Conducting the Economic Census

Shirin A. Ahmed*
Lawrence A. Blum*
Mark E. Wallace*

"'Conducting a census' to most people conjures up images of questionnaires, interviewers, and respondents. So much more is involved. This article describes the organization and planning that precedes the questioning, and shows what happens after the answers come in."

John Govoni, Chief
Economic Planning and Coordination Division
U.S. Census Bureau

This article is divided into five parts:

- The Organizational Structure Behind the Economic Census describes the organization of the U.S. Census Bureau with particular reference to the roles of divisions that impact upon the Economic Census. These divisions stipulate the policies underlying the programs that Economic Census data are to serve, prepare the questions and the questionnaires, arrange for the census to be taken, process the data after they are collected, and disseminate the results to the public. Without such an organizational apparatus, there could be no Economic Census.

* *Direct all correspondence to: Shirin A. Ahmed, Assistant Division Chief for Post-collection Activities, Economic Planning and Coordination Division, U.S. Census Bureau, Washington, D.C. 20233-6100 <sahmed@census.gov>; Lawrence A. Blum, Assistant Division Chief for Collection Processes, Economic Planning and Coordination Division, U.S. Census Bureau, Washington, D.C. 20233-6100 <Lawrence.A.Blum@ccMail.Census.GOV>; Mark E. Wallace, Chief, Economic Planning Staff, Economic Planning and Coordination Division, U.S. Census Bureau, Washington, D.C. 20233-6100 <mwallace@census.gov>.*

Government Information Quarterly, Volume 15, Number 3, pages 275-302.
 ISSN: 0740-624X

- Basic Concepts and Methodology elaborates the terms through which the Economic Census is conducted—the language of the census. These concepts include the (business) establishment, the company, the North American Industry Classification System (NAICS), and the Standard Statistical Establishment List (SSEL).
- Centralized Collection Processing covers the actual launching of the Economic Census. It highlights the role of the new DocuPrint technology in mail collection of data, electronic reporting, forms mailout, receipt and check-in, completeness and coverage of multi-establishment data, microfilming, data screening/keying and control file match.
- Decentralized Post-collection Processing covers the intricate processes that take place after the Economic Census data are collected. Accompanied by several graphics, this section shows the series of tasks that must be performed to make the answers to the questions usable: assigning of classification codes, determining the validity of the answers, creation of databases, tabulation of data, and the like.
- Publication of Data/Distribution of Results emphasizes one of the chief reasons for which censuses are conducted in the first place: that is, to make the results available to both public and private sector users. As such, this part of the article treats the topics of new, standardized table formats for 1997 Economic Census data, the development of a streamlined data production and dissemination system, the use of CD-ROMs and the World Wide Web (one part of the Internet) in making results available, and the advantages of access, speed, and consistency that users will enjoy as a result of these improvements.

Title 13, *U.S. Code*, mandates the taking of the Economic Census once each five years, for years ending in "2" and "7". The law directs the taking of the census for manufactures, minerals, construction, retail trade, wholesale trade, selected service industries, plus finance, insurance, real estate, and transportation. In addition to this basic coverage of business establishments, the Economic Census includes supplemental programs providing special information about the transportation industry, women- and minority-owned businesses, and the characteristics of business owners. The geographic scope of the census also is specified in the census law. The census covers each of the 50 states, the District of Columbia, the Virgin Islands, Guam, the Commonwealth of the Northern Mariana Islands, and the Commonwealth of Puerto Rico.

Census Bureau Organization

The Economic Census can be conducted successfully only because of the breadth of specialized resources and talents throughout the Census Bureau. As can be seen in Figure 1, the Bureau is organized along generally functional lines.

The *Director*, together with the *Deputy Director* and the *Principal Associate Directors*, serve as the Bureau's chief administrators, with the *Associate Directors* reporting to the *Principal Associate Directors* or the *Deputy Director*. The *Associate Directors* formulate the policies and provide the direction for carrying out the programs in their areas of responsibility. For the Economic Census, that person is the *Associate Director for Economic Pro-*

grams. Organized under that Directorate are the *Assistant Director for Economic Programs* and the divisions that plan and conduct the Economic Census.

The processing operations of the Economic Census are planned and coordinated by the *Economic Planning and Coordination Division (EPCD)*. EPCD also performs all maintenance activities of the Standard Statistical Establishment List (SSEL), which is the source list for mailing the Economic Census forms. This includes obtaining and processing administrative records from other federal agencies. In addition, EPCD is responsible for developing the census lists; mailing the questionnaires; planning and coordinating all centralized processing in connection with data collection operations, as well as decentralized post-collection operations; and formulating and implementing the plans for marketing and disseminating data results.

Also reporting to the *Associate and Assistant Directors for Economic Programs* are three subject divisions and the *Economic Statistical Methods and Programming Division (ESMPD)*. These subject divisions are responsible for planning the content[1] of the Economic Census and for analyzing, tabulating, and providing clearance for publication of the data. They include the *Manufacturing and Construction Division* (*MCD*) (with overall responsibility for census coverage of Construction, Manufacturing and Mining); *Service Sector Statistics Division* (*SVSD*) (with overall responsibility for census coverage of Retail Trade, Wholesale Trade, Finance, Insurance and Real Estate, Transportation and Utilities, and selected Services Industries); and *Agricultural and Financial Statistics Division* (*AFSD*) (with overall responsibility for census coverage of Outlying Areas, women- and minority-owned businesses, and the characteristics of business ownership). [Editor's Note: As of May 1, 1998, AFSD was renamed Company Statistics Division.] In turn, ESMPD has overall responsibility for the systems design and programming functions for the Economic Census.

Several other divisions of the Census Bureau participate in conducting the Economic Census:

- *Data Preparation Division (DPD)* is the Bureau's processing center in Jeffersonville, Indiana, and is under the *Associate Director for Field Operations*. Staff in DPD perform the large-scale clerical and related operations crucial to the census such as labeling questionnaires, assembling mailing packages, entering the reported data into the computer, and corresponding with respondents. [Editor's Note: As of May, 1998 DPD was renamed the National Processing Center.]
- *Geography Division*, under the *Associate Director for Decennial Census*, develops the geographic coding system and assigns geographic location codes to each establishment in the Economic Census.
- Awareness of the Economic Census involves the efforts of three offices under the *Associate Director for Communications*. These offices are the *Public Information Office* with responsibility for informing the general public or business sector about the Census Bureau's work; the *Congressional Affairs Office* which monitors congressional actions that may affect the Census Bureau and which provides a central point for congressional inquiries about Bureau activities; and the *Customer Liaison Office* whose mandate is to help data users learn about, acquire, understand, and use Census Bureau statistical reports. Moreover, 12 U.S. Census Bureau Regional Offices within

Note: As of May 1, 1998, the Associate Director for Administration and Comptroller was renamed the Associate Director for Finance and Administration. Also, the Data Preparation Division was renamed the National Processing Center and the Agriculture and Financial Statistics Division was renamed Company Statistics Division.

Figure 1
Organizational Structure of the U.S. Bureau of the Census

Field Division, under the *Associate Director for Field Operations*, also publicize and disseminate Economic Census data products.

- Supporting staff manage and operate the computers and peripheral equipment used to process the Economic Census, to develop and maintain operating systems and associated software, and to plan and perform various engineering services. These functions are under the responsibility of the *Associate Director for Information Technology.*
- Under the *Associate Director for Administration and Comptroller* are the divisions performing many of the other support services required to conduct the census. These include budget and finance, personnel, and the preparation of publications. [Editor's Note: As of May 1, 1998, this Associate Directorship was renamed the Associate Director for Finance and Administration.]
- The divisions and staff under the *Associate Director for Methodology and Standards* are responsible for ensuring that appropriate statistical methods and techniques are followed in taking the Economic Census.

BASIC CONCEPTS AND METHODOLOGY

Certain crucial concepts underlie the methodology of Economic Census-taking. These are basic to both the preparation and use of statistical information. First, statistics for the Economic Census are collected and summarized for publication primarily in terms of the *establishment.* Second, establishments are classified and statistics are summarized using the new *North American Industry Classification System (NAICS).* Finally, the Census Bureau's *Standard Statistical Establishment List (SSEL*), which is compiled by reference to administrative records of businesses, provides the frame (i.e., universe) from which the list of establishments is selected for conducting the Economic Census. To understand the census it is essential to understand these concepts.

