

---

# LONGITUDINAL EMPLOYER - HOUSEHOLD DYNAMICS

---

## LEHD DATA DOCUMENTATION LEHD-APSHOT CES-S2004

### Snapshot CES Data Description (Nonconfidential)

---

Date	:	June 7, 2007
Prepared by	:	Kevin McKinney and Lars Vilhuber
Contact	:	U.S. Census Bureau, LEHD Program Room 6H136C 4600 Silver Hill Rd. Suitland, MD 20233 USA

---

---

This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant 5 R01 AG018854-02, and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the author(s) and do not represent the views of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers use of these data through the Research Data Centers (see [www.ces.census.gov](http://www.ces.census.gov)). For other questions regarding the data, please contact Jeremy S. Wu, Program Manager, U.S. Census Bureau, LEHD Program, Data Integration Division, Room 6H136C, 4600 Silver Hill Rd., Suitland, MD 20233, USA. <http://lehd.did.census.gov>.

# Contents

1 Overview	3
2 Updates	3
3 Treatment of Federal Tax Information	3
4 Identifiers	5
5 Availability of files	5
6 Processing files	9
7 Disclosure rules	9
8 BRB	10
9 ECF	13
10 EHF	17
11 ES202	20
12 GAL	25
13 ICF	31
14 QWI	36
15 Acronyms used	40

Table 1: LEHD components

Name and abbreviation	CES abbr. if different	Name of FTI version	CES abbreviation of FTI version
Business Register Bridge ( <a href="#">BRB</a> )		(all)	
Employer Characteristics File ( <a href="#">ECF</a> )		ECFT26	ect
Employment History Files ( <a href="#">EHF</a> )			
ES-202 ( <a href="#">ES-202</a> )	es2	ECFT26	ect
Individual Characteristics File ( <a href="#">ICF</a> )		ICFT26	ict
Geocoded Address List ( <a href="#">GAL</a> )		GALT26	gat
Quarterly Workforce Indicators ( <a href="#">QWI</a> ) (establishment level)			
Successor-Predecessor File ( <a href="#">SPF</a> )			
Unit-to-Worker Impute ( <a href="#">U2W</a> )			

## 1 Overview

The Longitudinal Employer-Household Dynamics ([LEHD](#)) Infrastructure files available in the Research Data Center ([RDC](#)) is structured as individual components. A big-picture overview of it can be found at <http://lehd.dsd.census.gov/led/library/techpapers/tp-2006-01.pdf>. Figure 1 provides an overview of the flow of data elements through the [LEHD](#) data creation process.

Currently, the core outputs of the data creation process are the Quarterly Workforce Indicators ([QWI](#)), shown in Figure 1, and the OnTheMap ([OTM](#)) data. The [LEHD](#) Infrastructure files in the [RDC](#) environment do not contain any public-use data (both the aggregated [QWI](#) and the [OTM](#) data are available to the general public), nor does it contain any information related to the disclosure-avoidance measures used in the [QWI](#) (for more information on the disclosure-avoidance techniques, see [Abowd et al. \(2006\)](#) and [Abowd, Stephens and Vilhuber \(2006\)](#) for a discussion).

