building Permit Offices (BPOs) for new housing unit construction built since April 1, 2000.

The paper listing operation of the 1990 Sample Redesign included the GQ Frame, comprised of all the GQs in the Unit Frame, and the Area Frame, which includes housing units and GQs. The Lists/Updates Keying (LUK) system was used to track Area and GQ paper listing sheets. These listing sheets were completed by field representatives and mailed to the appropriate regional office. On a flow basis, the regional offices would mail the listing sheets to the Census Bureau's National Processing Center (NPC) for processing. The NPC manually checked-in and reviewed the listing sheets, keyed each listing sheet, and transmitted the keyed data to headquarters. Once all the data for a particular due date was received by headquarters, sampling was conducted.

The paper operation required several months to complete due to the many phases involved. Often, listing sheets were lost in the mail, misplaced or arrived at NPC past the due date. These late listing sheets had the potential to effect sampling.

III. Demographic Area Address Listing (DAAL) Operation

The DAAL operation is composed of two listing instruments and one management system.

The Automated Listing and Mapping Instrument (ALMI) is an instrument developed to list and mapspot housing units and GQs. Field representatives are assigned specific blocks to update addresses, mapspots, unit status, and contact information. These Update Block assignments allow the field representative to verify and update census data by working from what is on the ground to what is in the ALMI.

The GQ Automated Instrument for Listing (GAIL) goes a step beyond the ALMI in listing GQs. Group Quarters are a type of living accommodation where residents share common facilities or receive authorized care or custody. These structures contain no units or one or more units within them that should be listed. After completing the Update Block assignment in the ALMI, the field representative launches the GAIL and completes a more thorough investigation of the GQ. The field representative collects the location address, obtains contact information from a knowledgeable respondent, determines the specific type of GQ, and deciphers whether the GQ is eligible to be listed. After determining eligibility, the units within the GQ can be listed.

Case Management is the system that manages the workload of the field representative. The status of each case can be reviewed including the outcome code, address and notes made by the FR in updating the case.

IV. Non-sampling Error

The GAIL was designed to reduce non-sampling error within the instrument. Non-sampling errors are caused by factors other than sampling. Non-sampling error can be categorized into the following groups:

- specification error
- coverage error
- response error
- processing error
- non-response error

Specification errors occur during the planning stage. Unclear and ambiguous requirements can result in inadequate listing outcomes. To minimize specification error, programmers, subject matter experts, and project team members reviewed the GAIL specification. The format of the specification allows even the first-time reader to understand the instrument. The format used worked well for the programmers. Complex procedures were not difficult to understand and were programmed correctly.

Response error results from data that have been requested, provided, received or recorded incorrectly. Response errors may arise from the collection of data from inappropriate respondents, respondent memory or recall errors, misrecording of responses, or misunderstanding of data collection procedures. To minimize collecting data from inappropriate respondents in the GAIL, extra screens were developed. For example, in determining whether college housing should be listed, two college housing screens were added. If it was determined that the dormitory, apartment, or other college housing was leased by the college or university, and/or married or married and unmarried students lived in the structure, the structure was changed from a GQ to a housing unit in the GAIL. The GAIL then prompted the field representative how to proceed. During the paper listings, field representatives had to make that determination on their own.

Non-response errors are due to non-interviews or non-responses to individual questions. There are two types of non-response errors: complete and partial. Complete non-response errors occur when the listing fails to collect data for a unit in the selected sample. Partial non-response errors occur when incomplete information is obtained from the respondent. To minimize complete non-response errors in the GAIL, the field representative has the ability to return to a GQ several times to obtain a complete listing of the units. Field representatives are instructed to speak to any knowledgeable person at the facility to capture the requested data.

When part of a listing is completed, the GAIL looks at specific entries to see if they have been obtained. For example, if the number of units at a GQ has been captured, but the specific information on each unit has not been given, this listing is considered partially complete and will be sampled. However, if the number of units at a GQ has not been captured, the GQ is considered incomplete, and if not completed by the assigned due date, it is assigned to the next earliest listing period to be completed.

