

**Design and Conceptual Issues in Realizing Analytical Enhancements through Data  
Linkages of Employer and Employee Data**

**John M. Abowd**

**Cornell University, U.S. Census Bureau, CREST, and NBER**

**Julia Lane**

**American University, Urban Institute and U.S. Census Bureau**

**Ronald Prevost**

**U.S. Census Bureau**

**Presented at the November 2000 Meeting of the Federal Committee on Statistical  
Methodology, in Bethesda Maryland**

## **Abstract**

The maturation of the information age has created new challenges. Governments, faced with rapidly changing economies and societies are forced to make far-reaching economic and social policy decisions. These decisions, however, are based on limited, static, and expensive survey data. At the same time, large administrative data sets are available which are derived from data collected from households, business establishments and governmental entities. These data, which could describe the dynamic interactions of workers, businesses, government and society, are not fully used in the United States, not simply because the unique advantages of these data have not been clear, but also because key issues of confidentiality and access have not been fully addressed. This outcome has come at substantial cost. Linked data, if used wisely, has a great deal of scientific importance and can enhance our basic social data infrastructure in a number of ways. It can be used to reduce respondent burden, increase data quality, and enhance the information available to the federal, state and local agencies which rely on Census Bureau data for decision making. In addition, these data can provide a valuable tool to the research community.

The Longitudinal Employer-Household Dynamics Project (LEHD) creates data sets based upon the Census Bureau's demographic and economic products and using link information that permits the data sets to be longitudinal in both the household/individual and firm/establishment dimensions. These data sets, while of immense use in their own right, will also advance knowledge in two additional ways. First, by creating the data and developing the access modality, the research team explicitly addresses a series of confidentiality and access/availability issues, using internal Census Bureau expertise and that of external researchers at Cornell University, American University, the University of Maryland and National Center for Supercomputing Applications. Second, LEHD advances the knowledge of both linkage technology and the statistical properties of linked data so that researchers in all disciplines can use these techniques.

An innovative set of access support tools that combine a complete simulated environment (at the Cornell University support site) with the latest in web-based collaboration tools (from the NCSA at the University of Illinois) are being developed. These tools will be integrated with video/computer teleconferencing access to the Census Bureau and Cornell sites. Under carefully specified rules that encompass established confidentiality and disclosure review procedures, the Census Bureau will support external research of the linked data, using the web-based support tools and the video teleconferencing facilities.

The Cornell University component also includes the support of a restricted-access data site that will house confidential linked data from the national statistical agencies of other countries. Linked French data from INSEE have already been approved for this site. Statistics Sweden has expressed a willingness to complete an agreement that would allow linked Swedish data. The DIW in Germany will release a linked version of the GSOEP to the site, if it is funded. Efforts continue to negotiate restricted access agreements with other national agencies. The restricted access data site provides linked data analysis tools and a supercomputer facility for use with these data.

The new knowledge that can be generated from these data is potentially far reaching. The prototype American data sets provide the capacity to address fundamental questions in social and economic behavior. The restricted-access data from other countries provides a laboratory in which to test the generality of results found for a particular country. All of the data advances in the LEHD provide the opportunity to discover technological advances in confidentiality protection while enabling new partnerships to be formed across disciplines that focus on understanding social and economic systems, organizations and institutions.

## ***Introduction***

The maturation of the information age has created new challenges. Governments, faced with rapidly changing economies and societies are forced to make far-reaching economic and social policy decisions. These decisions, however, are based on limited, static, and expensive survey data. At the same time, large administrative and survey data sets are available which are derived from data collected from households, business establishments and governmental entities. These data, which could describe the dynamic interactions of workers, businesses, government and society, are not effectively used in the United States, because key issues have not been fully addressed. In particular, although American statistical agencies have the legal means and mandates to use such data to improve the scope and quality of statistical modeling, and have continually worked to maintain the confidentiality of these data sources and to protect the privacy of American citizens, they have made far less scientific and policy use of restricted-access, linked data than have many other national statistical agencies.

This outcome has come at substantial cost. Linked data, if used wisely, can reduce respondent burden, increase data quality, and enhance the information available to the federal, state and local agencies which rely on Census Bureau data for decision making. These data also provide a valuable tool to the research community. Indeed, other societies, notably in Europe, have made careful and wise use of such data and have demonstrated the enormous scientific gains from developing restricted-access files [SC99], [HK99]. Consequently, while for more than two decades, public policy around the world has been dominated by quantitative analyses based upon public-use American microdata samples, the very real prospect is that U.S. social policy will in the future depend heavily upon quantitative analyses derived from restricted-access European data. Just as public-use American microdata files have inspired countless research programs in the social sciences, medical sciences, biology, environment, statistical methods, and engineering, the European restricted-access linked data are inspiring new research programs. These new programs are focused on the interaction of employers and employees over time, the dynamics of social program participation, social organization, population demographics, and health care delivery, to name but a few.

The demonstrated success of the European approach over the past decade, combined with technological advances in confidentiality protections, suggest that the time has come for the American national statistical agencies to join its European counterparts in investigating many more potential uses of linked data. The Census Bureau needs the partnership of researchers in many disciplines from the scientific research community to improve linkage technology, to advance the understanding of the statistical properties of linked data, and to fully exploit the many different research opportunities in the social and behavioral sciences that are opened up by access to such data.

