Bureau of the Census

# LEHD Newsletter

We'd like to update you on what we've done about the key issues raised at the last workshop.Please remember that nothing in this is "official" Census Bureau data: this is an informal update on the LEHD program's activities.

Volume 2 Issue 1 October 2002

The LEHD program has moved ahead by leaps and bounds since the last newsletter. We have received funding to enable us to expand the program; we are getting ready to move to production mode in time for our next state workshop (January 29-30 2003); we are developing a web-based delivery system for our flagship product—the quarterly workforce indicators; and we have made significant steps forward in our research: low-wage work; the demand for older workers; human capital and immigration. Our work has been recognized with both national and international accolades.

In all of this, we depend on you, our state partners. We need your help, advice and support. In this edition of the newsletter, in addition to our regular updates, we provide you with a delivery schedule for the first pass of our Quarterly Workforce Indicators (formerly known as the EDE's), and a tentative production schedule. We'd like your feedback!

## Special points of interest:

1. The Employment and Training Administration has provided funding to add eight more states to the LEHD program. We welcome Oregon as a new partner
2. The Bureau of Transportation Statistics has provided pilot funding to improve ES-202 coding and develop origin-to-destination information on commuting patterns for workers of different income levels. Florida and Illinois have agreed to pilot the program.
3. The LEHD program is partnering with the Bureau of Labor Statistics to investigate whether LEHD data can be used to inform official productivity statistics.
4. We have developed a web based delivery system which we hope will enable Workforce Investment Boards, Chambers of Commerce and a wide variety of clients to immediately use the Quarterly Workforce Indicators (formerly known as Employment Dynamics Estimates)
5. The LEHD program has been endorsed by the Governor of Illinois, the California Secretaries of Health and Human Services and Workforce Development, the Florida Secretary of Education; the National Association of State Workforce Board Chairs, and numerous other users—workforce boards, transportation agencies and institutional researchers.

## Inside this issue:



## I. Quarterly Workforce Indicators (QWI's)

### Delivery Schedule

*Version 2.3.*

This version has already been delivered to each partner state. It consisted of
1. State x Industry (1 Digit) x Demographic Characteristics x Indicator
2. State x County of Work x Demographic Characteristics x Indicator

*Version 3.0:*

Run 1 (Delivered Fall 2002):
1. State x Industry (4 digit) x Demographic Characteristics x Indicator
2. State x WIB of Work x Industry (1 digit) x Demographic Characteristics x Indicator

Run 2 (Delivered after Run 1, CPU cycles, disk space, etc. permitting):
1. State x County of Work x Demographic Characteristics x Indicator
2. State x County of Work x Industry (1 digit) x Demographic Characteristics x Indicator

Run 3 (Delivered after Run 2, CPU cycles, disk space, etc. permitting):
1. State x MSA of Work x Industry (1 digit) x Demographic Characteristics x Indicator

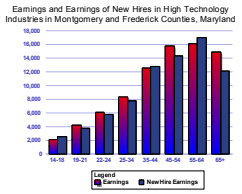
*Version 3.1 ( Delivered April/May, 2003):*

Run 1:
1. State x WIB of Work x Industry (4 digit) x Demographic Characteristics x Indicator
2. State x WIB of Residence x Industry (1 digit) x Demographic Characteristics x Indicator

State x WIB of Residence x Industry (4 digit) x Demographic Characteristics x Indicator

Run 2 (Delivered after Run 1 and as soon as CPU cycles, disk space, etc. permit):
1. State x County of Work x Industry (4 digit) x Demographic Characteristics x Indicator
2. State x County of Residence x Industry (1 digit) x Demographic Characteristics x Indicator
3. State x County of Residence x Industry (4 digit) x Demographic Characteristics x Indicator

Run 3 (Delivered after Run 2 and soon as CPU cycles, disk space, etc. permit):
1. State x MSA of Work x Industry (4 digit) x Demographic Characteristics x Indicator
2. State x MSA of Residence x Industry (1 digit) x Demographic Characteristics x Indicator
3. State x MSA of Residence x Industry (4 digit) x Demographic Characteristics x Indicator




*Version 3.3*

1. State x Commuting Distance by Geo: WIB, County, MSA

2. State x Traffic Analysis Zone (TAZ) of Residence x Industry (4 digit) x Demographic Characteristics x Indicator—IL and FL only. 1 digit only for BTS, reduced list of Indicators.