## 2 Updates

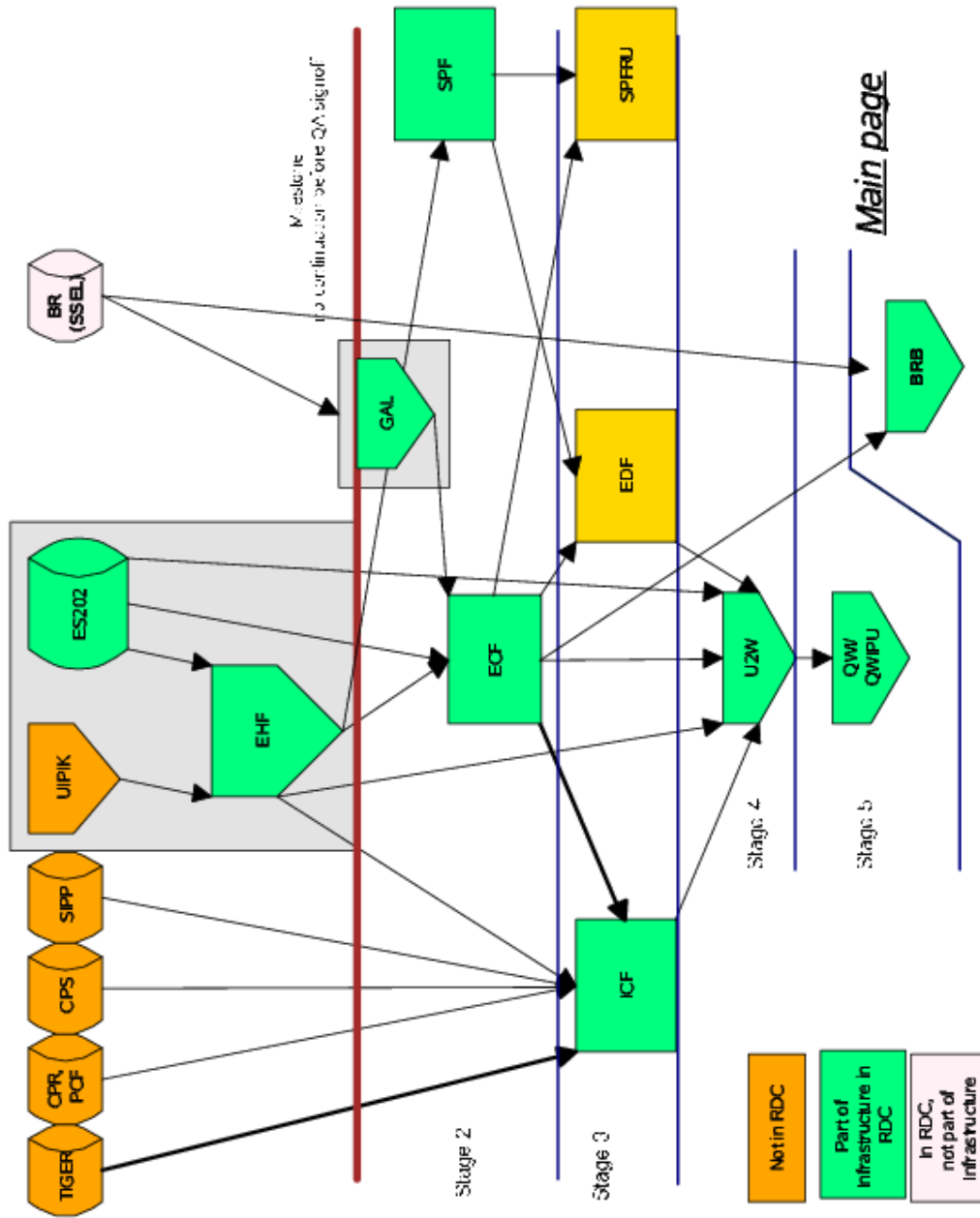
This is the first release of the [LEHD](#) Infrastructure files. It contains data that covers the years up to and including 2004Q1. We refer to it as the ‘S2004’ snapshot of the [LEHD](#) Infrastructure files. The data was pulled from [LEHD](#) archives as a coherent ensemble over the course of 2005 and 2006.

Plans are to update the [LEHD](#) Infrastructure files every two-three years. The next update is planned to occur starting in 2007Q4, containing all available states at that time in regular [LEHD](#) production, and covering years up to and including 2007Q1. Due to the post-processing required to make researcher-friendly files, and the quantity of data that needs to be moved into the [RDC](#) environment, we expect the ‘S2007’ version to become available in the Summer of 2008.

## 3 Treatment of Federal Tax Information

Some components have Title-26 protected variables, which are kept as separate components for tracking and monitoring purposes, but are not documented separately. Such T26 components need to be requested separately, and as of the writing of this documentation, will trigger additional proposal review. Table 1 shows the nine components and their Federal Tax Information ([FTI](#)) counterparts, if present, as they are available in the [RDC](#).

Figure 1: Data flow view of LEHD Infrastructure



## 4 Identifiers

In general, linkages between the different files are created using deterministic match-merge techniques. Person, firm, and establishment identifiers allow users to link all LEHD Infrastructure files. Throughout, all Social Security Numbers (SSNs) have been replaced by Protected Identity Keys (PIKs) - no SSNs are available anywhere in this data. In addition to within LEHD identifiers such as the PIK, the ICF also contains additional person identifiers linking to Census survey data: (Current Population Survey (CPS), and Survey of Income and Program Participation (SIPP)). Note that these are generally the Census-internal identifiers and may not have a direct correspondence to the identifiers on the public-use files.