These gains cannot be realized, however, unless the partnership addresses additional issues of privacy, confidentiality, and data security. Scientific research of all kinds gains its legitimacy from the free exchange of ideas and the process of replication and extension. To protect the scientific integrity of the research based on new linked data sources, it is essential that

their use not be restricted exclusively to government employees and contractors, and that external scientists who agree to access restrictions and rules can conduct additional research using the linked data.

This adds one more critical research dimension to the partnership – we must address additional issues of privacy, confidentiality, and data security that arise as these new restricted-access data products are developed. Consider first the responsibility to maintain the confidentiality of the respondent’s data. Some Census Bureau data were acquired with the explicit promise that they would not be used for certain purposes. Such promises must always be respected—thus limiting the linkages that can be developed with those data. On the other hand, substantial amounts of Census Bureau data were and are collected with the explicit acknowledgement, notification, and informed respondent consent that they would be used in conjunction with other data for the purpose of improving the quality of statistical analyses of the economy. This informed consent permits the linking of these data under controlled, secure conditions without violating the confidentiality of the respondent data. Consider next the Census Bureau’s responsibility to maintain the confidentiality of the respondent’s identity as directed by Federal laws including the Privacy Act, the Trade Secrets Act, and Title 13 of the United States Code. Safeguarding this confidentiality means that only authorized persons may see Census Bureau data with identifiers and that any published results must mask the identity of individual respondents. It is vitally important to respect such promises while recognizing that there are a variety of legitimate ways to authorize tabulations and research using such data. Thus, an additional program of research into confidentiality and disclosure risk management must accompany any efforts to link data from multiple sources. Finally, we must consider the data security issues that arise because the conduct of legal, authorized research using linked confidential data requires the use of networked, large-scale computing systems. The research program must take the necessary precautions to ensure that unauthorized persons cannot gain access to these systems.

Having laid out the basics of our new partnership, we now discuss several examples of the benefits of linked data. Consider, as a first example of the research potential of linked data, a question that comes from our labor economics roots. How do American employers respond to an economic downturn? The correct answer appears to be “by laying off workers.” The work of [DH91] has shown that establishments respond to negative economic shocks by destroying jobs, a fact learned from linked annual censuses and surveys of manufactures. This destruction of jobs is statistically associated with an increased inflow to unemployment, which produces an increased rate of unemployment because unemployment durations do not decrease at the start of a recession. Using matched administrative data from state unemployment insurance systems [AM94] confirmed this statistical relation by directly observing that the individuals separated from their jobs in the destruction process did, indeed, pass into unemployment, [BL99] also using linked unemployment insurance data, further documented that there were important flows into and out of unemployment even when there was no job destruction (or creation). Does the process work the same way in other countries? The correct answer appears to be “no.” French linked data show that employers in that country respond to an economic downturn by imposing a hiring freeze [ACK99] The hiring freeze reduces employment because there are enough employees on fixed short-term employment agreements who must leave the firm at the end of

those agreements. Adjustment by hiring freeze is cheaper than adjustment by layoff, which involves large fixed costs in France. [AKR98]

The previous example focussed on the worker—but one key element of these data is their ability to characterize the interaction between the worker and the firm. In particular, what are the economic inputs of businesses in service industries? Historical data collection for such businesses has focused on measures of physical capital, material purchases, employment and payroll. An important source of the differences among service businesses is the quality of the workforce, as measured by its education and its opportunity wage rate. Direct collection of such data is enormously expensive. Dynamic, linked data permit the calculation of firm-level time series that summarize both observable workforce characteristics and unmeasured heterogeneity, potentially enabling us to characterize the impact of the workforce on firm performance, competitiveness, and survival.

These examples show the analytic benefits associated with linked demographic and business data quite clearly. A partial list of the uses of such data would include, in addition to the analysis of firm-specific contribution to pay determination, the effect of technological and structural change on earnings and employment—in particular, on earnings inequality; the effect of the firm wage setting and turnover policies on welfare recipients; the effect of firm expansion and relocation decisions on neighborhood demographic composition, worker commuting patterns and worker mobility; and a full accounting of the returns to both firm and worker of investments in enterprise training and research and development. Indeed, the advantages of combining information on the demand and supply side of the labor market in the form of linked data have been eloquently stated elsewhere [HA99].

Linked employer-employee data are also essential to help unravel some of the striking findings in the industrial organization and productivity literatures regarding the nature of business dynamics and the sources of micro and aggregate productivity growth. For example, recent research using establishment and firm level data has shown that a large fraction of aggregate productivity growth is accounted for by the reallocation of outputs and inputs from less productive to more productive establishments (including the substantial contribution of more productive entering establishments displacing less productive exiting establishments) [FHK98]. Thus, it is apparent that understanding both micro and aggregate productivity growth requires understanding the heterogeneous dynamics of micro businesses. Moreover, recent theoretical and empirical work suggests that there is a close link between the dynamics of businesses and productivity at the micro level and changes in the structure of the workforce within and between businesses [HLS99]. While there has been substantial use of household data and business microdata separately to seek to understand the connection between skill-biased technical change and changes in the structure of the workforce [DDT97], [DHT97], the connections between technical change, productivity growth and changes in the workforce at the micro and aggregate level can only be directly evaluated with linked data on businesses and workers.

