25th Meeting of the Wiesbaden Group on Business Registers
- International Roundtable on Business Survey Frames

Tokyo, 8 – 11 November 2016

Brandy L. Yarbrough
U.S. Census Bureau
Session No. 2

Role of Business Registers

The Business Register as a Service Provider: Examples from the U.S. Census Bureau

Disclaimer: Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau

1. Introduction

The purpose of this paper is to illustrate how the U.S. Census Bureau leverages one of its most important corporate resources: the Business Register (BR). Although the primary mission of the BR is to offer survey methodologists a complete, unduplicated set of business entities from which survey sampling frames may be constructed, it also provides many other important functions. This paper will explore five key "services" that are BR-centric and highly important in terms of assuring that the Census Bureau's business statistics programs are administered in an efficient and cost-effective manner. Specifically, the following service-oriented roles of the BR will be discussed in the ensuing sections:

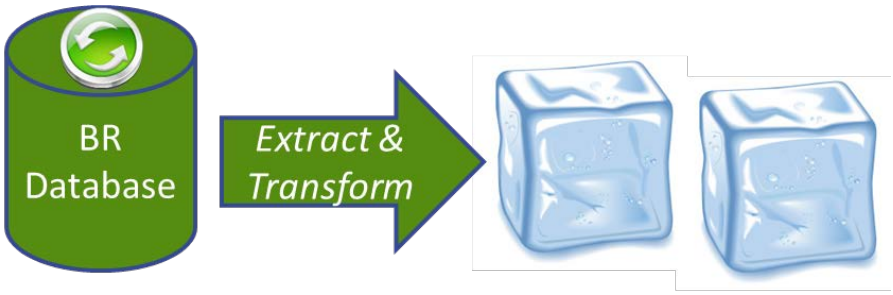
- *Sampling Support*
- *Operational Control System (OCS)*
- *Geographic Coding*
- *Company Reporting Calendar*
- *Workflow System*

The intent of this document is not to provide an exhaustive treatment on the technical implementation of these services—i.e., the *how*— but, rather, the focus will be on describing them in the context of statistical production and outlining the benefits that they provide. Toward this end, the *what*, *why*, and *when* of each service will be explained.

2. Sampling Support

What:
This service is intended primarily for the survey methodologist seeking to construct sampling frames but is also useful for any end-user that wishes to perform queries, analysis, or research on BR data.

As implied by the diagram below, the basic output of this service is a set of "frozen" rectangular data sets that are sourced directly from the BR database. In current practice, the BR is a dynamic, frequently updated Oracle database while the static data sets are delivered in a SAS format. SAS software, when properly configured, is capable of reading Oracle databases directly so the "extract and transform" process actually just takes the form of SAS code that is maintained centrally by the Register Operations staff. This staff ensures that any content or structural changes that are made to the Oracle database get properly reflected in the data sets as well.



Why:
There are a number of motivating factors for creating the data sets. In particular, the BR database itself is not necessarily structured for the types of analyses that are often demanded by end-users. Rather, the BR has been optimized for high-volume transaction processing and "normalized" in order to reduce data redundancies and minimize empty column values. Unfortunately, this emphasis on operational efficiency is sometimes at odds with analytical facilitation. For instance, in order to develop even the most basic record for a single-establishment company from the BR database one would need to navigate through at least four tables and have a good understanding of its data organization principles. Below is a simplified illustration showing data for a fictitious company.