Firm identifiers are called State employer identification numbers (SEINs) and generally reflect an entity reporting unemployment insurance (UI) taxes to state authorities. “Establishments” (more precisely: reporting units) are identified by SEIN reporting unit (SEINUNIT). Establishments and firms are structured as one would expect with establishments listed hierarchically within each firm. Therefore to uniquely identify an establishment both the SEIN and SEINUNIT must be used. The firm and establishment identifiers are state and firm-structure-specific - within the LEHD Infrastructure files, there is no straightforward method of linking units of a firm with multiple tax reporting entities (SEINs). Although the vast majority of firms have only one SEIN, a firm, depending on its structure may have multiple SEINs operating both within and across state boundaries. Although the federal Employer Identification Number (EIN) is available and can be used to link SEINs within and across states, the EIN suffers from similar problems as the SEIN. The identifier is not necessarily unique within a firm, is designed for tax reporting, and the structure of EINs within a firm is arbitrary. The Census Bureau recognizes the limitations of administrative identifiers and has addressed this problem on the Business Register (BR). The BRB is used to link to the Business Register (BR) and other Census economic data. Note that the BRB is a many-to-many link file.

It is also possible to use firm names (available on the ES202 component) and addresses (available with their geocodes in the GAL) to link to external data. In particular, the GAL contains rooftop coding of addresses (where available), but also contains standardized address components, facilitating further downstream matching exercises.

For any further information, refer to the component-specific documentation.

## 5 Availability of files

In general, LEHD Infrastructure files are available from 2000 onwards. However, the availability of historical data prior to 2000 varies significantly across states. At the time of writing of this guide, 31 states were available in the RDC-accessible LEHD Infrastructure. GAL files are available for all states, but certain crosswalks do not exist if ES-202 records were not available for that state. Table 2 tabulates the availability by component and state in the S2004 snapshot.

Table 2: Data availability, by state and process

	Data set group										
	ecf	ecft26	ehf	es202	gal	galt26	icf	icft26	qwi	spf	u2w
Covered states	62	3	219	1,604	882	714	31	31	28	95	32
	(227.54)	(9.91)	(408.92)	(72.63)	(65.94)	(14.59)	(25.16)	(15.93)	(652.74)	(72.00)	(41.28)
Alaska	.	.	.	.	9	14	.	.	.	.	.
					(0.12)	(0.04)					
Alabama	2	.	7	18	14	14	1	1	.	3	1
	(1.65)		(2.52)	(0.68)	(1.02)	(0.24)	(0.37)	(0.24)		(0.60)	(0.45)
Arkansas	2	.	7	11	13	14	1	1	1	3	1
	(0.44)		(1.02)	(0.26)	(0.57)	(0.15)	(0.20)	(0.13)	(4.08)	(0.25)	(0.17)
Arizona	.	.	.	.	8	14	.	.	.	.	.
					(0.99)	(0.25)					
California	2	3	7	57	24	14	1	1	1	3	1
	(45.20)	(9.91)	(74.48)	(14.79)	(8.23)	(1.52)	(4.24)	(2.61)	(87.68)	(12.71)	(8.13)
Colorado	2	.	7	60	25	14	1	1	1	3	1
	(6.72)		(14.62)	(1.81)	(1.26)	(0.25)	(0.84)	(0.53)	(42.76)	(2.47)	(1.29)
Connecticut	.	.	.	.	8	14	.	.	.	.	.
					(0.58)	(0.18)					
District of Columbia	.	.	.	.	8	14	.	.	.	.	.
					(0.12)	(0.04)					
Delaware	2	.	7	33	18	14	1	1	1	3	1
	(0.64)		(0.89)	(0.20)	(0.19)	(0.05)	(0.10)	(0.06)	(1.05)	(0.19)	(0.06)
Florida	2	.	7	66	25	14	1	1	1	3	1
	(21.81)		(38.31)	(6.75)	(4.74)	(0.96)	(2.27)	(1.43)	(36.12)	(6.96)	(4.13)
Georgia	.	.	.	.	17	14	.	.	.	.	.
					(2.10)	(0.44)					
Hawaii	.	.	.	38	20	14	.	.	.	.	.
				(0.53)	(0.29)	(0.06)					
Iowa	2	.	7	60	25	14	1	1	1	3	1
	(4.14)		(3.13)	(1.15)	(0.76)	(0.17)	(0.31)	(0.20)	(14.97)	(0.61)	(0.52)
Idaho	2	.	7	57	24	14	1	1	1	3	1
	(1.96)		(3.73)	(0.62)	(0.36)	(0.08)	(0.22)	(0.14)	(15.41)	(0.62)	(0.31)
Illinois	2	.	7	61	25	14	1	1	1	3	1
	(15.98)		(32.66)	(4.40)	(3.22)	(0.61)	(1.73)	(1.10)	(31.16)	(5.36)	(2.55)
Indiana	2	.	7	30	17	14	1	1	1	3	1
	(3.36)		(17.02)	(1.34)	(1.35)	(0.30)	(0.85)	(0.55)	(8.05)	(2.96)	(1.05)