Coverage errors occur when there is an omission, duplication, or wrongful inclusion of the units in the population or sample. The GAIL collects information for non-institutional GQs only. Non-institutional GQs are facilities where individuals are free to leave the grounds or change their address without permission. To minimize wrongfully listing institutional GQs, that is, facilities where individuals are not allowed to leave the grounds or change their address without permission, eligibility questions were added in the GAIL. Based on the field representative's response to these questions, the GAIL determines whether the GQ is institutional or non-institutional and if it should be listed. A confirmation screen informs the field representative of the GQ status.

Processing error results from incorrect data transfer, incorrect editing, or coding. Errors can be caused by software problems during computer processing which cause correctly keyed data to be lost. The GAIL is a response driven instrument. Answers to questions determine the path the instrument follows. Any changes to previously answered questions could drastically alter data previously captured. Once a field representative completes an assignment, access to the completed case is not granted. By limiting access to specific incomplete assignments, the chance of losing correct data is decreased.

V. Timing

The demographic area is still converting survey listing operations to an automated mode. Each major demographic survey listing operation will be converted separately. Because a complete conversion has not taken place, the 1990 Sample Redesign Lists/Updates Keying (LUK) operation is still in place. This section compares the two systems during the same time period.

The LUK system is a computer tracking system designed to track Regional Office (RO) assignments and listing sheets from the RO to the National Processing Center (NPC) to headquarters (HQ). All assignments for a given assignment period are released at one time. Field representatives list the assignments based on the RO due date. Since paper listing sheets are used, these listing sheets must be mailed to the RO for processing, shipped to the NPC for further processing and keying, and keyed data transmitted to headquarters. Mailing time must be considered when listing assignments. Once HQ receives the keyed listing data, sampling can be conducted.

The LUK operation involves the following steps.

- Field representatives list the housing and GQ units on paper listing sheets.
- Field representatives mail completed listing sheets to the regional office (RO).
- The ROs check in and review the paper listing sheets for errors.
- The ROs mail reviewed listing sheets to the National Processing Center (NPC).
- The NPC checks in and reviews the listing sheets for processing errors.
- The NPC puts the listing sheets into batches to be keyed.
- Batched listing sheets are keyed and transmitted to headquarters (HQ).
- HQ conducts sampling for demographic surveys.

The LUK process takes approximately 4-6 months from release date to HQ survey sampling.
See Table I.

The 2000 Sample Redesign Demographic Area Address Listing (DAAL) operation is similar to the LUK system, but involves fewer steps. Every three months, new assignments are release to the field representative's laptop. These assignments have three monthly due dates known as Field Representative Closeout. Field representatives can list and transmit any completed assignment to HQ at any time.

The DAAL operation involves the following steps.

- Field representatives receive assignments via transmission.
- Field representatives conduct listing operations and transmit nightly.
- For each closeout date, HQ processes listing data.
- HQ conducts sampling for demographic surveys.

The DAAL operation takes approximately 3-5 months from release date to HQ sampling.
See Table II.

By automating the listing operation, the NPC portion of processing is completely eliminated, thus reducing time and expense. Paper listing sheets are no longer needed. This eliminates tracking and storage. Field representatives now have the opportunity to complete assignments early, which allows travel and work time to be used more efficiently.

VI. Requirements

To convert paper listing GQ assignments to an automated environment, the following was needed:

- user-friendly software to accommodate field representatives who had little or no experience using an automated instrument for group quarter listing;
- software that could handle complex listing procedures; and
- a management system that would track the assigned workload and manage completed and partially completed GQ assignments. This management system had to incorporate all the information needed by the field representative on a GQ assignment and display it before the case was opened.

VII. Challenges

There were several challenges that needed to be addressed.

Software – What type of software would not only be user-friendly, but would also accommodate difficult listing procedures? Automated listing for demographic surveys had never been attempted. What system would manage completed and partially completed assignments?

The Census Bureau's expert in programming survey instruments is the Technologies Management Office (TMO). TMO suggested that the GAIL use the Blaise software. Blaise, a windows based software, was currently being used for one of the Census Bureau's surveys with much success. The software was user-friendly, had great functionality, and could be tailored to use for a listing operation.

TMO developed Graphical User Interface (GUI) Case Management to manage the GAIL and ALMI work. This system tracks the assignments, displays detailed information on specific assignments, allows updates to certain variables, and holds completed assignments until transmission. Case Management would be an essential part of the DAAL operation.

