

The Longitudinal Employer-Household Dynamics Program at the U.S. Census Bureau

Ronald C. Prevost, John M. Abowd, and Julia I. Lane
Longitudinal Employer-Household Dynamics Program
U.S. Census Bureau, Room 2138/3
Washington, D.C. 20233-8400

Ronald.C.Prevost@census.gov, John.Abowd@cornell.edu, JLane@ui.urban.org

1. Introduction

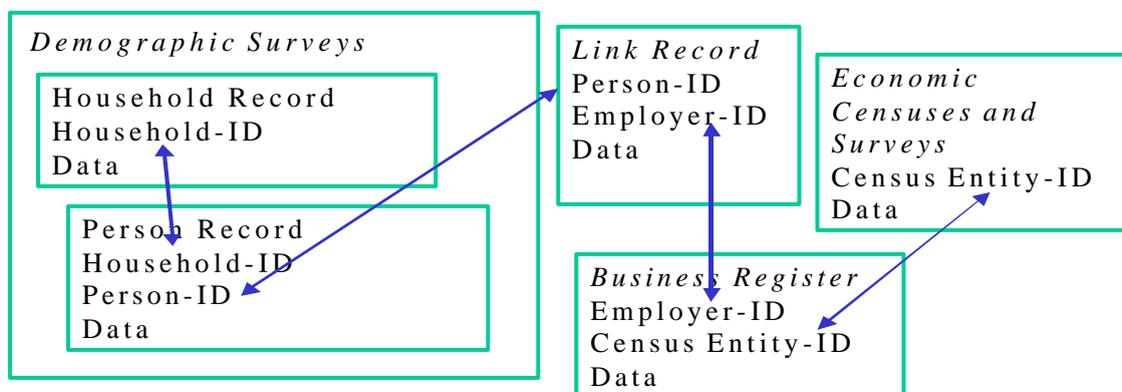
A number of technological breakthroughs have made it possible for statistical agencies to generate timely statistics at lower cost, with reduced respondent burden via the integration of administrative records, survey, and census data. First, profound advances in computing technology and record linkage accuracy have significantly increased the feasibility of expanded uses of administrative records for statistical purposes that avoids repetitive and burdensome inquiries of the public. In addition, the spiraling costs of, and public resistance to, traditional data collection have increased the opportunity for significant benefits through using administrative records in data collection, estimation, and evaluation systems [P1996], [PL1999]. Of course, this approach poses new challenges to statistical agencies. In particular, the management of activities that blend respondent information from multiple sources must weigh any, and all, privacy and confidentiality concerns against the potential benefits of such research. Statistical agencies that integrate administrative and survey data must also have a corporate approach to the development and application of policies ensuring that the privacy and confidentiality concerns of the public and their data providers are met. The Census Bureau is currently exploring expanded use of administrative records in a number of research activities designed to produce results that, if fruitful, can be used to implement major program activities.

2. Merging Business and Person Data

The Longitudinal Employer-Household Dynamics Program (LEHD) is the Census Bureau's latest effort to combine administrative and survey data. The mission of LEHD is to evaluate and improve the quality and understanding of data collected by the Census Bureau's core demographic and economic censuses and surveys through longitudinal analysis. LEHD accomplishes its mission by combining federal and state administrative data on employers and employees with core Census Bureau censuses and surveys. LEHD conducts policy-relevant research on labor force and employment issues, suggests improvements to the design and processing of census and survey instruments and frames, and creates value-added data products.

The LEHD Program represents a substantial investment made by the Census Bureau to permit direct linking of its demographic surveys (household-based instruments) with its economic censuses and surveys (establishment and business unit-based surveys). The LEHD uses state and federal administrative records as the mechanism to identify the link between employers and employees. The main data products included in our infrastructure are integrated versions of data protected by Title 13 of the United States Code, from the Current Population Surveys, Surveys of Income and Program Participation, Surveys of Program Dynamics, American Community Surveys, the Business Register, and Economic Censuses and Surveys.

Figure 1. The Longitudinal Employer-Household Infrastructure



The direct benefits to the Census Bureau, and the potential benefits to social science research are substantial. Generically, the structure of the LEHD data base consists of two types of links: (1) internal (not public use) versions of core Census demographic surveys that have been joined with information on

the employer obtained from sources in the Bureau's economic censuses and surveys; and (2) variables created for the economic data bases from linked information on the workforce characteristics.

