

**Distr.
GENERAL**

**CES/SEM.50/5
15 April 2003**

Original: ENGLISH

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR
EUROPE**

**STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (Eurostat)**

**CONFERENCE OF EUROPEAN
STATISTICIANS**

**Joint UNECE/Eurostat Seminar
on Business Registers
(Luxembourg, 25-26 June 2003)**

**THE CENSUS BUREAU'S BUSINESS REGISTER:
BASIC FEATURES AND FUTURE DIRECTION**

Invited paper submitted by the Bureau of the Census of United States*

I. INTRODUCTION

1. The Census Bureau is in the process of developing a new Business Register (BR). It will be fully operational by the summer of 2003. Although the new BR will provide several advantages over its predecessor, many of the concepts remain the same.

2. This paper describes the Census Bureau's Business Register, formerly known as the Standard Statistical Establishment List (SSEL). The first section presents an overview of the Business Register, covering topics such as purpose, scope, statistical units, content, and data sources. The next section of the paper will identify and discuss the differences and advantages of the Census Bureau's new BR and the new computer environment, and the paper concludes by briefly describing the focus of future work.

* Prepared by Thomas L. Mesenbourg, Edward Walker and Paul Hanczaryk, Bureau of the Census of United States.

II. BR OVERVIEW

3. The BR is the Census Bureau's master business list. We compile and maintain the BR on a continuous basis. The BR contains basic economic information for over 7 million employer businesses and over 16.5 million non-employer businesses, i.e., businesses without any paid employees other than the owner/operator. Its scope spans all economic activity, with the exception of private households. For the agriculture and public administration sectors, the BR contains information for legal entities, but no information for individual establishments or locations.

Purpose

4. The BR serves as the primary sampling frame for the Economic Census, a program conducted every five years, and most current economic surveys. In this capacity, it identifies the set of statistical units that represents the target population, as defined by reference period, statistical unit, industry codes, geographic codes, business size, and other factors. Secondly, the BR serves as:

- A central repository for administrative record information provided by the Internal Revenue Service and other agencies for both employer and non-employer businesses.
- Support for collection and processing, including processing modules that maintain organizational structure and linkages for all enterprises, update industry and geographic codes, monitor coverage, and so on.
- The source of basic employment and payroll measures summarized by industry and geography in the County Business Patterns and ZIP Code Business Patterns statistical series.
- A research resource for special studies.
- The basis for longitudinal studies that track units through reorganizations or changes in ownership.

5. The BR provides a coherent methodological framework; complete, unduplicated, and up-to-date coverage; consistent industrial, geographic, and size classifications; accurate contact information; and effective support for data collection and processing operations.

BR Data Sources

6. The BR contains information provided both through direct data collections as well as administrative data obtained from other agencies.

7. Every 5 years, the Census Bureau conducts an Economic Census, collecting data for reference years ending in "2" and "7". In December 2002, we mailed Economic Census report forms to 5 million business locations with paid employees. Approximately 1.5 single

location businesses (single units) received a short classification form, and an additional two million single establishment enterprises received one of the 500 plus Economic Census forms. All enterprises that own and operate more than one establishment (multiunits) were mailed census report forms for each location. Some 180,000 enterprises owning and operating almost 1.6 million different locations received individual census forms. For the remaining 1.5 to 2 million small, single location enterprises not mailed in the Census, we relied on tax data in lieu of direct reporting. In summary, the Economic Census provides detailed information on about half of the total number of business locations with paid employees, but these establishments account for over 85% of total economic activity as measured by receipts. The Economic Census also plays a critical role in updating and maintaining the organizational linkages within enterprises; it provides current information for all multi-establishment enterprises as well as identifying those single location enterprises that have become “multiunits” since the previous Economic Census.