The effect of linked data on other research fields is equally promising. The environmental impact of different firms' pollution levels has already been studied using micro firm data [MPS97]; linked data will enable environmentalists to examine the impact on workers. The medical field can also gain from linked data. For example, the Current Population Survey

has been matched to the National Death index to create a National Longitudinal Mortality Survey to examine, with large sample sizes, and without expensive clinical studies the effect of income and employment on death rates. [R92] Using further linked data, such studies can be extended to look at the effects of earnings and employment history, firm personnel practices, and the availability of health benefits on mortality. The education literature has often argued that firm-based training is an important way not only of reorienting workers to changing technology and exports but also making firms more able to compete in the global marketplace. [TB95] Linked data enable us to directly examine this impact. The dynamic aspects of absorbing immigrants into the workplace can be examined, as we see whether workers from different countries sort into firms that have hired their countrymen; what kinds of firms employ immigrants, the dynamics of immigrant mobility across firms as well as a host of other issues [SE97]. Finally, the use of linked data from three European countries: Sweden, France and Germany, permits, for the first time, detailed international comparisons of the dynamic interrelationships among firms and workers—thus extending the research possibilities to international as well as national scientists.

The remainder of this project description is as follows. We first describe the Longitudinal Employer-Household Dynamics Project (LEHD). We next describe the relationship between our project and the Center for Economic Studies, the Census Bureau entity that has developed the research versions of confidential establishment-level data. We continue by describing the resources that the Census Bureau has committed to the LEHD, beyond its regular commitments to data development in the demographic, economic and methodology areas. Next, we describe the Cornell University web support and restricted access data sites. The web support site provides the innovative access support for the Census Bureau linked data that we propose to develop. The restricted access site provides direct web-based access to confidential linked microdata from other countries, subject to access rules that are determined by the contributing country and Cornell University.

## ***The Longitudinal Employer-Household Dynamics Project***

### **Overview**

The core of the Census Bureau component of this activity is the development and potential extension of the linked data files known collectively as the Longitudinal Employer-Household Dynamics (LEHD) project. An overview of the structure of data from the project is presented in Figure 1. The LEHD represents a substantial investment made by the Census Bureau in order to permit direct linking of its demographic surveys (household-based instruments) with its economic censuses and surveys (establishment and business unit-based surveys). The main data products produced by this project are integrated versions of the Title 13 protected data from the Current Population Surveys, Surveys of Income and Program Participation, Surveys of Program Dynamics, American Community Surveys, Standard Statistical Establishment Lists and Economic Censuses and Surveys.<sup>1</sup> We propose linking the internal versions of the files because the final product will be Title 13 protected at restricted sites, not public access. The demographic variables will not be restricted to those on public use files.

---

<sup>1</sup> In Census Bureau terminology a “demographic” product uses a household as the basic unit of data collection while an “economic” product uses an establishment or business as the basic unit.

Such a vast collection of information from both sides of the Census Bureau’s traditional programs has never been attempted before. The direct benefits to the Census Bureau are substantial. The benefits to social science research are enormous. In this section we outline the fundamentals of the LEHD data linkages and describe the nature of the continuing investment required to make such a product accessible for general scientific research.

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

### Privacy and Confidentiality

Participants in all surveys used here gave informed consent to use the data that they provided in conjunction with data from other sources to improve the statistical analyses of the Census Bureau. The linked data files reside on the LEHD dedicated computer with all personal identifiers removed. The LEHD computer is a special-purpose machine, described below, isolated within its own firewall, within the LEHD and Center for Economic Studies firewalls, within the Census Bureau’s general firewall. Details of the data security plan are given below.

## The Longitudinal Employer - Household Dynamics Project

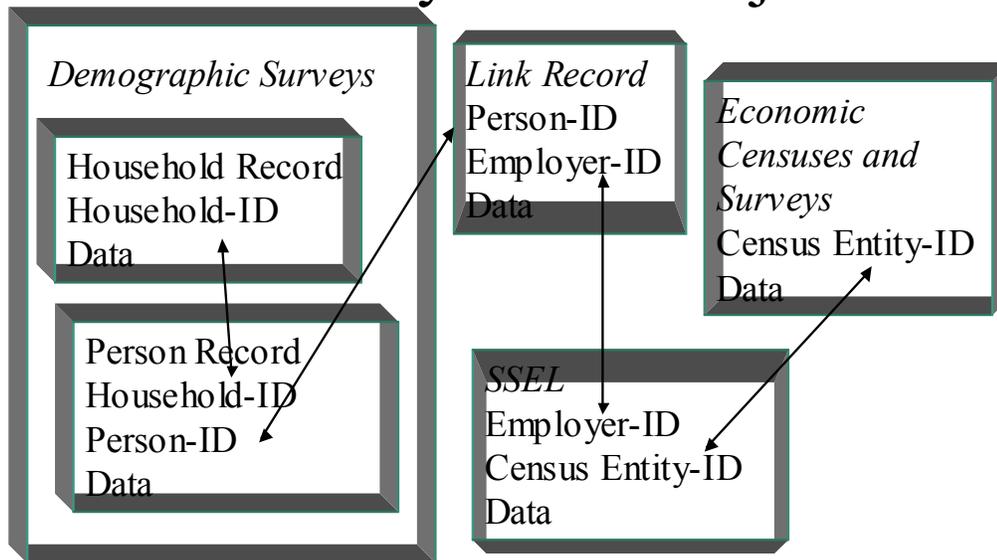


Figure 1

## Access Rules Governing the LEHD Data

Research conducted on the LEHD data will take place under a set of rules and limitations that are considerably more constraining than those prevailing in typical research environments:

- The solicitation of research proposals using the LEHD data is a public process.
- Proposals to use the LEHD data should be submitted to the National Science Foundation, the National Institute for Aging or another appropriate funding agency with strict peer-review procedures. (Permission to secure funding from an agency other than the two listed must be secured in advance from the Census Bureau.)
- Successful peer-reviewed proposals to carry out research using the LEHD data must then be approved by the Census Bureau, which will use an internal review panel consisting of subject matter, database and disclosure specialists. The Census Bureau review will confirm that it would benefit from the proposed research (a requirement for granting access to Title 13 protected data), that the proposed research is feasible using the LEHD data, that the variables and cases requested can be supplied, and that appropriate resources exist to undertake the analysis. The internal review will follow the procedures developed by the Census Bureau for the review of projects using administrative data with the review process managed through the Center for Economic Studies.
- Researchers using the LEHD data will be required to obtain Special Sworn Status from the Census Bureau. To obtain this status, researchers will be required to undergo a security check, including fingerprinting.
- Researchers holding Special Sworn Status are subject to the same legal penalties as regular Census Bureau employees for disclosure of confidential information. The penalties are a fine of up to \$5,000, imprisonment for up to five years, or both.
- All analyses of the LEHD microdata will be conducted using the Title 13 secure facility that the Census Bureau has provided for LEHD with access from Federal Building 3 and the Center for Economic Studies. Access is strictly limited to researchers and staff authorized by the Census Bureau. The computers and local area network connected to the LEHD secure facilities are not accessible from outside the Census Bureau's firewall. 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. 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.
- Researchers may not remove any Title 13 confidential data—whether recorded on any medium or merged with non-confidential data from the LEHD facility. All research findings must be submitted to Census Bureau personnel for disclosure review prior to public release of any kind

The relation of the Cornell web support site to the LEHD project is that the web support site will provide a complete simulated environment with real documentation and metadata but no Title 13 protected confidential data. Researchers may use the Cornell site to develop proposals and to prepare computer programs. Principal Investigators would be expected to conduct their actual analyses at the approved access sites at LEHD and CES. Video teleconferencing in the presence of Census disclosure specialists, on-site database specialists, and computer

programmers will be provided as a means of facilitating a Principal Investigator's completion of the approved research. A remote site that wishes to use the video conferencing facilities to communicate with LEHD or CES must use the Census Bureau standards, discussed below, which have been developed to preserve the integrity of the Title 13 confidentiality protections.

### ***The Center for Economic Studies***

The Center for Economic Studies at the Census Bureau supports research and development of Census microdata for analytical purposes by internal Census staff and the external research community. CES has been in operation since 1983 and is directed by the Office of the Chief Economist. CES maintains an external research lab at headquarters and a network of regional Research Data Centers that provide access to internal confidential Census microdata in a secure environment. Research Data Centers are now located in Boston, Carnegie Mellon University, University of California at Berkeley, UCLA, and the Research Triangle area located at Duke University. The Research Data Centers provide access to a wide range of internal confidential data on both households and businesses. Researchers gain access by submitting proposals to the Census Bureau that are evaluated by peer review panels from both the research community and Census. Research proposals are evaluated on the basis of scientific merit, feasibility, benefit to the Census Bureau (this is required by law), and disclosure risk (i.e., no research output can be removed from secure Census Bureau facilities that would reveal the data of individual or business respondents).

The initial database development efforts at CES focused on developing and providing access to the internal microdata on U.S. businesses (these data are drawn from the Economic Censuses and annual surveys of businesses conducted by the Census). More recently, CES has begun to acquire and provide access to internal microdata on households for analytical use by the research community. These database development efforts require substantial time and resources at headquarters by Census Bureau staff to acquire, document and to become sufficiently familiar with the internal microdata files to be able to provide access to external researchers. Adequate documentation and sufficient familiarity with the internal micro files are essential to provide adequate support to researchers and also to insure that disclosure analysis of research output can be appropriately and efficiently conducted. It is for these reasons that historically CES has developed files in headquarters and provided research access at headquarters before providing access to data at the regional Research Data Centers. Substantial progress has been made in developing research databases in this manner. Until recently, the only files available to Research Data Centers were data from manufacturing economic censuses and surveys but there is increasing focus on providing non-manufacturing microdata on businesses and selected demographic data (for additional information see <http://www.census.gov/ces/ces.html>). This pattern of database development efforts is important in the current context in that, initially, the only access to linked employer-employee databases developed under the LEHD will be at headquarters. These plans are consistent with past database development efforts at CES. The LEHD databases are sufficiently different in their scope, magnitude and concerns about confidentiality that the emphasis in the first five years of this project are to develop the data and to provide direct access to the microdata at headquarters (with the support of the Cornell University web site). After this initial phase of database development, documentation and

research access at headquarters, an evaluation will be conducted regarding the feasibility of providing access to the LEHD data at Research Data Centers.

Beyond access to the LEHD data, the support site at Cornell University and the video teleconferencing facilities can enhance the access to the wide range of internal microdata made available by CES at headquarters and at the Research Data Centers. For example, for projects involving internal microdata for which CES is currently providing access at headquarters or at Research Data Centers, a researcher could initiate a project at headquarters or at a Research Data Center. During this phase of a project, exploratory data analysis and construction of relevant research databases requires substantial hands-on examination of the data. In the later stages of the project, when statistical analyses are being developed and implemented on specific well-defined research databases, the support site and the video teleconference facilities could be used intensively. Thus, researchers would have to spend less time in residence at headquarters or at a Research Data Center. These innovative access support tools make Census Bureau microdata available to a significantly greater range and number of researchers across the country.