3. Managerial Controls

LEHD's approach to data management and control is to provide:

- A secure area where combined administrative and survey data on businesses and people can reside. LEHD has large, secure computer facilities; secure physical locations where research can be conducted with the limitation that its use is solely for statistical purposes; and assures the protections mandated by the laws governing both administrative and survey data sets [IRS1999]. Commingled administrative data, and Census Bureau survey and census respondent data are subject to the confidentiality provisions of all sources.
- The opportunity to explore mutually beneficial research and new data products for the Census Bureau and its research partners. LEHD possesses a number of advantages: a specialized staff that is experienced in data exploration, standardization, editing, and matching techniques; the existence of a core data infrastructure; the capacity to append new administrative and survey data sets required by our staff and research partners; and the provision of controlled access as part of the Census Bureau's data dissemination efforts.
- A process to facilitate the research activities conducted at the Census Bureau. All analyses of the LEHD micro data will be conducted using the Title 13 secure facility that the Bureau has provided for LEHD. Only a handful of Census Bureau employees will have access to the data files that contain the personal identifiers required to do the most sensitive part of the data linking [GC1999]. Specially processed research versions of these files, which contain no personal identifiers, will be used on the LEHD computer system. Researchers working on approved projects will be provided with extracts from the LEHD data containing only the variables and cases required for their analyses.
- A process to facilitate external research. A proposed example of this process would include the creation of synthetic micro-record databases that possess the qualities of real data but in no way identify individual entities. These databases could be used by researchers to develop programs to produce or simulate the statistics they require. Researchers would then submit their final programs to LEHD staff who would run production programs on the "real" data. LEHD would disclosure-proof the output and return the final product to the researcher.

The management of statistical information systems containing commingled survey and administrative data requires a multitude of control points. We must assure administrative data providers that we possess both the legal authority to acquire their data, and the administrative constructs and operations to assure the privacy and confidentiality of that data after receipt. These constructs are manifested in computer and physical security plans, practices, and reviews; employee oaths promising confidentiality; certification that employees understand the proper procedures for handling sensitive data; public notification for the needs and uses of information; completion of disclosure reviews for all releasable data products; and the maintenance of metadata libraries describing the contents and disposition of data files [HG1999].

We must also assure data providers that we will use their data for appropriate activities. This assurance is manifested in a project review, approval, and tracking process that requires the coordination of policies and operations throughout a statistical agency [S2000]. We must be prepared at any time to generate reports on the disposition of data files, the status of privacy and confidentiality controls, and the status of projects that use administrative data. The project review process includes the following criteria for evaluation of administrative projects: scientific merit; benefit to the Census Bureau; disclosure risk; conflict of interest; resource requirements; feasibility of project; compliance with interagency data agreements; public perception; and compliance with security standards and policies. In an agency the size of the Census Bureau, there are multitudes of projects at some stage of proposal or completion and each new project possesses the potential to set new policy precedents. Therefore, project reviewers work closely with the Census Bureau's Data Stewardship Executive Policy Committee to assure that all projects are consistent with the Census Bureau's Strategic Plan and that all individuals accessing data adhere to good stewardship practices.

4. Our Partnership and Its Potential

Governments faced with rapidly changing economies and societies often make far-reaching economic and social policy decisions using information derived from Census survey data. Understanding the increasing complexity of our nation's economy requires more information than survey data alone can capture. The LEHD Program, in its pilot phase, has formed partnerships with other governmental agencies such as the Internal Revenue Service, the Social Security Administration, the Department of Labor, and State Employment Security Agencies to unite information from disparate sources in solving the policy-relevant data needs of an ever-changing nation. This partnership produces high-quality data products, cheaply and efficiently, while maintaining stringent confidentiality protections and reducing respondent burden.

If Congress, the Department of Labor, and State Employment Security Agencies are interested in the dynamic interactions of workers, businesses, government, and society when implementing the Workforce Investment Act, ; the LEHD can meet these needs. If Health and Human Services and state agencies are interested in the policy implications of Welfare to Work, ; the LEHD can meet these needs. If the Social Security Administration and the National Institute on Aging are interested in individuals' retirement decision-making process and the impact it will have on Social Security benefits and the welfare of older citizens, the LEHD can meet these needs.

Although the LEHD program has only been in operation for two years, it has attracted widespread internal and external support. In addition to Census Bureau funding, LEHD received one of the few National Science Foundation Social Data Infrastructure grants. Because of its unique ability to examine two sides of the labor market, the National Institute on Aging has funded the LEHD development to enhance understanding of the demand for older workers. In addition, the Sloan Foundation has funded LEHD as part of its Federal Statistics program, the Bureau of Labor Statistics has supported the development of state- specific databases, and the Department of Labor (ETA) has funded policy-relevant research on labor force and employment issues.