BR Database

Parent table

ID	NAME	NAICS	LFO
2012345678	ABC CORP	211111	C

Data table

ID	PERIOD	EMP	PAY	RCPT
2012345678	2015A1	100	4000	50000
2012345678	2014A1	95	3900	48000

LINKS table

LINK_NAME	LINK_ID1	LINK_ID2	ACTV
EMPADR7	2012345678	987654321	Y
EMPADR8	2012345678	912345678	Y

Address table

ADR_ID	ADR_TYP	STREET	CITY	STATE	ZIP
987654321	MAIL	P.O. Box 10	WASHINGTON	DC	20233
912345678	PHYSICAL	123 MAIN	NORTH BEACH	MD	20714

Data Set

ID	2012345678
NAME	ABC CORP
NAICS	211111
LFO	C
EMP_CY	100
EMP_PY	95
PAY_CY	4000
PAY_PY	3900
RCPT_CY	50000
RCPT_PY	48000
STREET_MAIL	P.O. Box 10
CITY_MAIL	WASHINGTON
STATE_MAIL	DC
ZIP_MAIL	20233
STREET_PHY	123 MAIN
CITY_PHY	NORTH BEACH
STATE_PHY	MD
ZIP_PHY	20714

In order to construct the consolidated data set as shown on the right a user would need to understand the intricacies of BR database organization as displayed on the left. For example, to extract the mailing and physical addresses belonging to the BR unit one would first need to access the LINKS table in order to obtain the ID of each address record. The IDs would then be matched to the table containing the address elements and then correlated to the BR unit. This process is further complicated when the database record for the BR unit contains historic (inactive) addresses. Further, the measure-of-size variables for the current and previous reference periods are stored in the BR database as separate rows distinguished by specific reference period values—"2015A1" for the current period and "2014A1" for the prior period. When conducting comparative analysis it is obviously better to have these data in the same record (or, to use SAS terminology, *observation*). The sampling support data set creation service effectively resolves all of these sorts of complications for the end-user—i.e., they do not have to become BR database experts and can instead focus on meeting their core objectives.

The BR database is subject to near-continuous update from a variety of different sources including administrative records, survey responses, and interactive updates made by BR analysts. This dynamic, always-changing state is not always the best environment from which to construct sampling frames or conduct analysis. As such, the data set creation service provides frozen snapshots of the BR units at particular points in time which provides end-users with some stability

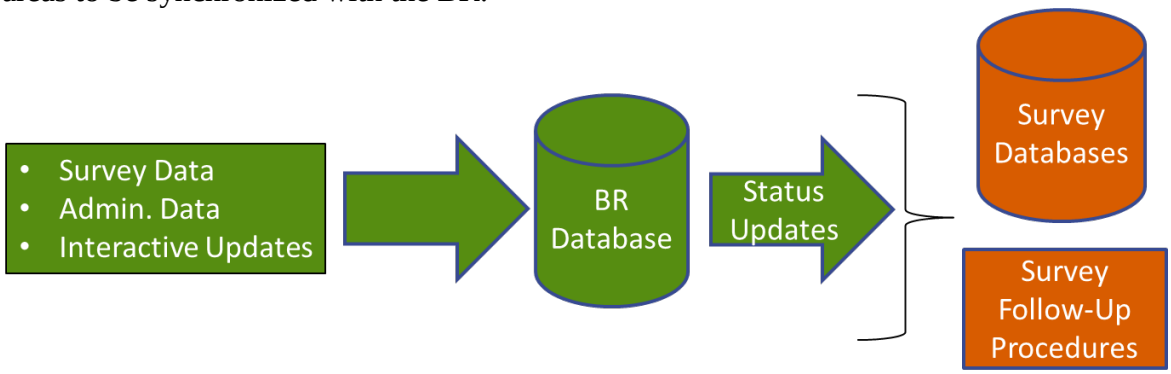
Another benefit of end-user data set creation is that it helps reduce the burden on the BR database and ensures that operational performance is maximized. If all consumers of BR data were to use the database directly it is likely that the performance of many production systems would be compromised. For example, database updates being made interactively by BR analysts might take longer to complete and large-scale batch operations might not finish in a timely manner.

When:
The data set creation service is performed monthly in conjunction with the completion of the administrative records posting cycle. This allows survey areas to augment or supplement their samples with births, deaths, and industrial classification changes for single-establishment companies (which are updated primarily from administrative records). The service is also performed annually in conjunction with the conclusion of the annual Report of Organization which is the primary instrument for updating multi-establishment companies.