### ***Research Agenda for the Census Bureau Site***

There are four separate dimensions that make up the research agenda. The first of these is to address the unique confidentiality and access issues raised by the use of linked data. Although some of these have been highlighted by papers given at a recent conference on linked data, we are working with internal Census personnel, the new National Academy of Sciences panel on confidentiality, and international agencies to further develop strong disclosure protocols—in particular, examining the potential for tailoring different access protocols to data sets with different degrees of sensitivity, while maintaining equal Title 13 protection for all the data. Different agencies, such as Eurostat in Luxembourg, Nomis/R-Cade in England; OECD in Paris and the World Bank at headquarters have come up with different solutions, which should be evaluated for their applicability to these kind of data. The Cornell University and NCSA components of the project, described in detail below, will examine further how technology can be used to maintain high confidentiality, keep sensitive microdata at the Census Bureau’s main site, but provide imaginative access solutions.

The second is to further develop linking technology—by improving basic matching algorithms, creating intermediary information and structures for linking, and enhancing internal validation procedures. The Census Bureau has been a pioneer in developing matching algorithms, introducing methods for creating analytically valid and confidential public-use, micro-data files, and using ideas for adjusting statistical and economic analyses for linkage error. The infrastructure project will update current matching techniques, as well as compare the performance of commercially available software to the customized versions of the general programs written by the Census Bureau [J89], [W95], [W97], [W98], [W99], [KW98], [WP97], [SW93], [SW97].

Census Bureau Staff will also help investigate alternative computational methods for the solution of very large linear statistical models (fixed and random); for example, of wages versus person and firm effects. These models may be based upon tens of millions of records with several

million person effects and several million firm effects. Usual computational methods for solving the full model are not practical, so alternative iterative or sparse computational methods will be investigated. Parallel computing methods will be used as needed. Algorithm development for the large-scale, parallel computing techniques and supporting programming tools will be available at both the Census Bureau and the Cornell support sites. [AKR99a]

The final dimension is to make programmatic and analytical use of the linked database. The programmatic uses, (a) improving the quality of current survey information, (b) adding information to the storehouse of social and economic knowledge without additional respondent burden, (c) improving the quality of longitudinal data, increase the quality of those Census data which are used by all researchers, not simply those involved in this project. The analytical uses of the data set will be determined by both the priorities of the external research community and by the two involved government agencies.

### ***The Role of Census Bureau Staff***

The Census Bureau sits at the center of this project, because the protection afforded by Title 13 provides the LEHD with the safeguards necessary to protect such sensitive data. In addition to the linkage technology, the confidentiality research, and the programmatic research which sit at the heart of the research agenda, Census Bureau staff will be involved in two additional ways:

- Protecting confidential records: A small core of Census Bureau staff will strip all identifying information from the files and replace them with a unique personal identifier. Census Bureau staff will be involved in all aspects of data processing, from evaluating and certifying the quality of the data links, to developing new, workable confidentiality and access protocols to conducting the disclosure review process.
- Metadata repository: A major internal aim of the project is to provide online documentation of variable definitions, surveys. Statistical metadata will be used to facilitate survey design, processing, management, and analysis. It is the electronic storage and organization of statistical metadata, which will allow statistical agencies to develop, automated survey design and processing systems.

### ***The Cornell University Component***

#### **Overview**

The Cornell University component of this activity provides for two distinct entities at the university: the web support site for the Census Bureau component of infrastructure project and the restricted-access data center. We describe each of these below.

The Cornell University web support site will consist of a complete simulated environment with full documentation of the LEHD data and access to the LEHD metadata. The Center for Economic Studies (CES) will also participate in the development of the web support site. At the request of the Chief Economist, the web site will also develop a support environment for other restricted-access data files maintained by the CES. This support will be substantially similar to that given to the LEHD—simulated microdata and computing environment, documentation of

variables, non-Title 13 metadata, video teleconferencing consultations with database and disclosure specialists. No Title 13 protected confidential data will be accessible through the Cornell University web support site. Disclosure review specialists, policy officers and legal staff (all Census Bureau employees) will carefully monitor the creation and use of the web support site, since there are policy issues regarding the Title 13 protection of data documentation that require Census Bureau review. We will initiate a review procedure at the Census Bureau to establish a policy for approving documentation and metadata for the web support site.

While the notion of the Cornell University web support site serving as a complete simulated environment with full documentation is quite appealing and would clearly enhance the ability of researchers to exploit the LEHD databases (while protecting the confidentiality of the individual records at the same time), its feasibility hinges largely upon the Census Bureau's ability to create and transfer a significant amount of metadata to this site. The Census Bureau is committed to creating a metadata repository and has assigned its metadata development team to work with the LEHD to facilitate the test development of appropriate metadata from the outset.

### ***Summary***

The scientific importance of this project rests on several dimensions. Our program extends the state of the art in developing new approaches to access highly sensitive and confidential data, which could otherwise not be used. We enhance linking technology to increase the value added from different data sets without incurring the increased cost and respondent burden from new surveys. We create, for the first time in the United States, three separate data sets which have the ability to analyze the dynamic interrelationships between firms, workers and their characteristics—that is, describe both sides of the labor market rather than each in isolation. Finally, we provide restricted access to confidential linked data from other countries that permits detailed international comparisons.