(cont)

Table 2 – Continued

	Data set group												
	ecf	ecft26	chf	es202	gal	galt26	icf	icft26	qwi	spf	u2w		
Kansas	2 (3.90)	.	7 (7.98)	61 (1.04)	25 (0.72)	14 (0.15)	1 (0.49)	1 (0.31)	.	3 (1.32)	1 (0.61)		
Kentucky	2 (1.22)	.	7 (4.92)	18 (0.53)	14 (0.86)	14 (0.20)	1 (0.45)	1 (0.29)	1 (2.78)	3 (1.07)	1 (0.36)		
Louisiana	.	.	.	.	8 (0.81)	14 (0.23)	.	.	.	.	.		
Massachusetts	.	.	.	.	8 (1.09)	14 (0.33)	.	.	.	.	.		
Maryland	2 (7.24)	.	7 (20.01)	62 (2.04)	25 (1.35)	14 (0.26)	1 (0.90)	1 (0.57)	1 (51.55)	3 (2.88)	1 (1.14)		
Maine	2 (1.48)	.	7 (1.89)	36 (0.49)	19 (0.36)	14 (0.08)	1 (0.15)	1 (0.10)	1 (10.24)	3 (0.35)	1 (0.19)		
Michigan	.	.	.	.	8 (1.78)	14 (0.48)	.	.	.	.	.		
Minnesota	2 (6.29)	.	7 (10.16)	43 (2.03)	21 (1.26)	14 (0.27)	1 (0.63)	1 (0.40)	1 (9.10)	3 (1.77)	.		
Missouri	2 (8.72)	.	7 (16.47)	61 (1.96)	25 (1.41)	14 (0.31)	1 (0.84)	1 (0.54)	1 (43.25)	3 (2.68)	1 (1.52)		
Mississippi	.	.	.	.	8 (0.53)	14 (0.15)	.	.	.	.	.		
Montana	2 (1.44)	.	9 (2.12)	50 (0.46)	22 (0.26)	14 (0.06)	1 (0.13)	1 (0.09)	1 (10.78)	5 (0.50)	3 (0.29)		
North Carolina	2 (10.02)	.	7 (20.51)	60 (2.46)	25 (1.97)	14 (0.44)	1 (1.21)	1 (0.77)	1 (22.07)	3 (3.63)	1 (2.87)		
North Dakota	2 (0.55)	.	7 (0.76)	25 (0.25)	17 (0.16)	14 (0.05)	1 (0.08)	1 (0.05)	1 (4.17)	3 (0.15)	1 (0.08)		
Nebraska	.	.	.	.	8 (0.30)	14 (0.11)	.	.	.	.	.		
New Hampshire	.	.	.	.	8 (0.23)	14 (0.08)	.	.	.	.	.		
New Jersey	2 (9.58)	.	7 (11.20)	40 (2.86)	20 (2.02)	14 (0.45)	1 (0.98)	1 (0.62)	1 (51.18)	3 (2.07)	1 (1.17)		
New Mexico	2 (2.16)	.	7 (2.86)	61 (0.63)	25 (0.47)	14 (0.09)	1 (0.24)	1 (0.15)	1 (12.04)	3 (0.56)	1 (0.27)		
Nevada	.	.	.	29	17	14	.	.	.	.	.		