3. State x Traffic Analysis Zone (TAZ) of Work x Industry (4 digit) x Demographic Characteristics x Indicator– IL and FL only. 1 digit only for BTS, reduced list of Indicators.

4. State x Block of Work x Industry (4 digit) x Demographic Characteristics x Indicator

5. State x Block of Residence x Industry (4 digit) x Demographic Characteristics x Indicator

Note: These tabulations will be completed in the Spring of 2003 as special projects with the states of Florida and Illinois in cooperation with the Bureau of Transportation Statistics

## C. Entity Demography Editing (Successor/Predecessor Firm Research)

The LEHD staff have returned historical successor/predecessor files to Florida, Illinois, California, Maryland and Texas for research and comment. Florida has also permitted LEHD to release the files to BLS for review and comment as well. Most states have indicated they would prefer to work with more recent files—so look for the 2000 and 2001 data to be returned in early spring.

## D. Cornell Support Site

***"The Cornell Site provides a new approach to working with micro-data that will not disclose the identity of any individual person or business in the dataset"***

Several states have requested access to the micro-data derived from the integration of Census and state data. Although the Census Bureau is not permitted by law to release data that could identify any individual or business, an important part of its mandate includes providing public use files—indeed, it was the first statistical institute in the world to do so. The public use micro-data files for the Decennial Census (PUMS) and the CPS are examples of this. The National Science Foundation funding which helps to support the LEHD program has been focused on the development of access to simulated data which can generate results similar to those derived using actual data. Those of you interested in reading up on this research should read John Abowd and Simon Woodcock's chapter in the recent book **Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies**, edited by Pat Doyle, Julia Lane, Jules Theeuwes and Laura Zayatz, Elsevier Science, 2001. This research has resulted in the development of the Cornell Support site. This site provides a new approach to working with micro-data that will not disclose the identity of any individual person or business in the dataset.

The Cornell University simulated LEHD site is now running in test mode. All of the data from the Quarterly Workforce Indicators V2.3 have been placed on the simulator in folders reserved for individual states. The simulated micro data will be created from the QWI v3.0 files, which are still being created at Census.

The QWI V2.3 data can be used now with all of the software tools that the LEHD staff use in Suitland. These include SAS, Stata, MatLab, Gauss, the Microsoft Office suite, and Scientific Workplace (LaTex). We are in the process of porting LEHD programming to the simulated site.

A few states have requested early access to the simulator and those accounts have been created. We appreciate the help these early users are providing in identifying parts of the site that need improvement.

## II LEHD Infrastructure

### A. Individual Characteristics, Employer Characteristics, and Work Histories.

The LEHD Program, as a part of its Title 13 mandate and under NSF, NIA and the Sloan Foundation support, creates, maintains and enhances data products that permit the integration of Census Bureau demographic products (surveys like the SIPP, CPS, and ACS), administrative data (Federal tax information, state UI wage records, ES-202 records) and Census Bureau economic data (Business Register, Economic Censuses, and Economic Surveys). These data products are then used directly to improve the Census Bureau's demographic censuses and surveys and the economic census and surveys.
Availability: ongoing.

### B. SIPP, CPS, SSA Integration.

Continuing Title 13 research projects study the quality of the administrative data integration into the SIPP and CPS.

Availability: report available upon request

### C. Business Register, Economic Censuses and Survey Integration.

Continuing Title 13 research projects study the quality of the integration of the business units defined by the Business Register, Economic Censuses, and Economic Surveys with the Census Bureau's demographic products. This research involves the testing of different methods of exact (identification number) integration and statistical integration (using name and address information). UI and ES-202 data to firm-level data files collected by the Census Bureau. The business files we integrate include: the Economic Censuses collected every five years, the Business Register, and data collected from various less extensive establishment surveys such as the Business Expenditure Survey (which collects detailed capital spending data among non-manufacturers). In addition, we have integrated in Compustat data on large employers.