3. Operation Control System (OCS)

What:
This service involves dynamic communication between the BR and survey data collection systems. In some respects, it is the opposite of the "frozen" data set concept in that the goal is to provide the survey operational infrastructure with near real-time status updates of BR units. In the most basic sense, this is done by sending "signals" from the BR to the survey databases and data collection processes. As indicated by the diagram below, these signals can originate from virtually any BR data input source including the posting of administrative records, the ingestion

of survey data, and updates made by BR analysts. In effect, the OCS service allows the survey areas to be synchronized with the BR.



Some examples:

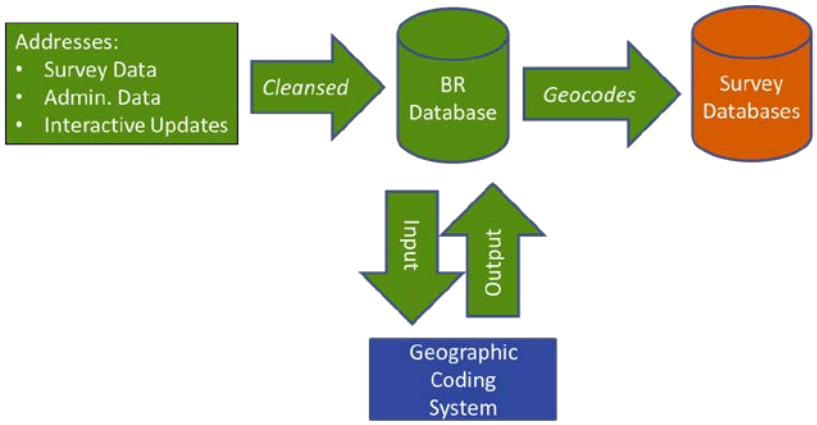
- Posting of payroll and income tax data of the BR that are within the survey reference period would result in a signal to retrieve this additional information. Similarly, an industrial classification update received from administrative records and applied to the BR would also generate a signal.
- An analyst adding a new location (establishment) to the BR that is in-scope to a survey might send a signal for this case to be extracted and loaded.
- If the tax-paying entity (EIN) of a single unit (SU) company becomes linked to an existing multi-unit company (MU) then the SU will effectively be removed from survey follow-up and reminder notifications. (In effect, the SU was an unlinked "duplicate" of the MU company). The SU will also be removed from survey tabulations. This type of linking and affiliation work is carried out by BR analysts and the OCS service automatically takes care of notifying the survey areas of such changes in status.

Why:
The primary rationale for the OCS service is that "frame maintenance" is often an integral part of survey data collection. For example, in the Annual Survey of Manufactures (ASM) single establishment companies are asked to report the number of locations in which they have operations. If more than one location is reported, the company is asked to provide some basic information about each one. These response data can be used by a BR analyst to convert the SU into an MU and effectively add establishments to the BR. These types of activities are happening on the BR concurrently with the processing and analysis of the ASM by survey specialists. The OCS allows "frame maintenance" updates made by BR analysts to flow into the ASM database as they occur. Such synchronization promotes efficiency by reducing survey costs, improving data quality, and also by helping to reduce response burden. This latter benefit is typically borne out as affiliations between unlinked units become known—e.g., an SU that is really part of an existing MU as explained above.

When:
In practice, the OCS is embedded in all BR updating processes so it runs continuously throughout the survey reference period.

4. Geographic Coding

What:
This service involves assigning tabulation codes, spatial coordinates, and quality indicators to BR unit addresses. The outputs of this service are used to tabulate data, perform geographic analyses, and create maps. As shown in the diagram below, addresses can originate from many different sources. When addresses are applied to the BR, they are cleaned, edited, standardized, and then sent into the Geographic Coding System[1]. This system parses the input address—street, city, state, and postal code (ZIP)-- and matches it to various reference files before returning a set of codes, coordinates, and flags that are collectively referred to as *geocodes*. The geocodes are stored in the BR database along with their associated address elements and ultimately passed on the survey areas (through the aforementioned OCS).