Establishment/Company Classification

For statistical purposes, an *establishment* is defined as a business or industrial unit at a single geographic location that produces or distributes goods or services—for example, a factory, store, or hotel. An establishment generally is the smallest basic unit for which key economic data—such as employment, payroll, and the value of products or services produced or sold—are available. Thus, the "establishment" concept provides for a highly detailed and definitive level of data collection and publication. The result is great latitude in how the data can be used: for example, information can be published not only on a very detailed industrial and geographic basis, but also summarized to much broader organizational, industry, and geographic levels.[2]

Aggregating data to broader organizational levels requires that a firm's ownership, affiliation, or structure be known. If an *establishment* is the physical location where goods or services are produced or distributed, then the *company* is the organizational entity (i.e., headquarters) that *owns the establishment or establishment(s).* For most businesses, the organizational structure is quite simple—one establishment constitutes the entire company. Approximately 85% of all establishments counted in the Economic Census are single-establishment companies. Other establishments, however, are operated by companies with

complex structures. Figure 2 shows the relationship between the two census concepts of *establishment* and *company*. the figure also illustrates the NAICS codes, which will be discussed in the next section.

The North American Industry Classification System (NAICS)

The 1997 Economic Census will be the first among the three North American nations—the U.S., Canada, and Mexico—to showcase the new classification system that was created to capture the many changes in the composition of economic activity in North America. This new system is the North American Industry Classification System (NAICS), and it replaces the Standard Industrial Classification (SIC) that was developed and used in the U.S. since the 1930s, albeit with periodic revisions.[3]

In vital economies, new goods and services are being offered constantly, and some of the old ones decline or disappear. The first step in measuring what goods and services are currently being offered is to classify them. The SIC classified the U.S. economy into 10 sectors: Agriculture, Forestry and Fishing; Mining; Construction; Manufacturing; Transportation, Communications, Electric, Gas, and Sanitary Services; Wholesale Trade; Retail Trade; Finance, Insurance and Real Estate; Services; and Public Administration. The subclassifications of these major groupings were modified a number of times throughout this century—the last time, in 1987— to enable the basic 10 SIC categories to accommodate new types of economic activity.

Finally, the point came at which the emergence of new industries and the decline of older ones made further adjustments to the SIC untenable. NAICS was developed over

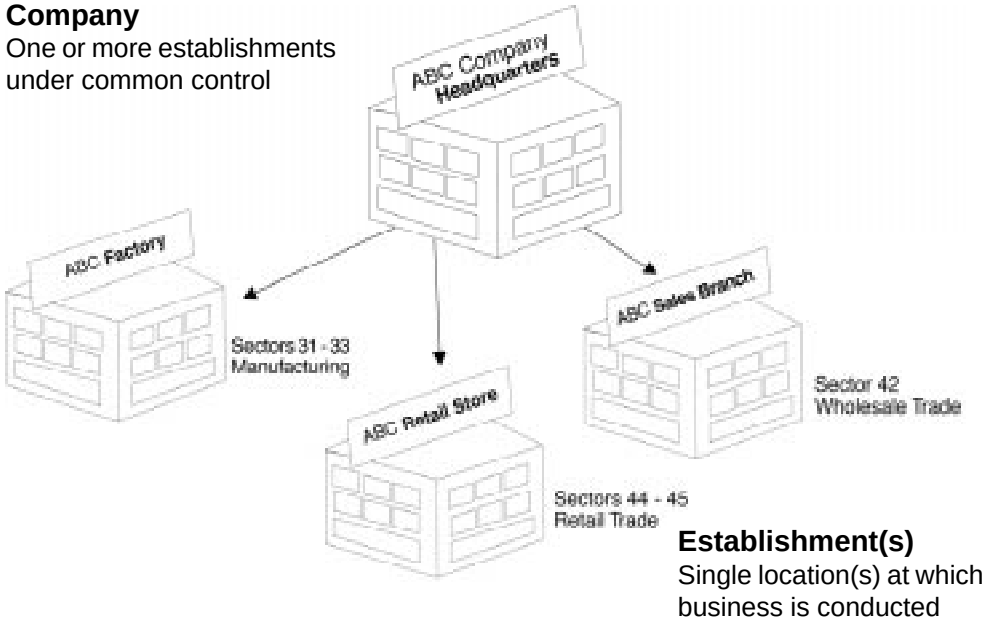

Figure 2
The Relationship of the Establishment to the Company
and the Classification of Economic Activity Using the New NAICS Codes

a five-year interval, 1992–1997, to replace the SIC and to consolidate the classification of all economic activity on the North American continent. With its 20 sectors, NAICS added a tremendous new vocabulary to "the language of the Economic Census." The 20 sectors of NAICS and their numerical classification codes are: Agriculture, Forestry, Fishing, and Hunting (11); Mining (21); Utilities (22); Construction (23); Manufacturing (31–33); Wholesale Trade (42); Retail Trade (44–45); Transportation and Warehousing (48–49); Information (51); Finance and Insurance (52); Real Estate and Rental and Leasing (53); Professional, Scientific, and Technical Services (54); Management of Companies and Enterprises (55); Administrative and Support, Waste Management and Remediation Services (56); Educational Services (61); Health Care and Social Assistance (62); Arts, Entertainment, and Recreation (71); Accommodation and Food Services (72); Other Services [except Public Administration] (81); and Public Administration (92).

NAICS will allow for the presentation of more detail for the rapidly expanding service sector that accounts for most economic activity, but accounted for only 40% of SIC categories. However, the objectives for the 1997 revision were even broader. Not only was the system designed to identify new industries, but NAICS also reorganized industry classifications according to a more consistent economic principle. That principle was the *type of production activity performed*. By contrast, SIC classified industries according to a mixture of production-based and market-based activity. Moreover, NAICS was developed in collaboration with Canada and Mexico to produce comparable statistics for the three North American Free Trade Agreement trading partners.[4]

Figure 2 illustrates the relationship of classification systems to the measurement of business activity—both across industrial sectors and across levels of organization. Using Figure 2 as an example, a 1997 Economic Census questionnaire package— using DocuPrint technology, which will be discussed below— would be mailed to ABC Company headquarters. In the package would be three separate census forms: one questionnaire *each* for the factory, the retail store, and the sales branch—that is, the "establishments" owned by ABC "company."[5] The data collected on the three separate questionnaires would be classified, respectively, in NAICS sectors 31–33 (Manufacturing), 44–45 (Retail Trade), and 42 (Wholesale Trade). In this way, the Economic Census accurately measures *what type* of economic activity is taking place (industrial sector), *where* (geographically), at *what level of organization* (establishment, company, or combination of companies—which are called "enterprises"), and *what the economic activity is:* employment rolls, size of payroll, volume of sales receipts or value of shipments, types of raw material used, costs of operating expenses, costs of capital expenditures, and the like.

From this example, it is clear that accuracy of data collection relies upon accuracy of classification. We cannot measure any economic activity occurring in a context whose classification characteristics we have not clearly identified.[6] As we will see in the following section, the next task in conducting the census is to arrive at a valid list of to whom to mail the questionnaires. Armed with the new NAICS, we can more precisely classify establishments than we ever did before—but first, we need a complete list. The SSEL points us in the right direction.

The Standard Statistical Establishment List (SSEL)

The SSEL is a database compiled and continuously updated by the U.S. Census Bureau. In it are maintained the records of about 6.5 million corporations, partnerships, sole proprietorships, and other organizations with employees. Its scope spans all economic activities (e.g., wholesale, retail, and services).[7]

The SSEL database contains basic economic data on U.S. businesses. Many of these data are obtained by the Census Bureau from administrative records filed by businesses with the Internal Revenue Service (IRS) or the Social Security Administration (SSA). These data include employment, payroll, sales and receipts, geographic location, industrial activity, and legal form of organization (e.g., corporation and partnership). It should here be mentioned that the relationship between the Census Bureau and the IRS is *not reciprocal*; that is, *the Census Bureau does not share its data with any other agency*. Census data are confidential, and are protected by Title 13 of the *U.S. Code*.

There are two chief reasons that the Census Bureau uses administrative records to obtain data on businesses:

- To ascertain the basic list of business establishments to which Economic Census forms should be sent. How would the Census Bureau know to whom to send census report forms without a means of determining the (universe of) businesses that are in existence? Administrative records provide one such means.
- To collect economic data on very small businesses—typically those "without paid employees"—without having to burden the small business owner with census forms. Stated otherwise, administrative data can often (but not always) be used in lieu of directly-collected information.

The use of administrative data makes possible a significant reduction in the size of the mail canvass that is used to conduct the Economic Census. In 1997, Economic Census questionnaires were mailed to approximately 3.7 million companies representing five million business establishments. But there were an additional 1.5 million small business establishments "with paid employees" and 14 million establishments "without paid employees" for which data were collected purely through administrative records of both (or either of) the IRS and the SSA. By using these already available tax records, the census could be conducted more efficiently, at less cost, and with a lighter burden on the small business community. Moreover, these businesses accounted for less than 10% of total economic activity measured in the 1997 Economic Census.