As part of this work, we have developed new measures of human capital that can be added to the Business Register. Measuring such intangibles has been an important challenge for the federal statistical system, particularly given the advent of the New Economy. Our new measures of human capital can be introduced into firm-level production functions. It is this work that forms the basis of our partnership with the Bureau of Labor Statistics.

***"Our new measures of human capital....form the basis of our partnership with the Bureau of Labor Statistics"***

## III. Special Projects

### A. Low Wage Work

We provided you an interim report on the low wage work project in August, and sent state specific tables in September. This represents the midpoint of the two year project We are currently working with North Carolina to convert the report into "WIB friendly" format. The main results are:

- A significant fraction of prime-age adults with regular labor force attachment have very low earnings (i.e., $12,000 per year or less) that persist over a period of at least three years;

2. These low earnings are associated both with their own demographic characteristics (i.e., race/gender and where they were born) and many characteristics of the firms for which they work (i.e., industry, size, turnover and net employment growth rates, and firm wage premia);

3. Of those with persistently low earnings, nearly half manage to escape this status in subsequent years, though earnings improve only partially for most of them (i.e., they continue to earn less than $15,000 in at least some years);

4. Of those with persistently low earnings, white males enjoy the highest subsequent earnings gains and highest rates of "escape" from this status of any race/gender group, while blacks endure the lowest improvements;

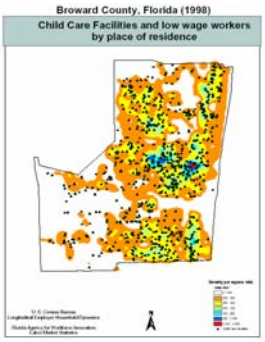
5. Job and industry changes are associated with large percentages of the observed improvements in earnings, though a significant fraction (i.e., roughly a fourth to a third) of all escapes from low-earning status also occur among those who stay on initial jobs;

6. Most earnings improvements for low-earning women occur within the service sector— in areas such as financial services, health care and education - while a larger fraction of those for males occur in the "traditional industries" like construction, manufacturing, transportation and wholesale trade;

7. Significant parts of the lower subsequent earnings of black and other (mostly Hispanic) males among initial lower earners are accounted for by their lesser access than white men to high-quality jobs;

8. Temporary help agencies are associated with lower pay for low earners while they work for them but higher subsequent wages and better job characteristics afterwards

***"A significant portion of prime-age adults have very low earnings that persist over a period of at leat three years...Temporary help agencies are associated with..higher subsequent wages and better job characteristics"***




These findings have some important implications for the low-wage labor market. For one thing, *some degree of upward mobility for persistently low earners is certainly possible*, and in fact is being achieved – even if these improvements remain fairly modest in most cases. Also, there is *no single path* for achieving earnings growth. Job changes are important to many who achieve earnings improvements, though staying on the job also works in a significant percentage of cases. What matters most is not job mobility *per se* but whether or not the individual ends up in a good job, either with or without an intervening job change.

A range of characteristics also seems to be associated with these good jobs – including not only firm wage premia (which are not observable to workers or labor market practitioners) but also industry, firm size, rates of turnover and employment growth (which are observable). Thus, it is useful to try placing low earners into high-wage sectors, firms with low turnover, and larger firms that provide job ladders and possibilities of upward mobility. The fairly positive results observed here for low earners who have worked with temp agencies might also lead us to suggest that more workers should work with such agencies, or at least with some type of labor market intermediary organization. Of course, any such recommendation is subject to the strong caveat that these agencies may work for some but not for others, and that those for whom they are successful may already be self-selecting into them. On the other hand, the results do provide some useful labor market information for intermediaries that are working with low earners, and they are supportive of the ongoing efforts of temporary help agencies with their current workforces.