For reference, provided below is a simplified example of an address that has been geocoded:

INPUTS	
STREET	9132 C BAY AVE
CITY	NORTH BEACH
STATE	MD
ZIP	20714

OUTPUTS		
Tabulation Codes	**Coordinates**	**Quality Indicators**
ST= 24	X = -76.5314	LEVEL = 8 (*Block-level*)
COU= 009	Y= 38.70786	MATCH=M (*Exact*)
PLACE= 98009		
CSA= 548		
TRACT=860401		
BLOCK= 1053		
CD= 05		

[1] This system is based on the Census Bureau's MAF/TIGER database. "MAF" is the Master Address File and "TIGER" is the Topologically Integrated Geographic Encoding and Referencing system. The MAF/TIGER database is created and maintained by the Census Bureau's Geography Division. The author of this paper is certainly no expert on this topic and suggests that those interested in learning more about it consult this website: https://www.census.gov/geo/reference/

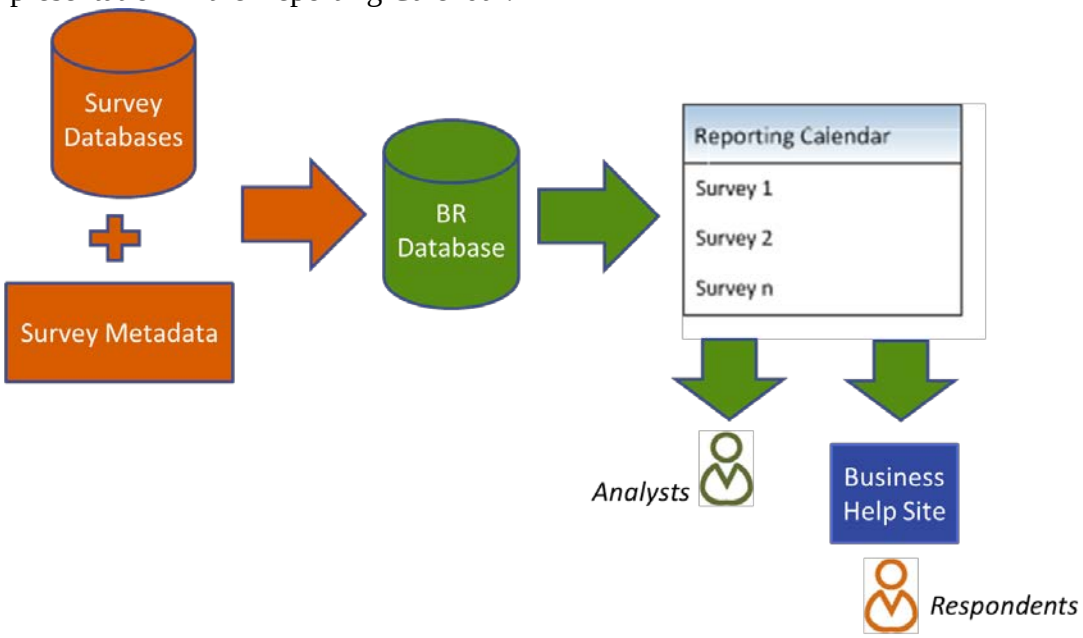
Why:
The benefits of having a centralized geographic coding service are efficiency and standardization across statistical programs. Using the BR as a platform to implement this service makes sense in that the database already contains the address elements of the units and provides a mechanism for delivering the output to the survey areas via the OCS. In essence, area tabulation codes for any statistical unit must be derived from the address that resides on the BR. The cleansing and standardizing of these addresses and the centralized coding services helps ensure that area-based tabulations are consistent and coherent. This is a particularly important in the Economic Census which produces tabulations with many geographic levels.

When:
BR addresses from all sources are processed through the Geographic Coding System each night.