5. Specific Products

LEHD is creating its first set of integrated data, using state records from California, Florida, Illinois, Texas and Maryland. We are negotiating agreements with Ohio, Pennsylvania, North Carolina and Minnesota, and hope to add the balance of the 50 states, the District of Columbia, Puerto Rico, and the Virgin Islands in the next five years. The LEHD program's preliminary work has used demographic survey and administrative data to: develop measures of worker quality for firms; understand changes in firm ownership; examined the demand for skill by different types of firms and examine the reasons for differences in employment totals from household and establishment based surveys.

LEHD is creating its second set of integrated data using federal tax data to unite Census household-based surveys to Census business data sets. The federal data sets have enabled the LEHD Program to begin work on place of work and industry coding in the ACS; non-response bias in CPS and SIPP/SPD; correcting work histories in the 1990-1993 SIPPs; examining earnings measurement error in the SIPP, CPS, and ACS; and the development of SIPP Public Use Files expanded with lifetime earnings histories. In addition, the LEHD program is examining the potential to use the health and pension benefit information that is reported by firms to measure the availability of these plans to workers and improve estimates of health and pension reporting on household surveys.

Specific Program Goals

FY2002

- Finalize production methodologies and operations to create quarterly worker and employment dynamics for states, counties and industries.

FY2003

- Production and dissemination of quarterly worker and employment dynamics for states, counties and industries for five states.
- Development of methods and procedures for external researchers to understand our data infrastructure, production and disclosure-avoidance techniques and products.

FY2004

- Completion of new or revised methods for improving place of work coding operations.
- Creation of expanded historical SIPP public use microdata products to include lifetime measures of income and poverty useful for the estimation of potential Social Security benefits.
- Production and dissemination of quarterly worker and employment dynamics for states, counties

and industries for nine states.

FY2005 and beyond

- Completed analysis of worker disability and access to healthcare and pension plans.
- Production and dissemination of quarterly worker and employment dynamics for states, counties and industries for fifteen states in FY2005 and expansion to a national program.
- Development of synthetic data products and a process for external researchers to submit programs that rapidly produce disclosure-proofed results.

REFERENCES

- [GC1999] Gates, Gerry and Cynthia Clark (1999), "Policy for Access and Use of Administrative Records", U.S. Census Bureau Internal Document, June 1999.
- [HG1999] Highsmith, Samuel N., and Daniel W. Gillman (2000), "Building a Corporate Metadata Repository at the U.S. Bureau of the Census", ISIS-2000, Riga, Latvia, May 29-31, 2000.
- [IRS1999] United States Internal Revenue Service (1999), "Tax Information Security Guidelines for Federal, State, and Local Agencies: Safeguards for Protecting Federal Tax Returns and Return Information", Publication 1075, 1999.
- [P1996] Prevost, Ronald (1996), "Administrative Records in Official Statistics - How Can Our Research on Administrative Records in Census 2000 Lead to Greater Use in 2010", Council of Professional Associations on Federal Statistics, November 1996.
- [PL1999] Prevost, Ronald and Charlene Leggieri (1999), "Expansion of Administrative Records Uses at the Census Bureau: A Long-Range Research Plan", Federal Committee on Statistical Methodology, November 1999.
- [S2000] Streitwieser, Mary (2000), " Approval Process for Projects Using Administrative Records", U.S. Census Bureau Internal Document, April 2000.

RÉSUMÉ

Le Programme des Dynamiques Longitudinales des Ménages et des Employeurs (Longitudinal Employer-Household Dynamics Program, LEHD) est un nouveau programme du Bureau du Recensement des États-Unis (U.S. Census Bureau) dont le but est d'apparier les données provenant de produits démographiques et économiques. Le LEHD représente une nouvelle approche pour (a) améliorer la qualité des enquêtes présentement disponibles, (b) rajouter des informations supplémentaires à l'inventaire des connaissances sociales et économiques sans l'imposition d'un fardeau additionnel aux répondants des enquêtes, et (c) améliorer la qualité des données longitudinales. Ayant recours à des techniques d'appariement, des bases de données sont produites qui sont longitudinales dans deux dimensions : celle des ménages et des individus, et celle des firmes et établissements. Ces bases de données, qui sont déjà fort utiles tel quel à des fins de recherche, avancent nos connaissances par deux voies additionnelles. D'abord, en créant des données tout en les rendant accessibles à d'autrui, l'équipe de recherche doit porter une attention particulière à toute une série de questions concernant la confidentialité des données, les procédures pour éviter l'identification de répondants, et les méthodes d'accès aux données et de disponibilité des données. En outre, les connaissances acquises aux LEHD sur les méthodes d'appariement elles-mêmes, et sur les propriétés statistiques des données appariées, sont mises à la disposition de chercheurs de tous les domaines, qui pourront alors les utiliser dans leurs propres travaux. Nous présentons les méthodes utilisées pour empêcher l'identification des répondants, les méthodes d'appariement, ainsi qu'un survol de la sécurité de nos données.