For various reasons, administrative record information cannot be used for larger firms: rather, Economic Census questionnaires are the only way to gather important information on products or activities, expenditures and assets, and operating characteristics that are not necessarily available from administrative sources. Moreover, the classification codes—now in terms of the new NAICS—from administrative records are less reliable than those that can be assigned based on information about the firm's products or activities that are reported on the census questionnaire. This is particularly true of larger, more complex businesses. Therefore, for the 1997 Economic Census, questionnaires had to be sent to all multi-establishment companies. These 165,000 companies accounted for 1.5 million establishments.[8] With this, the discussion turns to *levels of company organization* and the

importance of the SSEL in determining those levels so that the Economic Census can be conducted on the basis of accurate lists.

Conducting the Economic Census efficiently and accurately requires that all components of a company be identified and linked together. This process involves uniquely identifying three *levels of organization* for each *multi-establishment company*: (1) the enterprise or parent company; (2) each legal entity or subsidiary which, for tax reporting purposes, has been assigned an Employer Identification Number by the IRS; and (3) each establishment operated by the company. The means through which this linkage is accomplished is the SSEL.

The mailing list for the Economic Census comprises both single- and multi-establishment companies selected from the SSEL database. It is this database that provides information that administrative records alone cannot: for example, administrative records may cover entire firms, or other legal entities, in a way that provides no information about the location and kinds of businesses at the separate locations within multi-establishment companies. From the information in the SSEL, however, such crucial information about company structure can be derived.

For example, an *enterprise* is an entire economic unit consisting of one or more *companies.* Composition may vary from *a single legal entity*—including, for example, a corporation, partnership, or individual proprietorship with only one establishment—to *a complex family of legal entities* and their constituent establishments. If an enterprise is owned or controlled by another enterprise, all establishments of the subsidiary company are included as part of the "parent" enterprise.[9]

Following this line of reasoning, while it is true that the Economic Census collects data on *each establishment,* nevertheless, the questionnaires usually are mailed to the *company headquarters* (as shown in Figure 2). This is to ensure complete, unduplicated coverage of all of a company's establishments and activities. Stated otherwise, the Census Bureau obtains individual reports for each establishment of multi-establishment companies or enterprises.

Therefore, for Economic Census purposes, there are at least three objectives in linking together all legal entities and their establishments:

- To facilitate centralized mailing, collection, and correspondence relating to questionnaires for a company or an enterprise;
- To collect and publish data for enterprises as well as establishments; and
- To ensure, by linking all related companies under a master number, that there is complete coverage of changes within companies and to ensure that there is no publication of data that might reveal information about that company.

CENTRALIZED COLLECTION PROCESSING

A Short Overview

Plans for processing the 1997 Economic Census began in 1994. A thorough analysis of the processing operations used in previous censuses was conducted to determine where delays occurred and what operations needed to be improved or re-engineered to reach the goal of more timely publication of results. Questionnaires were mailed out in November and December 1997, and as completed questionnaires were received at the Census Bureau's processing center (DPD) in Jeffersonville, Indiana—near Louisville, Kentucky—the

reported information was keyed and transmitted electronically to headquarters in Suitland, Maryland. In turn, the information that was derived from administrative records also was processed at the Census Bureau's computer center in Suitland. After the questionnaire data were transmitted from Jeffersonville, the data were then merged with tax data from administrative records and given to the subject divisions for further processing and analysis.

Mail Collection of Data: The Role of the New DocuPrint Technology

The advanced operations of conducting a vast mail canvass require systematic planning and cooperation among a number of Census Bureau organizational units. Questionnaires must be designed, printed, and assembled into mailing packages. In 1997, the entire multi-establishment mailout system was re-engineered using DocuPrint technology. This technology made possible the simultaneous printing of all forms pertaining to a single company. This included cover letters, inserts, and the like, so that minimal clerical assembly was required. Previously, the Census Bureau relied on a commercial organization to provide preprinted forms. This required a massive clerical assembly operation to put the different forms of a company together in one package. In addition, an extensive quality control operation was needed to ensure that the company package included the proper forms.

By contrast, the DocuPrint technology stored all forms as "images," thus eliminating the dependency on the outside organization, as well as the intense manual assembly line needed in the past. Moreover, there were about 475 variations of the 1997 Economic Census questionnaires, and *only* inquiries about the operations of, or products produced by, a particular industry, or closely related industries, were included on a questionnaire for that industry. A further advantage was that, since the DocuPrint technology produced a company package at the outset, the former large quality control requirement was virtually eliminated.

The multi-establishment mailout system also was redesigned to identify easily two other contingencies: "split mail companies" and those who wished to report electronically. This was accomplished through the design of a parameter driven identification system. This system allowed an analyst to identify which parts of a company needed questionnaires to be mailed separately to *another* location within the company (i.e., "split mail") and/or allowed companies who wished to report electronically also to receive the proper electronic reporting medium within the company package.

For single-establishment companies, the Census Bureau relied on commercial organizations to provide pre-assembled mailing packages. Using specialized labeling equipment, these packages were then addressed—by the Census Bureau—through the open window envelope.

Classification Forms Mailout

As mentioned, questionnaires were mailed to approximately 3.7 million companies representing five million business establishments in November and December 1997. This included about 1.5 million classification forms generally designed only to obtain the more detailed kind-of-business classification information than was available from the administrative records. Because of the first-time use of NAICS, the 1997 mailout of classification forms was much larger than usual for a U.S. Economic Census. For these cases—and for the remaining 1.5 million nonmail businesses with paid employees, as well as for the 14

million businesses without paid employees—the basic statistics were obtained from administrative record tax sources.

Finally, it should be noted that supplemental mailings—over the several months succeeding the first mailouts—covered approximately 200,000 firms that had gone into operation *after* the basic mail list was developed.

Electronic Reporting

Electronic reporting initiatives for the 1997 Economic Census, like previous censuses, focused primarily on large retail, food service, and hotel chain enterprises. The Census Bureau contracted with a private company to develop an enhanced Windows-based Computerized Self-Administered Questionnaire (CSAQ) to facilitate electronic reporting in the 1997 Economic Census. CSAQ is a diskette-based questionnaire which runs on a personal computer.

Benefitting from prior experiences with CSAQs, the Bureau established challenging requirements. The Economic Census CSAQ had to be able to accommodate reporting for multiple establishments within an enterprise; cover reporting for 21 different report forms with content varying by form; handle a variable number of data items within a form; and perform selected interactive editing. Additional requirements included these: enterprise capability to export and manipulate information provided by the Census Bureau such as establishment identifiers, and name and address information; multiple import capabilities that provide the enterprise or an establishment within an enterprise the capability of linking to internal corporate spreadsheets and databases; and self-contained communication software that would permit modem transmission of the census data.

As in past censuses, the Census Bureau continued to offer magnetic tape reporting within the retail, services, finance and insurance, and utilities sectors, plus a CSAQ in the finance and insurance sectors.

Electronic reporting media were mailed to approximately 600 companies representing over 250,000 establishments during the original mailout process. The Census Bureau received inquiries from an additional 100 companies after mailout.

The actual mail/nonmail establishment counts for the 1997 Economic Census are shown in Table 1.

Receipt and Check-In

Completed questionnaires are returned to the Jeffersonville, Indiana processing center. Supermarket-type bar-coded labels and high-speed sorting equipment using laser

Table 1
Mail/Nonmail Establishment Counts

Economic Census/SSEL	*Mail/Nonmail Establishment Counts*		
	Total	*Mail*	*Nonmail*
Total Establishments	6,500,000	5,000,000	1,500,000
Multi-establishments	1,500,000	1,400,000	0
Electronic Reporters	250,000	250,000	0
Single-establishments	5,000,000	3,500,000	1,500,000
Classification Forms	1,500,000	1,500,000	0

technology are used to rapidly record their receipt. The high-speed equipment is augmented by hand-held wand stations linked to microprocessors to record the bar codes on the large mailing packages containing multi-establishment questionnaires (plus any single-establishment reports not neatly placed in the return window envelopes that were provided with the questionnaires). In addition to eliminating all hand-sorting and manual counting of questionnaires, the system ensures that the receipt of questionnaires and related correspondence are reflected in control records almost immediately, and that only those establishments for which reports are delinquent are included in subsequent follow-up mailings. Next, all multi-establishment report forms are transmitted for the completeness and coverage operations. Single-establishment forms are transmitted for microfilming.

Multi-establishment Data—Completeness and Coverage

The questionnaire data received from multi-establishment companies undergo extensive computer review to ensure that: (1) all establishments of the company are accounted for; (2) all essential individual establishment and company data are reported; and (3) company affiliation information—such as new adds, mergers, and sold cases—are properly handled. These operations produce a current, unduplicated list of companies and their establishments in the SSEL, and assure that total company employment and payroll are consistent with administrative record tax data.