## References Cited

- [AK93] Abowd, John M. and Francis Kramarz, "A Test of Negotiation and Incentive Compensation Models Using Longitudinal French Enterprise Data," in J.C. van Ours, G.A. Pfann and G. Ridder, eds. *Labour Demand and Equilibrium Wage Formation Contributions to Economic Analysis* (Amsterdam: North-Holland, 1993), pp. 111-46.
- [AL93] Abowd, John M. and Thomas Lemieux, "The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada," *Quarterly Journal of Economics* 108 (November 1993): 983-1014.
- [AB95] Abowd, John M. and Michael Bognanno, "International Differences in Executive and Managerial Compensation" in R.B. Freeman and L. Katz, eds. *Differences and Changes in Wage Structures* (Chicago: NBER, 1995), pp. 67-103.
- [AA96] Abowd, John M. and Laurence Allain, "Compensation Structure and Product Market Competition," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 207-217.
- [AKM96] Abowd, John M., Francis Kramarz and Antoine Moreau, "Product Quality and Worker Quality," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 300-322.
- [AK96a] Abowd, John M. and Francis Kramarz. (eds.) *The Microeconometrics of Human Resource Management*, special issue of *Annales d'économie et de statistique* 41/42 (Paris: ADRES, January/June 1996).
- [AK96b] Abowd, John M. and Francis Kramarz, "The Microeconometrics of Human Resource Management: International Studies of Firm Practices, Introduction and Overview," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 1-9 (French), 11-19 (English).
- [AK96c] Abowd, John M. and Francis Kramarz, "Les Politiques Salariales: Individus et Entreprises" (Compensation Policies: Individuals and Firms), *Revue Economique* 47 (May 1996): 611-622.
- [AK96d] Abowd, John M. and Francis Kramarz, "The Economic Analysis of Compensation Systems: Collective and Individual" in Norman Bowes and Alex Grey, eds. *Job Creation and Loss: Analysis, Policy and Data Development* (Paris: OECD, 1996), pp. 47-54.
- [AKMT96] Abowd, John M., Francis Kramarz, David Margolis and Kenneth Troske, "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," in *Comparaisons internationales de salaires* (Paris: Ministère du travail et des affaires sociales and INSEE, 1996), pp. 315-327.
- [AKM97] Abowd, John M., Francis Kramarz and Thierry Magnac, "Labor Demand, Quits, and Adjustment Costs," October 1997.
- [AFKR97] Abowd, John M., Hampton Finer, Francis Kramarz and Sébastien Roux, "Job and Wage Mobility: An Analysis of the Dynamics of Employment Durations Using Matched Employee and Employer Data from the U.S. and France," July 1997.
- [AKR98a] Abowd, John M. and Francis Kramarz, "Internal and External Labor Markets: An Analysis of Matched Longitudinal Employer-Employee Data" in J. Haltiwanger, M. Manser, and R. Topel (eds.) *Labor Statistics and Measurement Issues* (Chicago: University of Chicago Press, 1998), pp. 357-370.
- [AKR98b] Abowd, John M. and Francis Kramarz, "The Costs of Hiring and Separations," July 1998.

- [AKM98] Abowd, John M., Francis Kramarz and David Margolis, "Minimum Wages and Employment in France and the United States," July 1998 (under review at the *AER*).
- [AKMT98] Abowd, John M., Francis Kramarz, David Margolis and Kenneth Troske, "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," August 1998.
- [ACOK99] Abowd, John M., Patrick Corbel and Francis Kramarz, "The Entry and Exit of Workers and the Growth of Employment: An Analysis of French Establishments" *Review of Economics and Statistics* (1999), forthcoming.
- [ACRK99] Abowd, John M., Bruno Crépon and Francis Kramarz "Moment Estimation with Attrition" January 1999 (under second review at *JASA*).
- [AFK99] Abowd, John M., Hampton Finer and Francis Kramarz, "Individual and Firm Heterogeneity in Compensation: An Analysis of Matched Longitudinal Employer-Employee Data for the State of Washington" in J. Haltiwanger *et al.* eds., *The Creation and Analysis of Employer-Employee Matched Data*, (Amsterdam: North Holland, 1999), forthcoming.
- [AKA99] Abowd, John M. and David Kaplan, "Executive Compensation: Six Questions That Need Answering," *Journal of Economic Perspectives* (1999): forthcoming.
- [AKR99a] Abowd, John M. and Francis Kramarz, "Econometric Analysis of Linked Employer-Employee Data," *Labour Economics* (February 1999): forthcoming.
- [AKR99b] Abowd, John M. and Francis Kramarz, "The Analysis of Labor Markets Using Matched Employer-Employee Data," in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3, Chapter 26 (Amsterdam: North Holland, forthcoming, 1999).
- [AKR99c] Abowd, John M., Francis Kramarz, Thomas Lemieux, and David Margolis, "Minimum Wages and Youth Employment in France and the United States," in D. Blanchflower and R. Freeman (eds.) *Youth Unemployment and Employment in Advanced Countries* (Chicago: University of Chicago Press, forthcoming).
- [AKM99] Abowd, John M., Francis Kramarz and David Margolis, "High Wage Workers and High Wage Firms" *Econometrica* (1999): 251-334.
- [A96] Allain, Laurence "Essays in Compensation and Unemployment Insurance," Cornell University dissertation in Industrial and Labor Relations, specializing in Labor Economics, 1996.
- [AM84] Anderson, P. and Meyer, B. (1994) 'The Extent and Consequences of Job Turnover' *Brookings Papers on Economic Activity*
- [BHC92] Baily, Martin N., Charles Hulten, and David Campbell, "The Distribution of Productivity in Manufacturing Plants," *Brookings Papers: Microeconomics*, (1992).
- [BDH96] Baldwin, J., T. Dunne and J. Haltiwanger (1996) "Job Creation and Destruction in the Manufacturing Sectors of Canada and the United States," working paper, November.
- [B98] Balmaceda, Felipe "Essays on Incentives in Organizations" Cornell University dissertation in Industrial and Labor Relations, specializing in Labor Economics, 1998.
- [BA99] Bayard, K, Hellerstein, J Neumark, D and Troske, K "Why are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation Using the New Worker-Establishment Characteristics Database" in J. Haltiwanger *et al.* eds., *The Creation and Analysis of Employer-Employee Matched Data*, (Amsterdam: North Holland, 1999), forthcoming.