8. Between censuses, the Company Organization Survey (COS) is the primary vehicle for maintaining the organizational structure of the larger, multi-establishment enterprises. We survey approximately 50,000 of the largest multiunit enterprises annually, plus another 5,000 or so enterprises that have shown significant changes in their administrative record employment or payroll data compared to the previous year. The COS prelists all establishments owned by the enterprise along with their establishment identifier, name and address, and the appropriate North American Industry Classification System (NAICS) industry code. For each establishment, the enterprises provide March 12 pay period employment and payroll, as well as annual payroll. The enterprises are asked to report operational status for each location, report any new locations, closings, divestitures, or acquisitions. For divestitures and acquisitions, we ask the enterprise to report the name and address of the company or companies involved in the acquisition or divestiture. The employment and payroll data for each of the multiunit establishments is carried to the BR. Another source of organizational changes is our monthly, quarterly, and annual economic surveys.

9. In non Economic Census years, the administrative record information, in conjunction with the Company Organization Survey (COS), are the primary inputs to the BR. The Census Bureau makes extensive use of administrative records from the Internal Revenue Service (IRS), Social Security Administration (SSA), and the Bureau of Labor Statistics (BLS), to construct and maintain the BR. In particular, the IRS information provides first quarter employment and payroll, annual payroll, and receipts data for the millions of single location businesses on our register as well as the same data for the legal entities associated with all multi-establishment enterprises. These data offer an efficient, low-cost alternative to the direct mail canvass of small businesses.

10. The IRS data provide a degree of coverage and quality that would be difficult and too expensive to duplicate through other means. Further, the IRS data provide a valuable tool for editing and refining the reports of large businesses. The uses of business-related administrative record information increases the efficiency and quality of the Register data, while lowering the cost of Register operations.

BR Information Content

11. For each organizational business unit, the BR contains the current information on:
- BR identification number – A randomly assigned number that uniquely identifies the establishment; the BR ID is the survey unit ID for the Economic Census. The BR also contains links to enterprises, Employer Identification Numbers and other current survey IDs.
 - Employer identification number (EIN) – A IRS-assigned number, which Census links to the parent enterprise and the establishments they comprise.
 - Primary name of the firm – The legal name of the business.
 - Trade name of the firm.
 - Contact information – The primary reporting official’s name.
 - Mailing addresses – Includes the street, place, state, and 9-digit ZIP Code.
 - Physical location addresses – As with mailing address above, includes the street, place, state, and 9-digit ZIP Code.
 - Geographic codes – Automated geographic coding systems classify each establishment based on physical location address / for:
 - State
 - County
 - Place
 - North American Industry Classification System (NAICS) codes – The BR currently uses the 1997-based NAICS code. This census-based NAICS code is an 8-digit code that represents the 5-digit NAICS industry suffixed by a 1-digit U.S. industry extension and the Census Bureau’s 2-digit sub-industry extension. In addition, the BR carries type-of-operation codes that distinguish different categories of wholesale trade activities, and that distinguish various types of auxiliary (or support) activities.
 - Legal form of organization (LFO) code – Each statistical unit is classified as corporation, partnership, sole proprietorship, cooperative, government-owned or controlled, other (for example trust, estate, foundation, etc.).
 - Taxable/tax-exempt status.
 - First quarter employment – as of the pay period that includes March 12th.
 - First quarter payrolls – Consists of wages, salaries, bonuses, commissions, exercised stock options, taxable fringe benefits, and other compensation paid to employees.
 - Total annual payrolls – For single-establishment firms, it represents the sum of the four quarters of payroll. For multi-establishment firms, it represents annual payroll as reported by the respondents.
 - Value of IRS reported sales and receipts, or revenue – The specific content varies, depending on the income tax return the business/organization files.
 - Assets.
 - Interest income, gross rents, parent EIN, activity status, filing requirement codes, and much, much more.

Statistical Unit Definitions

12. The BR explicitly represents three basic types of statistical units – the establishment, EIN, and the enterprise. It also provides a link to a fourth unit – the Alternative Reporting Unit (ARU). Definitions for each unit provided below.