A dedicated computer system linked to interactive terminals is used to meet the completeness and coverage processing requirements. By using a series of video display screens, each terminal operator is guided through the steps required to take the following types of actions based on returned questionnaires:

- Correcting company affiliation information such as address information; adding new or acquired establishments, moving sold establishments of a company to the acquiring company; inserting codes to indicate that establishments were closed or idle; and adding new multi-establishment companies and their establishments to the files.
- Researching "postmaster returns" for the latest addresses for remail of questionnaires.
- Correcting unusual employment and payroll data relationships as compared to administrative record tax data. Comparing reported data to administrative record tax data usually uncovers many new locations for which the company has not reported.
- Correcting establishment reports of the company where the company combined all the data on one report. This involves splitting the combined data across the establishment reports based on prior year employment and payroll relationships.

These problems are usually resolved by calling the company to correct obvious data errors when compared to IRS data, or by using the SSEL interactive routines to process the company affiliation changes or the combined reports. After the completeness and coverage operations, all multi-establishment questionnaires, along with those for single-establishment firms, are microfilmed.

Microfilming

As part of the microfilming process, a serial number, which becomes an integral part of the establishment record, is imprinted on each questionnaire. Analysts may later reference this frame number at an interactive terminal and locate any census questionnaire for review. This process avoids the costly and error-prone filing of the actual reports. Further, copies can always be printed from the microfilm when needed. After microfilming, all questionnaires are transmitted for data screening/keying.

Data Screening/Keying

Data screening/keying, or data entry, is the typing of information from the questionnaire onto video display screens. As the data are keyed, the computer performs simple range checks on entered data. The video terminal displays messages alerting the keyer to unacceptable entries along with what procedures to follow to correct records. After data screening/keying, the keyed data are transmitted electronically to the Census Bureau's computers at headquarters for the control file match operation.

Control File Match

After keying, questionnaire data are merged by computer with existing administrative records and Census Bureau historical data contained in the SSEL database. The primary functions of this operation are to: (1) update company structure and affiliation based on changes reported on the census questionnaire; (2) identify and delete duplicate establishments and add new establishments to the control file; (3) capture any address changes made to the census questionnaires so that geographic coding operations (explained below) can be performed; and (4) assemble all keyed data from both the questionnaire and the SSEL in a single file for later in-depth computer analysis.

For the geographic coding operation, the Geography Division of the Census Bureau maintains an up-to-date computer address reference file with corresponding state, county, place, and census tract numerical codes. To publish tables on economic activity by these detailed geographic areas, address information is needed on the physical location of each establishment. Prior to mailing the questionnaires, the list of establishment addresses—both those that would be sent questionnaires and those for which information would be obtained from administrative records—are coded based on the address reflected in the SSEL. Any address changes identified during the control file match are recoded based on the information reported by the respondent.

Records for which problems are identified are held until the problems are corrected. The merged file of "good" records in terms of organizational and coverage checks are released to the economic subject divisions for post-collection processing.

To summarize this section, Figure 3 depicts the main data collection operations for the 1997 Economic Census.[10] It includes the following operations: forms mailout, receipt and check-in, multi-establishment completeness/coverage, microfilming, data entry, and the control file match. The next section will cover the processing of the reported data through post collection operations.

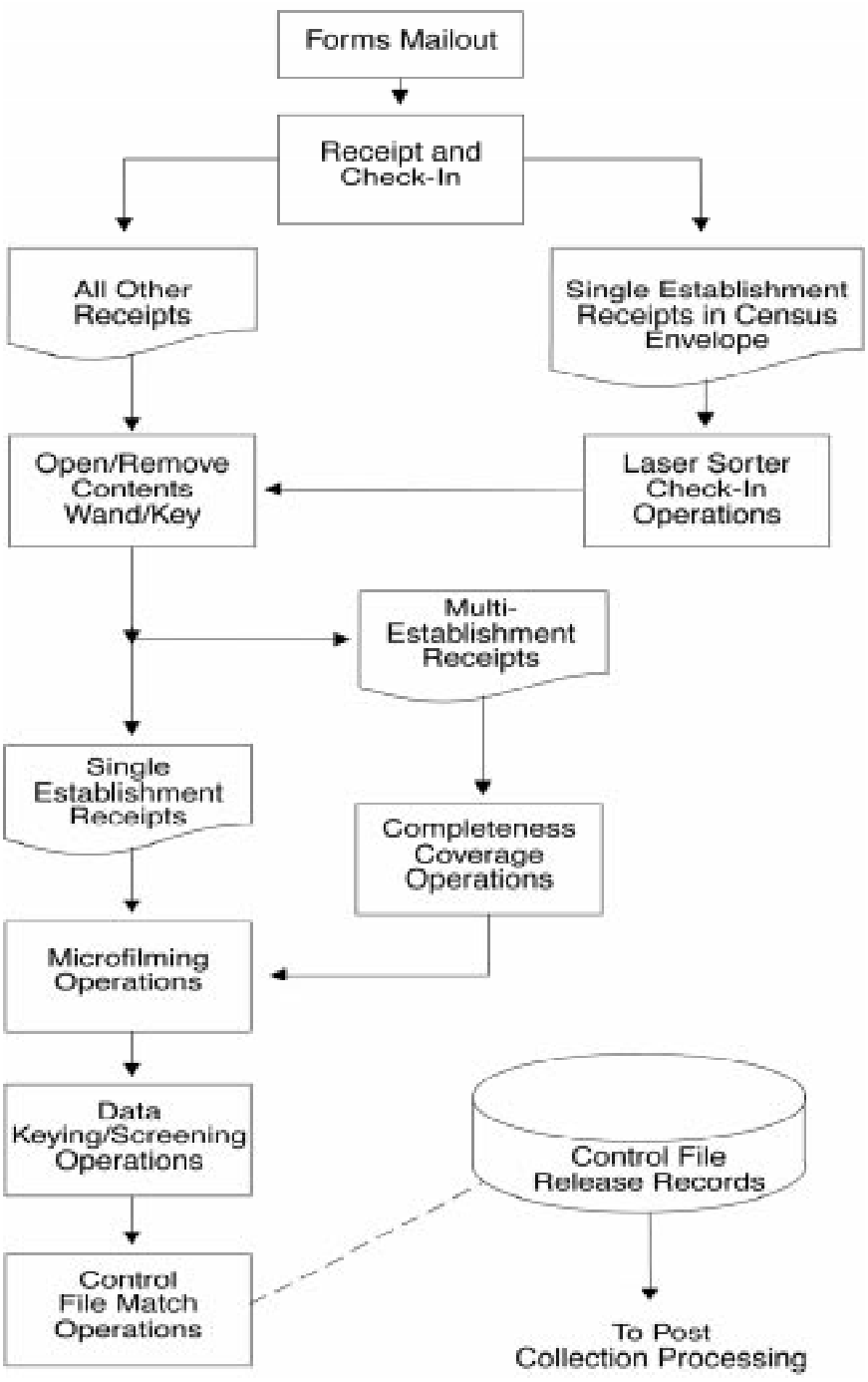

Figure 3
1997 Economic Census Data Collection Operation

DECENTRALIZED POST-COLLECTION PROCESSING

In the Economic Census the post-collection processing represents the steps conducted after the control file match, the last phase associated with the collection processing. The objective of the post-collection processing is editing, cleaning, summarizing, and analyzing the response data in preparation for dissemination to the public. The subject matter analysts in MCD, SVSD, and AFSD (now called Company Statistics Division) have the responsibility for this work, relying on additional analytical and clerical resources at the Jeffersonville, Indiana site. The two distinct phases of post-collection are the micro edit referral processing and the macro analytical processing.

Micro Edit Referral Processing

Micro edit referral processing involves further editing and cleaning of records on an establishment basis. The micro edit referral processing occurs in conjunction with the collection cycle, representing the next step in the pipeline. Figure 4 provides an overview of the micro edit referral processing.