5. Company Reporting Calendar

What:
This intent of this service is to provide both respondents and internal analysts with a comprehensive status of all business surveys in which a given company participates as a respondent. As the diagram below illustrates, the BR serves as a common platform for ingesting reporting unit status updates from survey databases as well as survey metadata and then producing a "Reporting Calendar". (The term "calendar" is used because one of the primary uses is to provide respondents with a list of all filing due dates for their surveys).

A key role of the BR is taking reporting units from disparate surveys, some of which use non-standard or proprietary identifiers, and matching them to the appropriate enterprise for grouping and presentation in the Reporting Calendar.



The calendar itself is essentially a two-part report. The first part of the calendar essentially provides key information about each survey—i.e., survey metadata. Included in this section are things like: name, purpose, and identifier of the survey, reporting media that are available, reporting unit type, mailing frequency, due dates, reporting burden hours, and the primary contact for the survey at the Census Bureau. The second part lists specific details for each reporting unit belonging to the company, including: reporting unit identifier, survey identifier, collection instrument (form number), business name, mailing address, company contact data, current filing status, and medium used for filing.

An example of a calendar for a fictitious company is provided below. This company is participating in four surveys and has six different reporting units.

Disclosure Prohibited Title 13/26 U.S.C. - Census Confidential

08/04/2016 17:48:15 **Company Reporting Calendar Surveys**

ID: 2134567890 NAME1: ABC INC

Form Number	Total Reporting Units	Survey Identifier	Survey Title	Form Description	Media Available	Reporting Unit Type	Mailing Frequency	Mailout Schedule	Due Date Schedule	Census Contact
ACE-1(M)	1	ACES	ANNUAL CAPITAL EXPENDITURES SURVEY	Collects capital expenditures from employer businesses for new and used structures and equipment.	Electronic,Fax,Paper	MU Enterprise	Annual	MARCH	30 DAYS	IMA KNERD 3017632222 IMA.KNERD@CENSUS.GOV
SA-44S	1	ARTS	ANNUAL RETAIL TRADE SURVEY	Collects detailed industry measures of retail company activities	Electronic,Paper	Part of an Enterprise	Annual	MARCH	APRIL	HOLLY WOOD 3017634444 HOLLY.WOOD@CENSUS.GOV
NC-99001	1	COS	REPORT OF ORGANIZATION	Updates company organization for multi-location businesses on Census Bureau's Business Register; Collects operating data by establishment which are used in annual County Business Patterns reports	Electronic,Paper	MU Enterprise	Annual	DECEMBER	FEBRUARY 12	NOAH BAWDY 3017635555 NOAH.BAWDY@CENSUS.GOV
MEPS-10	2	MEPS	HEALTH INS COST STUDY-ESTABLISHMENT	Collects data on enrollment in and characteristics of health insurance plans offered by employers for all businesses	CATI,Electronic,Paper	Establishment	Annual	JUNE 14	40 DAYS	LEE MALONE 3017631111 LEE.MALONE@CENSUS.GOV