The results also suggest a strong need to improve access to good jobs for many low earners – especially those who are not white males. Unfortunately, this analysis provides no direct evidence on what limits access to such jobs for many groups The results here do suggest that efforts by labor market intermediaries and other policymakers to reduce these barriers and improve access to better jobs for blacks could bear important fruit in labor market outcomes for these low-earning groups.

Report (by Anderson, Holzer and Lane) available upon request.

***"We have matched (form 5500) files to Illinois, Texas, Maryland, North Carolina and Florida data for 1996-98"***

## B. The Demand for Older Workers

The focus of this project has been to integrate form 5500 file information, with firm provided pension and health benefit information, to the Census Business Register and determine the quality of the links to the ES202 data. The files have been matched to Illinois, Texas, Maryland, North Carolina and Florida data for 1996 –1998. Preliminary analysis of the states for which we have usable data show the following match rates:


Preliminary Match Rate between the 5500 file and the Census Business Register

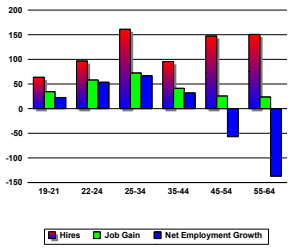
| | Number of Businesses | Number of Employees |
|---|---|---|
| Florida | 10.66% | 39.48% |
| Illinois | 16.44% | 48.00% |
| Maryland | 16.08% | 41.08% |
| North Carolina | 14.32% | 44.93% |
| Texas | 10.66% | 38.86% |

Not surprisingly, the match rates are very low for small firms, and much higher for big firms. Our current research examines the quality of the links for small firms, compare workforce composition and turnover of firms that offer coverage vs. those that don't.

### C. Sloan Foundation Workshop and Grant Proposal

The proposal was submitted July 15. The decision will be made October 8. We were pleased with the referees' reactions—one of whom said

"In summary, this is one of the most outstanding research efforts in empirical economics and provides an opportunity for an investment by the Sloan Foundation that would have a major impact on economics as a discipline".




## IV. Demographic Survey Improvements

### A. SIPP and CPS Earnings Improvements

The March Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) produce different information about annual wages. An analysis of the March CPS distribution shows a large number of high wage earners, and fewer low wage earners; the SIPP shows the opposite. LEHD staff use exactly-matched Detailed Earnings Records (DER) from the Social Security Administration to compare actual earnings information for respondents in the two surveys. Findings include the following.

1. Respondent to the March CPS and SIPP differ little in their true wage characteristics.
2. March CPS wages are typically underestimated by respondents who have one job and relatively low hours per week.
3. Respondents to the March CPS have a higher level of "underground" wages than do respondents to SIPP. This trend increased in the 1990s.
4. Respondents to the March CPS have a higher level of self-employment income "misclassified" as wages than SIPP. This trend increased in the 1990s.

These trends may explain one-third of March CPS's 6-percentage-point increase in aggregate wages relative to independent estimates from 1993 to 1995. .

***In summary, this is one of the most outstanding research efforts in empirical economics and provides an opportunity for an investment by the Sloan Foundation that would have a major impact on economics as a discipline*** *S loan proposal reviwer*

### B. SIPP/SSA/CBO Public Use Data Project


LEHD's Title 13 mandate includes researching the feasibility of creating a public use file that combines some SIPP variables with federal information from the Social Security Administration on employment histories and earnings from the SSA master earnings and benefit files. This research is being done in collaboration with the SIPP branch at the Census Bureau SSA, and the CBO. A "gold standard" file has been created for SSA and CBO to use in identifying both the key variables that are needed to do retirement and disability research and in validating the quality of the public use file. John Abowd, Martha Stinson and Julia Lane presented the preliminary work at a seminar at SSA in September.

## V. Wage Record Editing

An update to Texas edited files was shipped, for the first time implementing an update mechanism that greatly reduces the amount of data that needs to be shipped. North Carolina shipment was also sent to the state LMI. In each case, the shipment is accompanied by example decryption and readin programs, as well as with the offer of extensive help if needed. Pennsylvania wage record editing was completed internally and is in QA. Shipment occurred in the second half of September, as usual on CDROMs. Turnaround time from the moment the data set is complete to shipment has been greatly improved.