Each subject area performs the tasks associated with the basic flow in Figure 4. Each processing block from Figure 4 is described below:

- Get Release Records: The first step in the micro edit referral process is to get the "good" records from the control file match. Recall that these records passed computer checks for organizational structure, coverage, and completeness as described in the

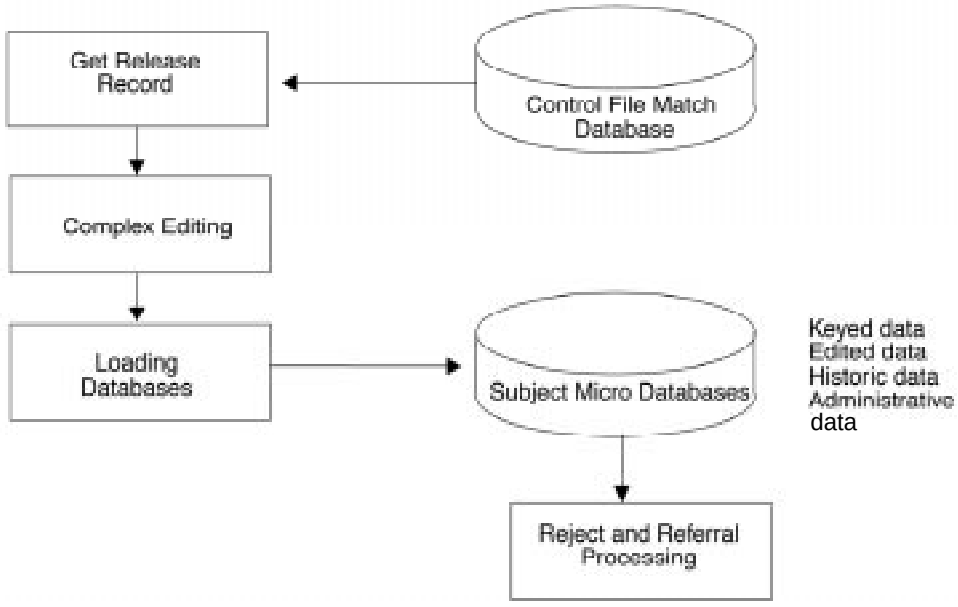

Figure 4
Basic Flow for Micro Edit Referral Processing

collection section. Once the organizational structure is reviewed and approved, the establishments are ready for post-collection processes, where review of response data becomes the focus. While the collection phase is concerned with organizational integrity, the post-collection phase is concerned with data integrity. Each subject area gets its own release records from the control file match database. Thus, post-collection becomes a decentralized process at this point in the census cycle.

- *Complex Editing*: Computer programs subject the retrieved records to a series of "complex" data edit modules. These complex edit programs have two major purposes. First, they assign a valid kind-of-business or industry code to classify the establishment. Assigning a valid classification code depends on computer evaluation of the responses to specific items on the census questionnaires. These items cover many aspects. For example, they include responses to merchandise lines sold by a retail establishment, products manufactured by a plant, entries written-in by the respondent explaining the establishment's activities, and self-designated check-box classifications. If this critical information is missing, the record is noted (i.e., flagged) as a "reject." It must be fixed (see discussion below) before further processing occurs.

  If critical information is available, the edit assigns the correct classification code. For the 1997 Economic Census the program assigns a "bridge" code that lets the Census Bureau ultimately tabulate the record on both the old SIC basis and the new NAICS. After classification codes are assigned, a "verification" operation is performed to validate the kind of business, industry, geographic, and zip codes.

  The second purpose of the complex edits is to evaluate the response data for consistency and reasonableness—for example, assuring that employment data are consistent with payroll or sales/receipts data. Evaluating the response data is done by industry. Additional checks compare data reported in previous censuses or from administrative sources. These checks are accomplished through generalized computer routines. The generalized routines are particularized by subject area to account for differences in industries and data collected on the questionnaires.

  The generalized routines cover modules to balance items and to conduct ratio checks.[11] Imputation options are employed to estimate noncritical missing data or to replace extreme outlier data. The balancing module allows for simple tests of details to totals; for nested tests of subtotals to broader levels; and for two-dimensional tests to compare additive rows to corresponding column information. The ratio module sets up comparisons of data for the establishment to parameters, developed by industry. These industry parameters are derived based on how the data have been correlated in past censuses or through administrative sources. The ratio module uses items from the questionnaires, administrative sources, or historically reported information. While suspicious records get through the pipeline, they have certain characteristics that question how the predefined computer rules applied to them. These establishments are flagged as "referrals" for manual review.
- *Loading Databases*: After the complex edits, computer programs load data into separate subject micro databases. Each subject area database contains the keyed responses from the questionnaire, the resulting data and flags after the complex edits, administrative data from the SSEL, write-ins from questionnaires, and historic 1992 census data (if available) for each establishment.

Table 2
Estimated Rejected and Referral Counts

Subject Area	*Total Establishments*		*Number of Rejects and Referrals*
	*Mail**	*Nonmail*	
Retail	1,350,000	220,000	115,000
Services Industries	1,740,000	450,000	185,000
Wholesale	455,000	0	100,000
Utilities	150,000	100,000	37,000
Finance, Insurance, Real Estate	405,000	125,000	87,000
Construction	130,000	405,000	80,000
Manufacturing	300,000	90,000	70,000
Minerals	17,000	10,000	4,000
Outlying Areas	40,000	0	20,000

Note: *Mail count includes the classification cards with the exception of the 328,000 unclassifieds.
Source: Based on 1997 mailout counts and expected referral counts.

- *Reject and Referral Processing*: The flags assigned during complex editing indicate establishments requiring further follow-up as part of the reject and referral processing. Table 2 shows estimated counts of the rejects and referrals expected throughout the processing for the 1997 Economic Census.

Establishments identified as rejects and referrals are handled through post edit correspondence and through manual problem resolution procedures. Where possible, the complex edits automatically identify establishments qualifying for post edit correspondence. These establishments receive correspondence requesting information on the missing items. In some subject areas, such as manufacturing, more manual procedures are used to mail correspondence to respondents requesting explanation to failed-edit information or to request more detailed information than what was previously reported. For example, detailed product data are needed to code manufacturing plants. Respondents reporting only broad product lines are mailed correspondence asking for the detailed product lines.

For manual problem resolution, analysts and clerks are involved in the data cleaning process, with clerks handling primarily the rejected, simpler cases. These are typically problems involving respondent write-in entries for kind of business descriptions, merchandise lines, or products. The clerks code the write-ins to the appropriate bridge code, type of operation, type of construction, or product code. The analysts and clerks use video-display terminals to review the rejects and referrals. They follow carefully prepared guidelines and procedures for evaluating and correcting problems. Update capabilities let them correct errors and reedit the data interactively. This means the analyst or clerk gets an immediate response to determine further review points. These sophisticated computer systems that operate in an interactive environment are distinguished from the batch computer process shown in Figure 4. The interactive environment is referred to as the “establishment review and correction system.”

Macro Analytical Processing

The objective of macro analytical processing is review of tabulated summaries in preparation for data release. The key macro analysis begins at the end of the collection cycle—in other words, after September 1998. Meaningful tabulations are not available until the Census Bureau receives a vast majority of questionnaires and completes most of the follow-up activities for missing and failed-edit establishments. Figure 5 shows the macro analytical process.

Starting in the left corner of Figure 5, the macro analytical process is described below:

- *Subject Micro Databases:* The starting point for the macro analysis is the tabulation of the establishment records housed in the subject micro databases. For the Economic Census, the tabulation of these records occurs along two paths. First, the micro records are tabulated as part of the main publication series associated with the Economic Census. Second, the micro records are tabulated for new cross-sector publications which showcase the introduction of the new NAICS.[12]
- *Data Tabulation:* The micro establishment records defined as "tabable" are used in the tabulations. Tabable records are those establishments which get through the com-

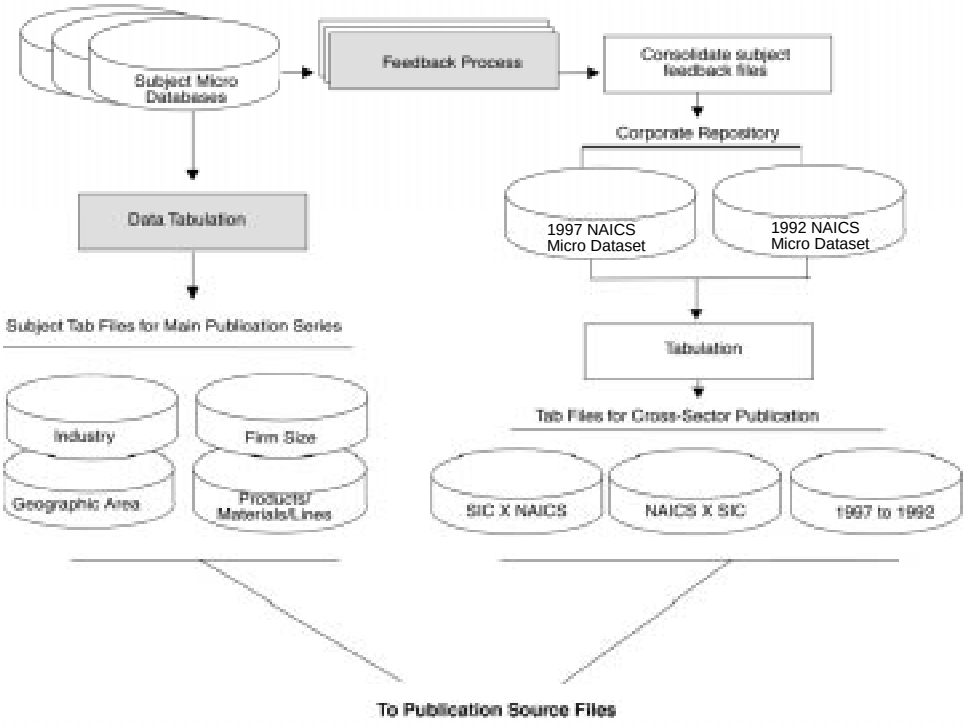

Figure 5
Macro Analytical Process

plex edits without being identified as a reject or referral. The objective of the reject and referral processing is to resolve problems such that the establishment becomes tabable. Invariably, the scope of this work leaves many problem establishments unfinished at the time of macro analysis. Once macro analysis starts, only the most significant offenders are corrected at this stage.