Disclosure Prohibited Title 13/26 U.S.C. - Census Confidential

08/04/2016 17:48:15 **Company Reporting Calendar Reporting Units**

ID: 2134567890 NAME1: ABC INC

Reporting ID	Survey Identifier	Form Number	Name1	Name2	ATTN	Street	City	State	ZIP	Reference Year	NAICS Code	Mailout Date	Reporting Medium	Check in Date	Census Form Count	Company Contact	Census Contact
12345600010	ACES	ACE-1(M)	ABC INC	ATTN JEAN BLEU	JEAN BLEU	46 MOLSON GOLDEN HWY	GOLDEN VALLEY	MN	55422	2015	459900	18-MAR-16	Electronic	25-APR-16		JEAN BLEU 1112224444999 JBLEU@ABC.COM	
00012345600	ARTS	SA-44S	ABC INC	ATTN: DOUG DURT		46 MOLSON GOLDEN HWY	GOLDEN VALLEY	MN	55422	2015	442110	18-MAR-16	Electronic	22-APR-16		DOUG DURT BEAN COUNTER 11122233330000 DDURT@ABC.COM	
8111111111	COS	NC-99001	ABC INC		DOUG DURT	46 MOLSON GOLDEN HWY	GOLDEN VALLEY	MN	554224822	2015		21-JAN-16	Electronic	26-APR-16		DOUG DURT STAFF ACCOUNTANT 1112223333 DDURT@ABC.COM	
00012345600	MART15	SM7212S-A	ABC INC	DOUG DURT	ATTN DOUG DURT	46 MOLSON GOLDEN HWY	GOLDEN VALLEY	MN	55422	2016	442110					FINANCE CENSUS BUREAU REPORTING 11122233330000	
013456789000001	MEPS	MEPS-10	ABC INC			7894 111TH ST SW	WASHINGTON	DC	200090000	2016	442110	21-JUL-16	CATI		1	JESSICA RABBIT WELLNESS COORDINATOR 7654321000	LEE MALONE 3017631111 lee.malone@census.gov
013456789000002	MEPS	MEPS-10	ABC INC	DELIVERY CENTER		7799 BLUEBERRY HILL DR STE 0	ELKRIDGE	MD	210750000	2016	493110	21-JUL-16	CATI		1	JESSICA RABBIT WELLNESS MANAGER 7654321000	LEE MALONE 3017631111 lee.malone@census.gov

For respondents, the Reporting Calendar is delivered through the Census Bureau's Business Help Site (BHS). The BHS essentially functions as a secure, internet-based respondent portal for business surveys. Respondents log on to the BHS via a user ID and password and, among other things, are able to download their Reporting Calendars.

Why:

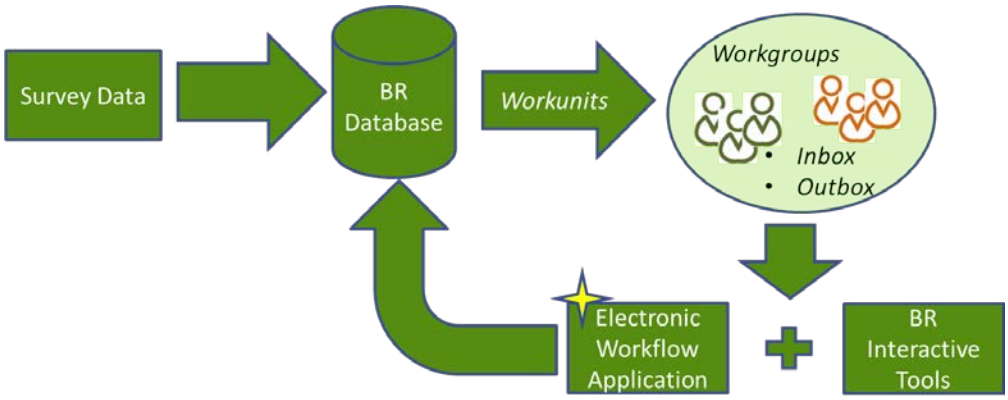
Currently, not all of the Census Bureau's business survey programs utilize standard BR units or identifiers. Further, not all surveys are conducted and administered using the same processing system. This, of course, makes it difficult to provide a consolidated point of reference for a given company. While efforts are being made to standardize reporting units and better manage

reporting burden across surveys, the Reporting Calendar offers an interim solution by providing a company with a comprehensive look at its filing status and obligations. In addition, it facilitates communication between business survey analysts and respondents by providing each party with the full picture of reporting burden, filing status, and due dates. For example, an analyst may be more inclined to accept later response for their particular survey if the company in question happens to be involved in many other surveys and are making relatively good progress on them. The Reporting Calendar is also a useful tool for the Census Bureau's Respondent Advocates who are charged with fielding complaints from companies that feel they are being over-burdened or unfairly included in too many surveys. The Reporting Calendar can be used to quickly confirm or refute such claims.

When:
The data that feeds the Reporting Calendar are updated in the BR on a weekly basis.