13. **Establishment**—An establishment is an economic unit, generally at a single physical location, where business is conducted or where services or industrial operations are performed. Examples include a mine, factory, warehouse, sales office, grocery store, bank, hotel, movie theater, doctor’s office, museum, and central administrative office. The establishment is the BR’s most basic statistical unit; it serves as the primary collection unit for the Economic Census and some economic surveys (for example, the Annual Survey of Manufactures) and as the building block for constructing more aggregated collection units used by other economic surveys. Further, the establishment is the fundamental unit to which industrial classifications apply.

14. **Alternative Reporting Units (ARU’s)** - An ARU is a consolidated unit made up of one or more establishments. ARU’s are established for the Economic Census for industries that do not maintain financial information for individual establishments. In mining, utilities, insurance, finance, and some information industries, establishments are aggregated by 5- or 6-digit NAICS code and companies report detailed financial information for these consolidated units while a separate form collects employment and payroll information for each establishment. Units similar to the ARU are frequently used in current economic surveys in the non goods producing sectors of the economy. The consolidated unit reduces business reporting but makes it difficult to publish subnational data. The new BR contains all Economic Census survey units, survey units for current surveys will be added in the future.

15. **EIN Entity**—An EIN (Employer Identification Number) entity is a legal entity set up by the IRS for tax reporting purposes. Each EIN has a unique identifier. In practice, this identifier is used more broadly by administrative records systems (for example, administrative records of the Federal/State Unemployment Insurance [UI] systems include the employer’s EIN). Therefore, its presence on the BR is critical for linking administrative records information to related establishment and enterprise units. The EIN entity also is used as a sampling/collection unit in some surveys (notably, annual and monthly surveys covering retail, wholesale, transportation, and a broad range of service industries).

16. **Enterprise**—An enterprise is an economic unit comprising one or more establishments under common ownership or control. On the BR, this unit represents the top-level U.S. parent company and all subsidiary U.S. companies of which the parent owns or controls a majority (more than 50 percent) interest plus all EINs and establishments affiliated with the parent and its subsidiaries.

17. The BR does not contain any information on foreign subsidiaries or their associated establishments.

18. A brief profile of the BR is shown in the table below.

Profile of the Business Register's Business Population

	Number of Active Enterprises	Number of Active EINs	Number of Active Establishments	Number of Employees (mid-March)
All Enterprises	22,182,499	N/A	23,600,003	114,064,976
Employers, total	5,652,544	5,779,683	7,070,048	114,064,976
Single-Unit	5,471,321	5,471,321	5,471,321	48,977,768
Multiunit	181,223	308,362	1,598,727	65,087,208
Non-employers	16,529,955	N/A	16,529,955	0

Source: County Business Patterns – 2000.
Non-employer Statistics – 2000.

Statistical Unit Relationships

19. A pictorial representation of single unit enterprises and a typical complex multi-establishment enterprise is provided in Figure 1 at the end of this paper. For single units the enterprise, EIN entity, and the establishment are one and the same. For complex multi-establishment enterprises the relationships can be extremely complex. The picture would be even more complicated if we overlaid current survey reporting units.

BR Maintenance Operations

20. Figure 2 at the end of this paper provides an overview of key BR maintenance operations.

21. As mentioned earlier, the BR places a heavy reliance on the administrative data obtained from the Internal Revenue Service. Information on almost 100 million records are processed annually. The Social Security Administration (SSA) provides information on all new businesses with paid employees. The SSA provides name, address, and NAICS classification information for all “births”. The IRS records sometimes have no or incomplete information about the principal business activity of the EIN. The Census Bureau provides the Bureau of Labor Statistics (BLS) quarterly with a file of EINs where the classification is incomplete. BLS matches these EINs against their BR and returns NAICS codes to Census for incorporation into the BR.