Given the size of initial shipments, and the time it takes to create CDROMs, we are currently evaluating the possibility of shipping on DVD. As an example, in compressed ASCII, 5.7 GB need to be burned to approximately 10 CDROMs for Pennsylvania. This would fit on two DVDs. Alternatively, the SAS-formatted data for the same state, ready to use (rather than requiring time-consuming readin), requires approximately 27 GB - which would require more than 40 CDROMs, but only 6 DVDs. A sample draft of the impact of wage record editing on the QWI and other, lower-level statistics is expected at the end of October.

***"We find that UI counts exceed Census counts in Agriculture, Wholesale Trade and Professional Services. We find Census counts exceed UI counts in Transportation and Other Services"***

## VI. Immigration/Population Estimates

Our ongoing work is identifying many interesting similarities and differences between UI and Census data.We find that UI counts exceed Census counts in Agriculture, Wholesale Trade and Professional Services. We find Census counts exceed UI counts in Transportation and Other Services. Our ongoing work requires the completion of the QWI research into place of work and industry, which will enable us to compare UI data with Economic Census, as well as Decennial (population) census results. In the interim, we are generating the following tables for each partner state:

Table 1: Worker Totals by Year for Native Born Americans and Immigrants (broken out by top five immigrant groups separately)

Table 2: Workers by Country of Origin for the latest year available

Table 3: Distribution of Employment by Industry for Native Born Americans, Immigrants and the top five immigrant groups

Table 4: Native and Immigrant Employment by County plus top five groups breakout.

Table 5: Native and Immigrant Employment by Fim Size Class with Breakout for top five groups

Table 6: Labor Attachment and Earnings for Natives, Immigrants (with Breakout for top five groups)

Availability: October 2002

## VII. Geocoding

The geocoding team has produced a prototype Address Master File (AMF) from the Census Bureau's Master Address File and business register and the state ES-202 files. LEHD's Address Master File contains unique residential and commercial addresses located at latitude-longitude coordinates and geocoded mostly to the Census Tract level. The file supports all of LEHD's statistical programs, including the Quarterly Workforce Indicators, and research on low-wage workers' distance to work. Current and future development of the file will geocode the addresses to the Census Block level.

Nicole Nestoriak is creating a longitudinal employer geography file that contains the best available geocodes for each reporting unit in the quarterly ES-202 data. This file will be used to enhance the employer characteristics that are used in the Quarterly Workforce Indicators and other LEHD products. It is based on a complex search algorithm applied to the AMF.

## Staff Notes

Erika McEntarfer defended her dissertation at Virginia Polytechnic Institute and began working at LEHD in August. She has been working on the LEHD website documentation, updating the QWI employment history data, cleaning up new IRS pension distribution data, and QWI version 3 documentation.

Martha Stinson defended her dissertation at Cornell University and will stay at LEHD. Paul Lengermann defended his at University of Maryland, and has started at the Federal Reserve Board of Governors (but will stay involved with LEHD).

Work has progressed in extending displaced worker analysis to other states. The presentations made at international conferences in Germany, France, Canada, and at the NBER in Boston were very favorably received.

PRESENTATIONS:

Lars Vilhuber: "The Creation of the Employment Dynamics Estimates". Presented at the 2002 NBER Summer Institute, Workshop on Empirical Personnel Economics."Displaced workers, early leavers, and re-employment wages". Presented at McGill University; First IZA/SOLE Transatlantic Meeting of Labor Economists, the 10th International Conference on Panel Data, the 14th Conference of the European Association of Labour Economists.

Simon Woodcock: "Modeling labor markets with heterogeneous agents and matches" at the Annual Meetings of the Canadian Economic Association. And Cornell University

Fredrik Anderson “Is it where you live or where you work?” The Urban Institute, and the European Association of Labor Economists.