- [BBG94] Berman, Eli, John Bound, and Zvi Griliches, "Changes in the Demand for Skilled Labor Within U.S. Manufacturing Industries: Evidence from the Annual Survey of Manufacturing," *Quarterly Journal of Economics*, 109 (May 1994): 367-398.
- [BL99] Burgess, S., J. Lane and D. Stevens 'Job Flows, Worker Flows and Churning' *Journal of Labor Economics*, forthcoming
- [CEH97] Caballero, R., E. Engel, and J. Haltiwanger (1997) "Aggregate Employment Dynamics: Building From Microeconomic Evidence," *American Economic Review*, March.
- [CEH95] Caballero, R., E. Engel, and J. Haltiwanger (1995) "Plant-Level Adjustment and Aggregate Investment Dynamics," *Brookings Papers on Economic Activity*.
- [C98] Caselli, Francesco, "Technological Revolutions," mimeo, University of Chicago (April 1998).
- [CHP95] Cooper, R., J. Haltiwanger, and L. Power (1995) "Machine Replacement and the Business Cycle: Lumps and Bumps," NBER Working Paper No. 5620, September.
- [DH91] Davis, Steve J. and John Haltiwanger, "Wage Dispersion Between and Within U.S. Manufacturing Plants, 1963-1986," *Brookings Papers on Economic Activity: Microeconomics*, (1991): 115-200.
- [DH96a] Davis, S. and J. Haltiwanger (1996) "Driving Forces and Employment Fluctuations," NBER Working Paper No. 5775, September.
- [DH96b] Davis, S. and J. Haltiwanger (1996) "Employer Size and the Wage Structure in U.S. Manufacturing," *Annales D'Economie et de Statistique*, January.
- [DH97] Davis, S. and J. Haltiwanger (1997) "Sectoral Job Creation and Destruction Responses to Oil Price Changes and Other Shocks" Working Paper, May.
- [DH98] Davis, S., J. Haltiwanger (1998) "Measuring Gross Worker and Job Flows", in *Labor Statistics Measurement Issues*. Chicago: University of Chicago Press.
- [DHS96] Davis, S., J. Haltiwanger and S. Schuh (1996) *Job Creation and Job Destruction*. Cambridge, Massachusetts: MIT Press.
- [DLM97] Davis, S., P. Loungan and R. Mahidhara (1997) "Regional Labor Fluctuations: Oil Shocks, Military Spending and other Driving Forces," working paper, January.
- [DH97a] Davis, S. and M. Henrekson (1997) "Industrial Policy, Employer Size and the Economic Performance in Sweden," in *The Welfare State in Transtion*. Chicago: University of Chicago Press.
- [DH97b] Davis, S. and M. Henrekson (1997) "Explaining National Differences in the Size and Industry Distribution of Employment," working paper, August.
- [D97] Davis, S. (1997) "Sorting, Learning, and Mobility When Jobs Have Scarcity Value: A Comment," *Carnegie-Rochester Series on Public Policy*, 46.
- [DDT97] Doms, Mark, Timothy Dunne, and Kenneth Troske, "Workers, Wages, and Technology," *Quarterly Journal of Economics*, 112 (1997): 253-290.
- [DHT97] Dunne, Timothy, John Haltiwanger, and Kenneth R. Troske. "Technology and Jobs: Secular Change and Cyclical Dynamics," *Carnegie-Rochester Public Policy Conference Series*, 46, 1997.
- [F98a] Finer, Hampton "Firm And Worker Heterogeneity In The National Longitudinal Survey Of Youth" August 1998.
- [F98b] Finer, Hampton "Wage Determination and Firm Performance in the Presence of Individual and Firm Heterogeneity" Cornell University dissertation in Industrial and Labor Relations, specializing in Labor Economics, 1998.