6. Workflow System

What:
This service takes the form of an electronic case management system that is used by analysts to identify and track BR referrals that originate with survey responses. When survey response data are ingested by the BR system, they are subject to a series of tests that are intended to identify conditions that may entail an update to the BR database. Cases meeting these conditions are flagged and delivered—i.e., *referred*—to BR analysts for research and resolution. A common example of a BR referral would be the identification of a newly reported establishment of a multi-unit company. The BR analyst would confirm that the location is not already on the BR database (perhaps under a different parent enterprise), validate the reported data, and then, if appropriate, officially add it. There are many other examples of referrals that stem from structural changes in a company or updates to characteristics of its establishments[2]. As illustrated by the diagram below, the basic purpose of the electronic workflow system is to efficiently group referrals into *workunits* and deliver them to *workgroups*.



[2] The annual Report of Organization is a Register Profiling survey that is designed to capture organizational changes in the largest BR enterprises. This survey, as well as the five-year Economic Census, typically results in several thousand BR updates and referrals to BR analysts.

A *workunit* is essentially a BR enterprise and all of its referrals while a *workgroup* is a collection of BR analysts. Workgroups are distinguished by the types of workunits that they receive. For example, workunits belonging to the largest and most complex enterprises would flow to a workgroup comprised of the most knowledgeable and experienced BR analysts.

The central component of the system is the Electronic Workflow Application (EWA) that provides interactive features and functionality which allows BR analysts in workgroups to review, document, transfer, and resolve referrals associated with their assigned workunits. Each analyst has their own virtual "inbox" and "outbox" that contains, respectively, workunits that have been assigned and workunits that have been completed. As such, progress and backlogs can be assessed in real time and, if necessary, supervisors can re-balance the workloads of individual analysts. The EWA is integrated with existing interactive tools for updating and maintaining the BR so it is effectively an extension of a system with which analysts are already very familiar.

Why:
The impetus for an electronic workflow system grew out of the need to become more efficient and cost-effective in the management and resolution of BR referrals. For many years, BR workflow was a paper-based system with physical hardcopy listings being printed and distributed to workgroups. In a typical year in which the Report of Organization survey is conducted, as many as 500,000 sheets of paper could be produced by this system. During the Economic Census, more than a million pages of output could be generated. The burden of tracking and managing of all of this paper fell to the workgroup supervisors and analysts who would spend a significant amount of time simply trying to account for their workloads. Further, given the enormous volume of physical paper and the need to often route it to multiple workgroups some of it would simply get "lost in the shuffle". In this type of environment it is also virtually impossible for managers to get a true status on progress and backlogs.

Also, in recent years at the Census Bureau, there has been an over-arching strategic plan to replace paper-driven processes with electronic-based solutions. This has been most prevalent in the area of survey data collection where paper questionnaires have been de-emphasized or completely replaced by web-based instruments. However, there are numerous other areas of the survey life cycle where this strategy can prove beneficial and electronic workflow is most definitely one of them.

Unlike the other BR-based services referenced in this paper, the electronic workflow system is currently under development.

When:
The electronic workflow system will run continually throughout the survey reference period. It will be deployed in production for the first time starting with the 2016 Report of Organization and is expected to pay great dividends throughout the 2017 Economic Census.

7. Conclusion

As this paper hopefully illustrates, the Business Register serves many valuable roles beyond simply being a "business list". It is, in fact, a fulcrum for establishing effective and efficient survey operations for the Census Bureau's business statistics programs. This central role that the BR plays is only expected to grow and expand even as the survey environment rapidly changes due to both internal and external factors. The emerging world of "Big Data" portends the development of additional services that use the BR as a platform for cleansing, matching, standardization, and delivery. Internally, as the Census Bureau looks toward corporate or enterprise-wide solutions for survey operations, the BR will be expected to offer services that interface and communicate with these new systems. Further still, there are thoughts of merging or incorporating some survey-specific databases directly into the BR. This, of course, would entail the BR absorbing the functionality of these databases and their underlying systems. Regardless of the future direction of the Census Bureau's business survey programs, the BR will continue to play many important roles and be an indispensable asset.