22. The payroll data reported to the IRS is usually complete and of high quality, but we may be missing first quarter employment for 12-15% of the employment/payroll records. For these cases, we impute employment based on prior payroll patterns. For enterprise with multiple establishments, we summarize by EIN the establishment employment and payroll data they report in the annual COS survey and compare it to the same information reported to

the IRS for each EIN. This “completeness” edit helps us identify possible changes in company organization and associated reporting problems.

III. CHANGES AND IMPROVEMENTS IN THE NEW BUSINESS REGISTER

23. The Standard Statistical Establishment List (SSEL) became operational in 1973. The SSEL’s remarkable longevity is a tribute to the soundness of its concepts and methods. However, as we approached the 2002 Economic Census it was clear that the SSEL was showing its age and a major redesign was needed.

24. Many of the SSEL’s processing subsystems were over 10-years old, and were difficult to maintain. Some of the old Digital Equipment Computer (DEC) software was no longer supported. The antiquated software and the complicated design significantly limited the Census Bureau’s ability to change and improve the Register. For example,

- There was no place to store new data content, such as the NAICS code and the additional administrative record data we now receive from IRS.
- It was difficult to support current surveys under the SSEL framework.
- We were restricted to three years of data, which hampered our analytical review capabilities.

25. The Census Bureau has developed several new features that will improve the functionality of our centralized Business Register:

1. **New Identification Numbering System** – We have developed a serialized number for establishment and enterprise units on the new BR. This new “survey-unit ID” will not contain any embedded ownership information, which was built into our previous SSEL numbering system. The new BR numbering system will improve our analyst’s ability to track establishments and enterprises over time.
2. **Additional Organizational Unit for Alternate Reporting Units (ARUs)** – We have updated the register’s conceptual framework to reflect current practices and trends in business organization. In particular, we have added a new organizational unit, the ARU. An ARU is a special consolidated unit established for the collection of multiunit enterprise data in certain networked industries, such as utilities, finance, and other areas where financial data are not available by establishment. ARUs will facilitate the collection of census and survey data for networked industries, while reducing respondent burden.
3. **More Data** – The Census Bureau will maintain many additional administrative data elements on the new Register, including Beginning and Ending Inventories, Expenses, Interest Income, Gross Rents, all components of quarterly wages (including Total Compensation, Medicare Wages, and Social Security Wages), Total Income, Company Contact information for single-establishment firms, and more. The amount of available administrative record source data has been increased dramatically. The expanded data will assist analysts in their review and reconciliation operations.

4. **More Years of Data** – We will maintain up to seven years of data on the new Register (the SSEL maintained only three years of data). This will enable our analysts to track data from the prior quinquennial economic census.
5. **Data for Employer and Non-employer Businesses** -- The new BR will carry data for all active businesses (about 24 million records), including employer and non-employer businesses. The SSEL only retained information for employer businesses. As a by-product, the Census Bureau will simplify and expedite the processing of the annual Non employer Statistics program. The steps to produce the non employer business universe has been built into our new BR processing, which will eliminate the need for external matching operations of huge non employer data files. We expect that these enhancements will reduce the non-employer processing time by over 3 months.
6. **Original and Corrected Data** – We will retain the original and corrected data for several key data fields, including the employment, payroll, and receipts information. We will retain the original data, the current “best” data after analytical updates, and the previous data for those cases where the original data has been changed. This will enable analysts to track a business’ reporting history, and determine if comparable changes should be invoked in the current processing cycle.
7. **Comparable Survey Data** – The new BR will maintain complete survey data from related economic programs, such as the Economic Census, Annual Survey of Manufactures, Company Organization Survey, and County Business Patterns. This will help to ensure comparability of all the Census Bureau’s economic data releases. The data from these economic programs will be easily accessible through the Directory Interactive Problem-Solving Environment.
8. **Additional Industry Detail** – The new BR will use the North American Industry Classification as our primary classification. The BR will maintain the 1997-based NAICS code initially, and will use the 2002-based NAICS code beginning in 2004. We have added fields to account for this revision, and will house both the old and the new NAICS code. This functionality is crucial in supporting all of our current programs, as they will convert to the new NAICS structure at slightly different time intervals.
9. **New NAICS Processing Structure** – We will improve the NAICS code assignment and maintenance, particularly from administrative sources. The Census Bureau will use the NAICS codes from the economic census or current survey collections, if it is available. If not, we use the NAICS code from administrative record sources. Under the new BR processing system, we have developed an enhanced methodology to select the “best” NAICS codes when conflicting codes are received from administrative records. We receive NAICS codes from three administrative record sources – the Bureau of Labor Statistics (BLS), the Social Security Administration (SSA), and the Internal Revenue Service (IRS).
10. **Enhanced Contact Information** – Contact information is critical in delivering collection instruments to potential economic census and survey respondents. We will maintain additional name/address information on the new Register, including the legal name and trade name of the business; the administrative record and respondent-based mailing address; company contact information for