Timing – Would we be able to have an automated instrument ready for production in 2004? Sample Redesign would require a lot of resources thereby limiting the number of resources available to work on group quarters automated listing.

Planning for the GAIL started in 1999. A sub-group from the DAAL operation was formed specifically to address the GAIL instrument. By late 2000, an early version of the GAIL instrument was ready to review. Throughout the next three years, the instrument went through a series of additions, changes, and enhancements. The 2004 production instrument was ready by July 2003 as scheduled.

Training – Could we train large field and regional office staffs to provide the quality of work currently attained using paper listing sheets? Group quarters automated listing would be combined with housing unit automated listing and supervised by regional office staff with little or no previous listing experience.

Verbatim training was developed to train all regional office and field representative staff. This type of training would ensure consistency. Subject matter experts on the GAIL, the ALMI, and Case Management were consulted. Each draft of training materials was reviewed for consistency and ease of learning.

Support – Could sufficient technical support be provided in a timely manner? Given the reduced listing time, hardware, software, and subject matter expertise, timing was critical. Headquarters, field and regional staff's full support would be needed.

TMO currently had technical support for data collection application and automation infrastructures. This support team would service the DAAL operation's technical support needs.

For subject matter expert support, a contact person was identified and contacted directly about subject matter issues. A 24-hour response rule, similar to what was expected for surveys, was established.

VIII. Automating Procedures

Automating the interviewing process increased the quality of data in several ways.

- Listing skip patterns were automated which minimized interviewer error in following complicated listing procedures.
- Data edits were incorporated to ensure a predetermined quality of data before a listing could be labeled complete.
- Including verbatim questions in the instrument for the field representatives provided a sense of professionalism in the listing operation.

- Listing group quarters and housing units with laptops instead of paper listing sheets appears more official.
- Quick identification and reassignment of problem cases is easier.
- A quick turnaround of completed listings to headquarters was achieved.

To further increase data quality, the GAIL included distinct paths for:

- initial and update listings for GQ Frame group quarters and Area Frame group quarters
- institutional, non-institutional and military group quarter listing
- group quarters converting to housing units
- non-interview group quarter listings
- seasonal group quarters
- various listing methods
- call back group quarters assignments, where the field representative returns on another day
- college dormitories and religious housing

IX. Training

Structured, homogenous training was developed for all twelve of the Census Bureau's regional offices. By developing standardized, verbatim training, each field representative, regardless of experience level, would receive the same training.

The goal of training was to introduce automated listing to every listing field representative so that upon completion, each person would possess a working knowledge of each DAAL instrument. This required a consistent training plan that could be implemented by trained regional office staff.

The initial training was done with the regional office staff at an off-site session. The five-day training session referred to as "train-the-trainer" allowed the regional office staff an opportunity to review the instruments associated with the DAAL system, ask questions, and provide valuable insight to the training process. As a result of the train-the-trainer off-site, the training was reduced to four days, instrument questions were revised, and training materials were updated.

A DAAL evaluation of training was developed and distributed at the first training sessions. Over 300 field representatives responded to the questionnaire.

The field representatives were asked to indicate whether they thought a change in the amount of time allotted to each section of the classroom training would help make the overall training more effective. The results were interesting. See Table III.

The DAAL training evaluation revealed some other thoughts. When field representatives were asked what training developers need to be aware of in working to improve the effectiveness of the entire DAAL training in the future, 20.5% stated that more field work needed to be included (n=365). 13.7% suggested that more time should be allocated for the entire DAAL training (n=365). Overall most field representatives felt that training was very good, and the trainers were excellent (n=264).

X. Lessons Learned

As with any project, upon completion, there are many lessons that have been learned.

Functional requirements for automated instruments should be decided early. Decisions on functions to include should be finalized as soon as possible. These requirements should contain great detail to ensure that the finished product operates efficiently. Although the GAIL was completed on time, several late requirements prevented the instrument from functioning correctly during the full systems test.

Using new software requires a certain level of up-front expertise to prevent unrealistic expectations. While new applications usually provide more functionality, there still can be drawbacks. These limitations should be discussed in the planning phase. The Blaise software was new to the project planners. Assumptions on how the software should work were not realistic. Having knowledge about these concepts early would have prevented a lot of frustration and wasted time.