For the Construction sector, which is sampled and does not reflect a complete enumeration, records are weighted to equal universe totals. In general, tabulations occur frequently once the collection cycle ends. This is necessary to account for those corrections to micro records which occur at the macro review stage.

- *Subject Tab Files for Main Publication Series:* Once data are tabulated, the totals are loaded into computer files, which analysts access via macro analytical tools. The Census Bureau tabulates the micro establishment records many ways based on data product and analytical needs. For example, as shown in Figure 5, tabulated data exists by Industry, Geographic Area, Firm Size, Products Produced, Materials Used, and Merchandise Lines Sold, to name the major ones. Note, each subject area has its own set of tab files for its industries.
- *Feedback Process*: New for the 1997 Economic Census is a process to feed tabable micro records from the subject area databases to a corporate repository. Unlike the subject-specific systems, the corporate repository allows analysts to share data across the subject areas. For example, those subject analysts assigned to review Manufacturing also have access to view establishments in Retail industries.

  As shown in Figure 5, the tabable establishments are pulled as feedback records and consolidated into one file, labeled the 1997 NAICS micro dataset. Only basic data items are pulled for each establishment. That is, data items reflecting Sales/Receipts/Value of Shipments, Payroll, Employment, Operating Expenses, and Wholesale Inventories. Once created, a process combines the 1997 micro dataset with the 1992 micro dataset. The latter reflects all tabable records from the 1992 Economic Census. The resulting current and historic dataset makes possible the creation of special cross-sector tabulations and reports that allow the analyst to compare data on the new NAICS basis as well as on the old SIC basis. The corporate repository is a SAS® dataset.

Figure 6 brings together the heart of the macro analysis. The inputs, as discussed previously, represent the subject micro databases, the subject tab files, and the corporate repository datasets.

Figure 6 shows the analytical tools available to the analysts. Detailed analytical review guidelines are prepared to instruct analysts on what to look for and on how to use the available tools. Typical review points cover analyzing large differences, assessing changes to the tab cell composition due to organizational changes (establishments bought, sold, or new); nonresponse; and errors in historic 1992 data. For tabbed cells confirmed as correct, analysts document their findings. For corrections, they modify data and review re-tabbed results. Descriptions of the tools follow:

- *Subject Table Review Systems*: As with the micro edit referral processing, the macro analysis has specially designed interactive systems that let analysts view the subject tabs in table formats. Each subject matter area has an interactive table review system

for analyzing tabulated cells on a NAICS basis. As analysts correct micro data through the establishment review and correction system, they can then use the table review systems to determine the macro effect of these corrections. The subject table review systems are instrumental in the disclosure analysis.

Disclosure analysis preserves the confidentiality of reported information such that no individual firm's operations or identity can be inferred from the published data disseminated to the public. For 1997 data, consistent rules across all the subject areas are used to identify possible disclosures. Often when tabulated cells are disclosed (and not published), it is necessary to suppress additional tabulated cells so that the primary (originally) disclosed cell cannot be obtained through subtraction. The process of suppressing these secondary tabulated cells is called complementary disclosure analysis. The analysts use the table review systems to review and interactively set the complementary suppression symbols.

- *Problem Identification Tools:* To review thousands of tabulated data cells, the analysts need mechanisms to help pinpoint specific problem cells. The two main methods of identifying suspicious tabulated cells are reconciliation and outlier detection. For reconciliation, the analysts compare cell totals to sources outside the Economic Census. The outside sources cover the current programs within the Census Bureau as well as information from related government agencies.

  For outlier detection, statistical techniques are used to identify the suspicious cells. New for the 1997 census is the use of automated graphical data review techniques to accomplish this. Graphical techniques provide analysts with visual displays that let

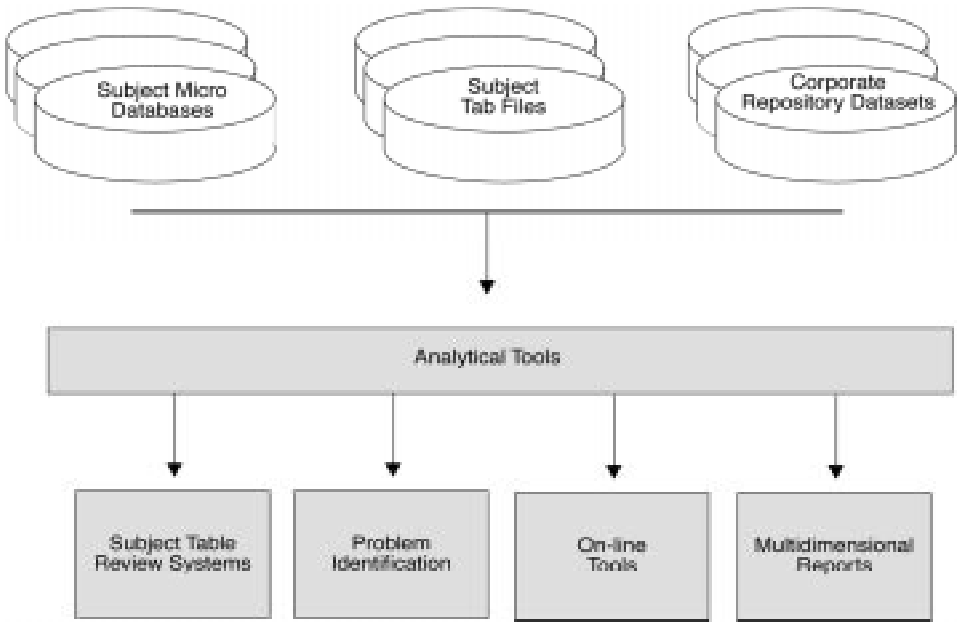

Figure 6
Analytical Tools

them see cells outside the normal distribution. The SAS ® product called "SAS Insight" makes this capability available. Of particular note, the graphical analysis is used to identify extreme cell changes for historic ratios (e.g., historic census employment compared to current census employment) as well as current ratios (e.g., current census annual payroll compared to current census employment).

In addition to graphical analysis, subject matter areas use manual techniques to identify cell outliers. Generally, analysts specify predetermined values for their historical and current ratios. For example, since the Manufacturing sector has prior year comparison points with the Annual Survey of Manufactures, analysts search "census to prior-year" totals exceeding 5% for basic statistics items and 10% for products.

- *Online Tools*: Analysts have tools that let them navigate through the tabulated cells and through the micro establishment records. The predominant tool is the capability to search computer files. Analysts develop "canned searches" which they execute interactively for typical inquiries about the data. Analysts can modify the canned searches for ad hoc, unplanned situations that surface in data review. They specify and incorporate ratios as part of their searches to select records on predetermined size.

  In addition to the canned and ad hoc searches, analysts have specially designed computer interfaces that let them do special ad hoc tabulations. For example, using the corporate repository, analysts can select and tabulate the establishments associated with specific companies — even if these establishments cross subject areas. They can use their subject micro databases to select and tab records by classification code (SIC, NAICS, or bridge), geography, and type of operation that differ in any way from the specified publication table formats.
- *Multidimensional Reports:* Used with the corporate repository, multidimensional reports provide an analysis tool to evaluate the new NAICS classification system. They provide, among other things, reports showing SIC industries based on their NAICS components and, vice versa, the NAICS industries based on their SIC components. Additionally, they provide the analysts with comparative SIC data showing the basic data from 1997 to 1992 at the broader levels.

Ultimately, the outcome of performing the various elements of macro analysis is the moving of the cleansed tab files to the next phase of census processing: the publication of data and distribution of results.

PUBLICATION OF DATA/DISTRIBUTION OF RESULTS

As for the centralized collection and the decentralized post-collection processing phases of conducting the Economic Census, the publication of data/distribution of results phase includes several innovations. Indeed, a series of improvements, over the course of earlier Economic Censuses, has culminated in several major breakthroughs as the Census Bureau prepares to release 1997 Economic Census data. These breakthroughs in the publication of data and distribution of results have come in response to requests from the data users. They have told the Bureau that they want timely data, available immediately on the day of release, with ready capability to find and use the requested data for multiple economic sec-

tors and years. With respect to the 1997 Economic Census results, the Census Bureau is well-poised to honor this request.