both single- and multiunit companies; and physical location address from three sources (from census collections, from the IRS, and from the BLS). As a point of note, the additional physical location sources will help us resolve geographic cluster problems (i.e., multiple businesses at the same address, quite possibly the address of the accountant), a problem that has required considerable resources in past censuses.

11. **Edit and Impute Administrative Record Data More Timely** – As part of the design of the new Register, we will impute administrative record data when posting the data from source files. Under SSEL processing, we did not impute and edit these data until the end of the processing year. This will help to make the Register continuous rather than a series of iterations, which will enable us to create updated target populations for current surveys more quickly.
12. **Edit and Impute Multiunit Data More Timely** – The Census Bureau will edit and impute for missing multiunit establishments and data items earlier. At the time when the data are received, we will impute for missing multiunit data.
13. **Enhanced Quality Assurance (QA) Measures** – We are developing quality assurance measures that will show the characteristics of key data items at regular intervals. This “snapshot” of the register will highlight crucial information, including the distribution of the number of NAICS code from the various sources, the employment and payroll data by category (reported, imputed, corrected), and other measures. When the QA characteristics deviate from past periods, we will evaluate the differences, as these differences may result from processing errors.
14. **Expanded Use of Administrative Records** – The Census Bureau will expand the use of administrative record data on the new Register, to improve processing efficiency and reduce costs. As an example, we have incorporated vital linkage information that uses data from IRS Form 851, Affiliations Schedule. This administrative record source links parent corporation EINs to its affiliated subsidiary EINs. These corporate links will help us to identify the structure for new multiunit companies. In addition, the Census Bureau will use data from a new IRS file that provides merger, acquisition, divestiture, and other organization changes for legal entities over the past five years.
15. **Administrative Record Data in a Separate Table** – The Census Bureau will maintain all administrative record data in separate tables, or relations. It is critical to segregate IRS Title 26 data from Census-collected data. This meets a requirement of IRS Publication 1075, Safeguards for Protecting Federal Tax Information (FTI), which stipulates that FTI should not be commingled with other data where possible. Additionally, the new BR establishes a separate table for selected demographic data (race and gender information) that is needed for our Survey of Business Owners (SBO) program. Access to this table, called the SSN_NUMIDENT table, is restricted to SBO analysts.

26. The changes listed above should improve the quality, functionality, and productivity of the new Business Register. The following table compares the computer environment for the old SSEL and the new BR.

Comparison of SSEL with Business Register

	SSEL	Business Register
Platform	Compaq 8420	Compaq GS320 (file server)/GS320 (web server)
Operating System	Open VMS	Open VMS (file server)/Tru64 UNIX (web server)
Redundancy/Failover	2 Node Cluster	2 Node Cluster (file server)/2 Web servers
Disk Used	40 GB (shadowed)	600 GB (raided) - will grow to 1T8t
Memory Available	8 GB	24 GB (file server)/8 GB (web server)
# of CPUs	4 @ 625 MHz each	8 (file server)/2 (web server) - 1 GHz each
Database Software	Rdb	Oracle 8i, Developer, Designer, Discover, PL/SQL, SQL Plus
Other Software	SAS, Decforms, Cobol, ACMS	SAS, C, Cobol, Java, TOAD, HTML
Interactive Applications deployment	Dumb terminal session (using Decforms)	Web
Max # of Users Supported	1000 – 1200	1000 - 1200