The implementation of Project Management is essential to success. Change control procedures, delivery of instruments, and testing results need to be monitored in an organized manner. Toward the end of the project, some of the project management concepts were adopted. These concepts would have made the entire project run smoother and more efficiently had they been established from the start.

The entire system should be planned while considering each specific instrument's requirements. One change could affect many systems adversely. Constant communication between the instrument experts on all parts of the systems, the GAIL, the

ALMI, and Case Management, should be planned. Many testing errors could have been prevented if changes to an instrument were communicated.

More time should be allotted for classroom practical exercises and field exercises. The findings from the DAAL General Evaluation clearly identify field representatives' desire for more hands-on training.

XI. Conclusion

The GQ paper listing operation conversion to an automated mode for collecting Group Quarters data was a success. Through planning and effort, all of the challenges were addressed and mitigated. All objectives were met as well. The overall listing time decreased by one month and the process eliminated an entire manual operation comprised of mailing and keying completed listing sheets. Non-sampling error issues were addressed and reduced through the complex design of the instrument, the training, and the complete DAAL operation.

XII. References

Demographic Statistical Methods Division, U. S. Bureau of the Census, Intranet site, Overview.

Form 11-8, Listing and Coverage: A Survival Guide for the Field Representative, Vol. 1, Demographic Statistical Methods Division, Intranet site.

Lee III, T.; Hayes, J.; (October 2003), Lists/Updates Keying (LUK) Ongoing Phase Timing.

Dietz, E.; (August 1999) "Reducing Nonsampling Error in the Bureau of Labor Statistics Occupational Compensation Surveys".

Byrne, R.; Adams, T.; Ramos, M.; (August 2001), "Results of Quality Assurance on the Person Interview Operation of the Accuracy and Coverage Evaluation of the Census 2000".

Ettu, M., (June 2003) "Demographic Area Address Listing (DAAL) General Evaluation of Training".

Krasko, N.; Dawson, J.; Prouse, H.; (August 2001), "Automating the Census 2000 Accuracy and Coverage Evaluation Field Operations".

Kennel, T.; Flanagan, T.; Hayes, J.; King, S.; (April 2004), "Specification for Creating the Demographic Area Address Listing (DAAL) Sample Control Input Files (SCIFs) for Current Surveys (Doc. #1.6-S-1).

U. S. Department of Labor, Bureau of Labor Statistics, BLS Handbook of Methods, via www.bls.gov.

Table I – Schedule for the Lists/Updates Keying (LUK) System

Release Date	Update & List	RO Closeout	NPC Closeout	HQ Sampling	Interview Date
July 8, 2004	July 26 – Sept 19	Oct 9	Nov 17	Nov 19	200501
	July 26 – Oct 19	Nov 9	Dec 15	Dec 17	200502
	July 26 – Nov 19	Dec 9	Jan 19	Jan 21	200503

Table II – Schedule for the Demographic Area Address Listing (DAAL) Operation

Release Date	Update & List	Field Rep. Closeout	HQ Sampling	Interview Date
August 11, 2004	Sept 1 – Nov 1	Nov 1	Nov 6	200501
	Sept 1 – Dec 1	Dec 1	Dec 6	200502
	Sept 1 – Jan 1	Jan 1	Jan 6	200503

Table III – Findings from the Demographic Area Address Listing (DAAL) General Evaluation of Training

Percent of Field Representatives Reporting that Training Time for Topic should be eliminated, reduced, increased, or kept as is.					
Training Topic		**Eliminate Time**	**Reduce Time**	**Increase Time**	**Keep Time As Is**
Case Management	n = 386	0.8%	8.3%	9.6%	81.3%
Listing GQs with the GAIL	n = 389	0.7%	15.4%	14.1%	69.8%
Preparing for Listing Assignments Using the ALMI	n = 387	0.6%	4.1%	19.1%	76.2%
Completing Tasks in the ALMI	n = 386	0.5%	2.6%	31.9%	65.0%
Classroom Practical Exercise	n = 385	0.3%	2.1%	33.5%	64.1%
Field Exercise	n = 387	0.3%	0.8%	42.1%	56.8%
Summary and Evaluation	n = 359	1.1%	2.5%	8.6%	87.8%