Harry Holzer “The Interactions of Workers and Firms in the Low-Wage Labor Market” Georgetown University, Economic Policy Institute and University of Maryland

Marc Roemer “Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation” American Statistical Association's Joint Statistical Meetings

Martha Stinson “"Estimating the Relationship between Employer-Provided Health Insurance, Worker Mobility, and Wages" presented at the 10th International Conference on Panel Data, Berlin, Germany.b

**Bureau of the Census**

LEHD Program
US Census Bureau

♦ 4700 Silver Hill Road

Washington DC 20233

Phone: 301-763-8303
Fax: 301-457-8430
Email:
Ronald.c.prevost@census.gov
jlane@ui.urban.org

john_abowd@cornell.edu

LEHD

## VIII. Proposed Production Schedule

**2003 Production Schedule**
**(States Participating in the LEHD Program in 2002)**

| **Quarterly Workforce Indicators (QWI)** | | | |
|---|---|---|---|
| Transaction End Date | State Data Shipment UI Wage and ES-202 | PRED Processing | LEHD Processing/ Return to States |
| September 30, 2002 | April 1, 2003 | May 1, 2003 | June 30, 2003 * |
| Dec 31, 2002 | July 1, 2003 | August 1, 2003 | September 31, 2003 ** |
| March 31, 2003 | October 1, 2003 | November 1, 2003 | December 31, 2003 |
| June 30, 2003 | Jan 1, 2004 | Feb 1, 2004 | March 30, 2004 |

* This processing cycle represents the first round of our production operations and reflects the timeframe agreed to in the MOUs. It assumes that the wage record and ES202 data are shipped in the file format specified in the MOUs (otherwise substantial processing delays will occur). LEHD will deliver QWI products (Version 3.0) in Fall 2002, and Version 3.1 in December 2003. This schedule allows us to implement the Change Management Process (below) in November 2002 and every subsequent November.

** This production cycle will include:

1. Methodological updates (jointly determined see the Change Management Process described below)
2. Updates for all annual files (Census Numident/Person Characteristic File and residential address information)
3. Predecessor/Successor Updates
4. Revised Work Histories based on annual Wage Record Edit updates. This once-per-year update provides the opportunity to incorporate past wage record information into the edit process and finally into the Quarterly Workforce Indicators. Limitations of our existing production environment preclude our ability to incorporate time series edits every quarter and produce the Quarterly Workforce Indicators at the same time.

| **Successor/Predecessor Files** | | | |
|---|---|---|---|
| Transaction End Date | State Data Shipment | PRED Processing | LEHD Processing/ Return to States |
| Dec 31, 2002 | July 1, 2003 | N/A | December 31, 2003 |
| **Wage Record Editing** | | | |
| Transaction End Date | State Data Shipment | PRED Processing | LEHD Processing/ Return to States |
| Dec 31, 2002 | July 1, 2003 | August 1, 2003 | December 31, 2003 |

| **Annual Methodological/Operational Updates (Change Management Process)** | |
|---|---|
| Proposals submitted (LEHD/State Partners) | November 1 |
| Feasibility review (LEHD/State Partners) | December 31 |
| Joint discussion/agreement on modifications | January 30 (Annual Meeting) |
| Implementation testing completed (LEHD) | April 1 |
| Production testing completed (LEHD) | July 1 |

## IX. Website, Data Delivery System and Bulletin Board

The website and data delivery system are still under construction. The ETA funding will enable us to add the mapping component you've all been asking for, so please keep checking the site and giving us your much needed feedback

The LEHD home page includes a link to a LEHD bulletin board called SiteScape, which should help us to work more collaboratively. The first page displayed by SiteScape will prompt you for a login name and a password. Both of these have been preset to your first initial plus last name (e.g. "jsmith"). After your intial login, you should change your password by clicking on "user profile" and then "modify profile" This bulletin board includes a number of discussion areas, covering broad topics, under which we can all post points of view. As with any new software, it will take some getting used to, but SiteScape does include a help facility. Please try it out today.