Publication Is Increasingly Electronic

To meet these needs, the Bureau has greatly augmented the electronic component of its publication system. Today, with CD-ROMs and the Internet becoming a large part of the public's information toolkit, data users are requiring an increasing amount of information—particularly in electronic form. The Census Bureau, therefore, will prepare and release the majority of data from the 1997 Economic Census via improved CD-ROMs (taking advantage of improved software technologies) as well as via the Web—with most data on the Web available free of charge.

In turn, the Census Bureau is reducing its output of paper products. To ease this transition from hardcopy to electronic media, some data, tabulated at the national level, will still be issued in hardcopy. In addition, the CD-ROM and Internet products issued from the 1997 Economic Census will provide the capability to do on-demand, publication-quality printing for all key reports that had been issued in print for earlier censuses. In technical terms, the data tables that used to be issued in print will henceforth be accessible in page image forms via portable document format (PDF) files. This will allow users with the free Adobe Acrobat reader to print high quality pages which will be identical in appearance to the previously printed Table Image Processing System (TIPS) generated tables.[13] Further, the Bureau is planning to archive the 1997 files, as had been done for the 1992 files, so that future users can refer back to the 1997 data in PDF files.

Thus, the Census Bureau has established electronic products as the primary focus of the 1997 Economic Census data dissemination program.[14] Data dissemination via the Web will ensure instantaneous global availability of Census results on the day of release. And users will have a wide range of improved CD-ROM and Internet data access software features at their fingertips, enabling them to locate and use the data with ease. The processes that led up to improvements in these products will now be discussed.

A New Streamlined System to Produce and Disseminate Data Products

An interdisciplinary group—called the Economic Product Team (EPT)—within the Census Bureau was formed early in the publication planning phase for the 1997 Economic Census. This group was assigned an imposing task: develop a streamlined data production and dissemination system, and make that system compatible with a focus on electronic products. The EPT had its work cut out for it: in 1992, there were two separate systems for products—one for printed products and one for manipulable or electronic products. The result of having two separate systems was that there were errors and inconsistencies between the two products. The errors and inconsistencies caused very high resource expenditures, processing bottlenecks, and ultimately, major delays in the release of data products. What would it take to resolve these problems?

The Publication Source File

If two systems created inconsistencies, then the use of one system should resolve them. It did: the "one source" or "publication source file" was devised by the EPT. *Only one*

source of data was used to create all Internet, CD-ROM, and printed products for dissemination to the public. How do data get into the publication source file, and what happens to them after they are put there? Figure 7 displays these relationships.

As mentioned in the post-collection discussion earlier in this article, it is the work of subject analysts to review data and approve it as final. Once tabulated data are approved, the programmers generate the publication source file based upon specifications by subject analysts.

Analysts next provide the instructions that tell programmers what to extract from the file for various data products (e.g., Internet, CD-ROM, and page image). The subject analysts provide these instructions, called "data dissemination parameters," to the programmers electronically via the Interactive Parameter System (IPS).

After this stage has been reached, data from the publication source file can be extracted by the programmers. *From this point onward, the creation of data products proceeds down two parallel paths, resulting in (1) the manipulable files and (2) the page image files.* That is,

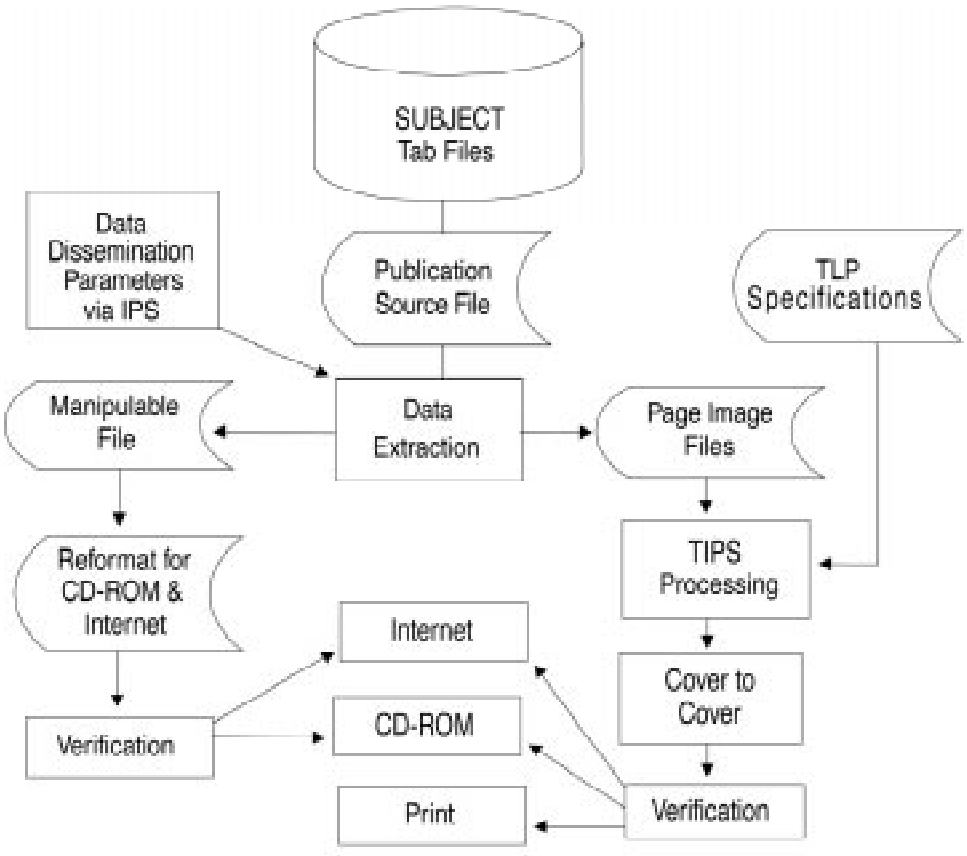

Figure 7
A New Streamlined System to Produce and Disseminate Data Products

- The manipulable files require only a brief format verification, taking not more than a couple of hours before they are ready to be released on the Internet. With this new system, the contingencies that had, in the past, to be met to produce a *printed* report do not delay the release of the data.
- The page image files, in turn, must be combined with the table text, then reviewed, and finally, converted for Internet and/or sent to print. These steps can take anywhere from several days to a number of weeks to complete.

Centralized Programming Efficiencies

The need to provide efficient programming to implement these specifications and instructions led to another important component of the new, streamlined production system: centralization. *Namely, the responsibility for all TIPS programming for the page image products has been centralized within the economic programming area.* Using the publication source file, together with the data dissemination parameters provided by the subject analysts, the programmers are developing a generalized system to extract automatically the appropriate data for TIPS tables. *This same system will, in turn, create the data files that will be the input for the final CD-ROM and Internet files.* This is a major innovation that will significantly reduce the programming workload, essential to the timely issuance of the publications.

The Meta-data Source File

"Meta-data" is an umbrella term used to refer to all supporting materials in a table, or about a table, except the actual numerical results for which the public awaits release every five years. Meta-data, therefore, includes table stubs and data labels, definitions, explanatory text, as well as other auxiliary files required to drive electronic software.

With respect to meta-data, it is the subject analysts that are responsible for providing the appropriate meta-data (for their respective subject areas) to a centralized database. A metadata coordinator then oversees the collection of the meta-data from the subject areas into a "meta-data source file" from which—in an operation similar to that for the publication source file—subject analysts and programmers coordinate the distribution of the appropriate meta-data to Internet, CD-ROM, and page image products. (Some of the meta-data required for the data products for 1997 already exists in an electronic format from 1992. Other meta-data needs to be generated anew for 1997 data).

Important examples of meta-data are the table "shells" or outlines. The Census Bureau's editors create the table shells based upon specifications (called Table Layout Parameters) by subject analysts: that is, how many rows and columns, and in what relationships to one another, will be needed to present the data? Will data be presented, for example, at the levels of nation, state, county, and/or metropolitan area? And, in turn, at what level of industry classification detail (in terms of the NAICS categories discussed earlier) will data be presented in the tables? Merging these supporting metadata with the actual data files is thus required to produce the complete data product.

Standardization of Major Data Product Components

Given the decided focus on electronic products for 1997 Economic Census data, standardizing data presentation is critically required in the age of the Internet, with its potential for instant comparison of data. Here, radical breakthroughs have been indeed achieved. Sector-by-sector conformity in terms of data-driven stub displays; methods of displaying geographic data; and units of measure will be notable features of 1997 data presentation. Certain artifacts (such as brackets, underlines, or data-driven footnotes) of printed reports previously varied by sector. These variations have been eliminated for 1997. Therefore, in addition to its improving the presentation of data, standardization saves processing resources central to the timely completion of the 1997 Economic Census data dissemination program.