IV. FUTURE DIRECTION

27. Even as we are in the process of finishing the initial implementation of the new BR, we have started thinking about future improvements. Over the next two years we will expend significant resources in incorporating and integrating the Census Bureau's 100+ plus monthly, quarterly, and annual surveys into the new BR. Undoubtedly, the integration of current surveys is going to require new functionality and tools to make the BR more accessible and useful across the Census Bureau.

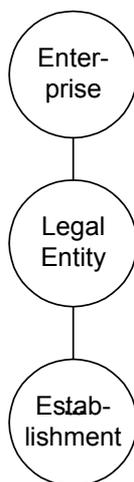
28. In 2002, the United States Congress passed legislation that would permit business information to be shared among three U.S. statistical agencies: the Census Bureau, the Bureau of Economic Analysis, and the Bureau of Labor Statistics. Shared data can only be used for statistical purposes. Data sharing provides numerous opportunities to reduce business reporting burden across the three agencies statistical programs while also improving data quality and content. As of April 2003, The Congress had not passed accompanying legislation that would permit tax data currently in the possession of any of the three agencies

to be shared with each other; the Census Bureau will not be able to share any of its data until the Treasury bill is passed. New functionality is going to have to be developed to make it easier to extract data from the BR and related data sets and as well identifying new opportunities to take advantage of BEA and BLS data.

29. Currently, the BR and most of our current economic surveys run on large computers, running either DEC VMS or UNIX operating systems. We are presently researching the feasibility of implementing a distributed network of commodity computers running in a Linux environment. Work to date has been encouraging and we believe this new environment can cut costs, improve performance, and increase computer redundancy.

Figure 1

Single Establishment Enterprise



Complex Multiestablishment Enterprise

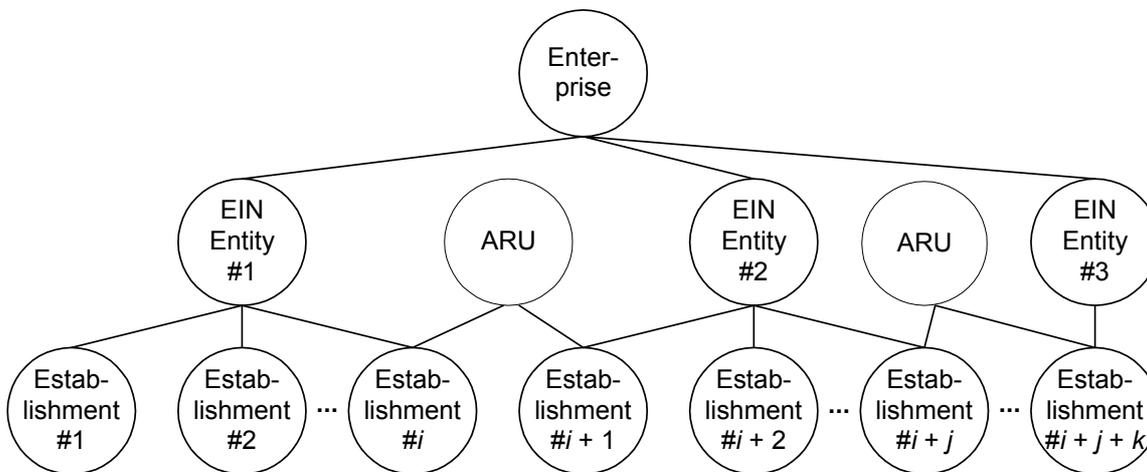


Figure 2

Business Register (BR) Maintenance Operations

Economic Directorate, non-census year