- [FHK98] Foster, Lucia, John Haltiwanger, and C.J. Krizan (1998) "Aggregate Productivity Growth: Lessons from Microeconomic Evidence," NBER Working Paper No. 6803, November.
- [HLS] Haltiwanger, John C, Julia Lane and James Spletzer, "Productivity Differences Across Employers: The Role of Employer Size, Age, and Human Capital" *American Economic Review* (May 1999), forthcoming.
- [HA99] Hamermesh, Daniel 'LEEPing into the Future of Labor Economics' *Labour Economics*, (February 1999), forthcoming.
- [HNT98] Hellerstein, Judith, David. Neumark and Kenneth. Troske "Market Forces and Sex Discrimination" mimeo, 1998
- [J89] Jaro, M.A. "Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida" *Journal of the American Statistical Association*, 89, (1989): 414-420
- [J82] Jovanovic, Boyan. "Selection and the Evolution of Industry," *Econometrica*, 50(3), 1982, 649-70.
- [JM94] Jovanovic, Boyan, and Glenn M. MacDonald. "Competitive Diffusion," *Journal of Political Economy*, 102 (1) 1994, 24-52.
- [JMP93] Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, "Wage Inequality and the Rise in the Return to Skill," *Journal of Political Economy*, 101 (1993): 410-42.
- [K98] Kaplan, David, "Essays on Incentives and Compensation: Theory and Evidence" Cornell University dissertation in Economics, specializing in Labor Economics, 1998.
- [KM92] Katz, Lawrence F. and Kevin Murphy, "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *Quarterly Journal of Economics*, 107 (1992): 1-34.
- [KW97] Kim, J. J., and Winkler, W.E., (1997), "Masking Microdata Files," Statistical Research Division research paper 97/03. Also appeared in *American Statistical Association, Proceedings of the 1995 Section on Survey Research Methods*.
- [K93] Kremer, Michael. "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 108 (1993): 551-575.
- [KM96] Kremer, Michael and Eric Maskin, "Wage Inequality and Segregation By Skill," NBER Working Paper, No. 5718 (August 1996).
- [KA93] Krueger, Alan, "How Computers Changed the Wage Structure: Evidence from Microdata, 1984-89," *Quarterly Journal of Economics* CVIII (February 1993)" 33-60.
- [LBT97] Lane, J, S. Burgess and J. Theeuwes "The Uses of Longitudinal Matched Employer/Employee Data in Labor Market Analysis" *Proceedings of the American Statistical Association*, 1998
- [L93] Lichtenberg, Frank R., "The Output Contributions of Computer Equipment and Personnel: A Firm-Level Analysis," NBER Working Paper, No. 4540 (November 1993).
- [LM99] Lumsdaine and Mitchell, "Retirement Behavior," in *The Handbook of Labor Economics*, O. Ashenfelter and D. Card, eds. (Amsterdam: North-Holland, 1999), forthcoming
- [M93] Margolis, David N. "Compensation Practices and Government Policies in Western European Labor Markets," Cornell University Dissertation in Economics, 1993.
- [MPS97] Morgenstern, R, W. Pizer and J. Shih "Are We Overstating the Economic Costs of Environmental Protection?" CES Working paper 97-12 (October 1997).

- [MB91] Morrison, Catherine and Ernst R. Berndt, "Assessing the Productivity of Information Technology Equipment in U.S. Manufacturing Industries," NBER Working Paper, No. 3582 (January 1991).
- [OS94] Oliner, Stephen and Daniel Sichel, "Computers and Output Growth Revisited: How Big is the Puzzle?" *Brookings Paper on Economic Activity*, 2, (1994).
- [OP96] Olley, Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64 (November, 1996): 1263-1297.
- [R92] Robot, E "Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study" *Public Health Reports* v. 107 (July/Aug. 1992): 457-61.
- [SW93] Scheuren, F.S. and Winkler, W.E. (1993), "Regression Analysis of Data Files that are Computer Matched," *Survey Methodology*, *Survey Methodology*, 23, 157-165. Also appeared in *Record Linkage Techniques 1997*, (1999), Alvey, W. and Jamerson, B. (eds.), National Academy Press, Washington, DC.
- [SW97] Scheuren, F.S. and Winkler, W.E. (1997), "Regression Analysis of Data Files that are Computer Matched, Part II," *Survey Methodology*, *Survey Methodology*, 23, 157-165. Also appeared in *Record Linkage Techniques 1997*, (1999), Alvey, W. and Jamerson, B. (eds.), National Academy Press, Washington, DC.
- [SE97] Smith, J and B. Edmonston (eds) *The New Americans: Economic, Demographic and Fiscal Effects of Immigration* (Washington D.C, National Academy of Sciences). 1997.
- [TB95] Tan, H and G. Batra, "Enterprise Training in Developing Countries: Incidence, Productivity Effects and Policy Implications" (Washington, DC: World Bank, 1995).
- [TR99] Troske, Kenneth "The Worker Establishment Characteristic Database," in J. Haltiwanger, M. Manser and R. Topel, eds., *Labor Statistics Measurement Issues* (Chicago: NBER, 1998).
- [W95] Winkler "Matching and Record Linkage" in B.G. Cox (ed.) *Survey Methods for Businesses, Farms and Institutions*, (New York: J. Wiley, 1995).
- [W97] Winkler, W.E. (1997), "Views on the Production and Use of Confidential Microdata," Statistical Research Division research paper 97/01.
- [W98] Winkler, W.E. (1998), "Producing Public-Use Microdata that are Analytically Valid and Confidential," Statistical Research Division research paper 98/02. Also appeared in *American Statistical Association, Proceedings of the 1997 Section on Survey Research Methods*.
- [W99] Winkler, W.E. (1999), "Re-identification Methods for Evaluating the Confidentiality of Analytically Valid Microdata," *Research in Official Statistics*, to appear.
- [WP97] Winkler, W.E., and Porter, E.H. (1997), "Approximate String Comparison and its Effect in an Advanced Record Linkage System," Statistical Research Division research paper 97/02. Also appeared in *Record Linkage Techniques 1997*, (1999), Alvey, W. and Jamerson, B. (eds.), National Academy Press, Washington, DC.