Creation of New Data Access Software

Economic subject analysts and programmers are redesigning and developing new CD-ROM software. Based both upon the results of several usability studies (in which data users came to the Census Bureau and tested software packages capabilities) and also upon specifications from the subject matter areas, the goal is to take the best capabilities of the previous software packages (EXTRACT and GO) and to create, instead, a new windows-based software that provides those capabilities and more.

In addition, a dedicated Census Bureau-wide team is working with the subject areas to create data access software for the Internet. The Internet application will allow the user to do most of the same things that CD-ROM software permits, but will provide the user access to the latest Economic Census data, as well as access to all the other data on the Census Bureau Web site.

To summarize this discussion of data publication and distribution of results, the Bureau has achieved greater processing efficiencies through these innovations:

- The introduction of the publication source file;
- An improved processing and programming infrastructure; and
- The meta-data source file.

The 1997 Economic Census demonstrates substantial trailblazing in product presentation with its standardization of major product components (table layouts, text, and file structures) to promote uniformity across the 1997 Economic Census Product line. This product standardization is critically important today, not only because of its potential for resource savings, but also because improved electronic access—which enables users to mix and match data from various sources more easily—calls for consistency in data presentation.

CONCLUSION

As outlined in the various major sections above, conducting the Economic Census is a very intricate process. It requires synchronization and sequencing of processes as well as the efficient use of staff time and dollar resources. It demands great sophistication in the use of advanced technology to realize the promise of such technology for data collection, post-

collection, and dissemination. Major improvements have been introduced in the form of the new DocuPrint technology and electronic reporting techniques. Similarly, the use of generalized editing methods across subject areas, shared data through the corporate repository, and new problem identification tools—such as graphical data review techniques—add efficiency and quality to the post-collection processing phase. And, the advent of new electronic data products and media—produced and delivered by a streamlined data dissemination system—break new ground in Economic Census data release.

Ultimately, users of 1997 Economic Census data will be able to derive the key messages behind the numbers—both for previously delineated economic sectors, as well as for new NAICS sectors—as never before.

NOTES AND REFERENCES

1. See Judy M. Dodds, "Determining Economic Census Content," *Government Information Quarterly,* 15 (1998):247–262, for a detailed treatment of all the criteria that the Census Bureau must take into consideration—and the important "players" (such as the Bureau of Economic Analysis, other federal agencies, and many other types of data users) that must be consulted—before determining which questions will be on the Economic Census forms.
2. See Mark E. Wallace, "Public and Private Sector Uses of Economic Census Data," *Government Information Quarterly,* 15 (1998): 321–336, which shows how organizational, industry, and geographic levels of measurement form veritable "building blocks" of data that can be combined in limitless ways by data users.
3. See Carole A. Ambler & James E. Kristoff , "Introducing the North American Industry Classification System," *Government Information Quarterly,* 15 (1998):263–273. This article—by the current chairperson of the Economic Classification Policy Committee (the interagency committee chartered to develop NAICS) and by a participant on the interagency subcommittee that developed the manufacturing sector of NAICS—authoritatively covers the emergence and characteristics of the new system.
4. See Paul T. Zeisset & Mark E. Wallace, "How NAICS Will Affect Data Users," published on the Internet at <*www.census.gov/epcd/www/naicusr.html*>. This brochure also is available in paper (Lanham, MD: Bernan Press 1997).
5. In previous Economic Censuses, "enterprise statistics" were collected by mailing a census questionnaire for ABC Company headquarters to fill out, too, on its payroll and other such characteristics. However, in the 1997 Economic Census, only Form 9901 is sent to ABC Company headquarters, a one-page form that asks for data on company ownership and control. If ABC Company (a multi-establishment company) *also* happens to have on its physical site a manufacturing plant, a warehouse, or other type of activity, then ABC Company receives not only Form 9901, but also the appropriate census questionnaire for the economic activity taking place at that site.
6. One aspect of measurement that will be greatly improved by using NAICS is the determination of market share. For example, under SIC, the Bed and Breakfast Inn industry was in Subsector 7011, Hotels and Motels. Sharing that category were the following: auto courts; cabins and cottages; casino hotels; hostels; hotels, except residential; inns, furnishing food and lodging; motels; recreational hotels; resort hotels; seasonal hotels; ski lodges and resorts; tourist cabins; and tourist courts. Under NAICS, Bed and Breakfast Inns now have not only their own, separate classification, but are classified as "bed and breakfast inns with 25 guestrooms or more" and "bed and breakfast inns with less than 25 guest rooms." This is the precision for which NAICS will be hailed by data users across the entire spectrum: public, private, individual, organizational, research-oriented or policy-oriented.
7. For the definitive treatment of this subject, see Edward D. Walker, "The Census Bureau's Business Register: Basic Features and Quality Issues." Presented at the Joint Statistical Meetings. Anaheim, CA, August 10–14, 1997.
8. The timely response to census questionnaires by big businesses is particularly important to the success of the Economic Census. The top 1,000 companies—those with over 5,000 employees—account for about 30% of all business activity, and nearly 10% of establishments. For a detailed account of the efforts of the Census Bureau to make large companies aware of their importance to accurate measurement during the

Economic Census, see Robert A. Marske "Increasing Large Company Response to the Economic Census." Presented at the Census Advisory Committee of Professional Associations, October 26, 1995.

9. The enterprise consists of all business organizations under common ownership or control, as well as any affiliated firms in which the enterprise has the power to direct or cause the direction of the management and policies. Although controlling interest is usually defined as ownership of more than 50% of the outstanding voting stock, in some businesses (e.g., the banking industry, which considers 25% of stock ownership as a controlling interest) control may be exercised with a smaller percentage. In general, the Census Bureau accepts the company's listings and judgments of which companies it controls.
10. See Dodds, "Determining Economc Census Content," for several graphics that summarize the equivalent operations for determining Economic Census content. The graphics (and respective accompanying discussions) in both the Dodds article and the present article form a unified account of the Economic Census—from deciding which questions to ask, to the collection and processing of data, to the distribution of results in newly standardized table layouts and formats.
11. See Richard S. Sigman, "Development of a 'Plain Vanilla' System for Editing Economic Census Data," *Working Paper presented at October 1997 United Nations Work Session on Statistical Data Editing*.
12. The Census Bureau will publish the *Core Business Statistics Series*, debuting for 1997. This series is critical to the success of the 1997 Economic Census program. It includes the unprecedented release of data for all economic sectors right after the year in which the census is taken. This series also is important because it will showcase the very first data published on the new NAICS basis, and is key to bridging data between the old SIC system and NAICS. These "bridge" tables will present new data cross tabulated by both old and new classification systems at the same time, identifying the lowest common denominators between the two systems—SIC and NAICS. See Zeisset & Wallace, *"How NAICS Will Affect Data Users"* and Paul T. Zeisset, "Disseminating Economic Census Data," *Government Information Quarterly*, 15 (1998):305–320.
13. The Census Bureau has successfully created an improved data production and dissemination system for 1997. It is a system that will require fewer resources to create a more standardized data product which can be accessed quickly and easily by innovative and user friendly CD-ROM and Internet software. To provide some perspective regarding this accomplishment, a brief history of the Economic Census publication system processing follows.

    Prior to the 1977 Economic Census, all publication tables had been typed or generated by the impact printer, with all titles, column headings, lines, footnotes, symbols, and corrections added manually. Much of this process was literally a "cut and paste" operation. The experimental computerized, photocomposition publication program (the Table Image Processing System [TIPS]), developed for the 1977 Economic Census, was a major innovation to accelerate the production of publications. Nonetheless, this system still had major limitations that slowed the publication process. Therefore, a redesign effort was undertaken for the 1982 Economic Census. Addressing these limitations led to the development of TIPS II. The improved system included automated table composition, minimal manual processing of corrections to text and data, and greater flexibility and timeliness in the release of publications.

    For most past censuses, Economic Census printed reports were organized in separate series for industries, geographic areas, and special subjects. For some of these series, preliminary reports containing selected basic statistics were issued several months prior to publication of the final data. The final reports, which were issued after further review and analysis of the reported data, presented more detailed information.

    However, even the improvements associated with TIPS II resulted in relatively minor differences in the timing of data release from the 1977 and 1982 Economic Censuses. Later, with advent of CD-ROM technology—allowing random access to database information—electronic data dissemination played increasingly important roles in contributing to continuous improvement in the data release from the 1987 and 1992 Economic Censuses. Also—beginning with 1982 data, increasingly with 1987 data, and tapering off with 1992 data—the Census Bureau released data online via CENDATA. This was a database of the most current and widely used data products and was provided through cooperating private online vendors. Finally, with the advent of the Internet and World Wide Web, the Bureau phased out electronic data dissemination via CENDATA. In fact, most products from the 1992 Economic Census are now available on the Internet.
14. Zeisset, "Disseminating Economic Census Data."