(cont)

Table 2 – Continued

		Data set group										
		ecf	ecft26	chf	es202	gal	galt26	icf	icft26	qwi	spf	u2w
New York	NY	.	.	.	(0.63)	(0.46)	(0.10)	.	.	.	.	.
Ohio	OH	.	.	.	.	8	14	.	.	.	.	.
Oklahoma	OK	2	.	7	25	16	14	1	1	1	3	1
Oregon	OR	2	.	7	61	25	14	1	1	1	3	1
Pennsylvania	PA	2	.	7	56	24	14	1	1	1	3	1
Rhode Island	RI	.	.	.	(3.88)	(3.00)	(0.58)	(1.50)	(0.96)	(18.57)	(4.44)	(2.04)
South Carolina	SC	2	.	7	29	17	14	1	1	.	3	1
South Dakota	SD	.	.	.	(1.32)	(1.00)	(0.22)	(0.48)	(0.31)	.	(1.05)	(0.55)
Tennessee	TN	.	.	.	.	8	14	.	.	.	.	.
Texas	TX	2	.	7	60	25	14	1	1	1	3	1
Utah	UT	.	.	.	(6.80)	(5.34)	(1.05)	(2.42)	(1.52)	(39.91)	(7.37)	(5.89)
Virginia	VA	2	.	7	41	20	14	1	1	1	3	1
Vermont	VT	2	.	7	20	15	14	1	1	1	3	1
Washington	WA	2	.	7	61	25	14	1	1	1	3	1
Wisconsin	WI	2	.	7	62	25	14	1	1	1	3	1
West Virginia	WV	2	.	7	61	25	14	1	1	1	3	1
Wyoming	WY	.	.	.	(0.71)	(0.41)	(0.08)	(0.18)	(0.12)	(10.15)	(0.37)	(0.27)
					.	8	14	.	.	.	.	.
					(0.09)	(0.04)						

Number of files for each data set group and state. Aggregate size of all files in GB in parentheses.

Files not currently available may become available in the next update to the LEHD Infrastructure (see Section 2). At the time of writing of this document, we expect the next version to contain up to 46 states. Availability of core Infrastructure files is dependent on a state’s participation in the Local Employment Dynamics (LED) program. The latest participant list can be found at the LEHD website at <http://lehd.did.census.gov>.

## 6 Processing files

LEHD Infrastructure files are significantly larger than even traditionally large research files such as the decennial census. In the current version, in all available states and years combined, information on 754,775,697 jobs is presented. There are 226,639,116 quarterly observations on firms. Careful planning is required to ensure that adequate resources are available. To facilitate researchers in this endeavor, the research versions of the LEHD Infrastructure files in the RDC environment have additional random variables that allow for the selection of uniform random subsamples of firms (SEIN), establishments (SEINUNIT), and individuals (PIK). No such random variable is available on the EHF, since there is no single good strategy for selecting jobs. Tables in the documentation for individual components also contains information about the size on-disk of each file.

## 7 Disclosure rules

Special disclosure and data use rules apply to analyses based on the micro-data from the LEHD Infrastructure file system. These data underlie the QWI, and disclosable research results are therefore subject to restrictions that insure the QWI disclosure protection mechanism is not compromised. Confidentiality protection for the QWI uses noise infusion of the micro-data. The Disclosure Review Board (DRB) does not allow the release of any tabulations for sub-state geography that do not use the QWI noise infusion process. In addition, the required noise factors have not been placed on the RDC snapshot files as part of the DRB’s normal rules limiting access to the specific parameters of its approved disclosure limitation methods. Only the DRB may approve the release of tabular output from the LEHD infrastructure file system. Sub-state geography tables will not be approved. National or multi-state tables may be approved provided they do not compromise the protection system. Model-based output is normally allowed. The chief disclosure officer for the RDC network will coordinate the reviews.

The underlying micro-data in the LEHD infrastructure file system were provided to the Census Bureau by state Labor Market Information (LMI) offices under Memoranda of Understanding (also called Data Use Agreements) negotiated with each state. This process is part of the LED federal/state partnership, and places additional restrictions on the results that may be published. Current members of the LED partnership are shown on the LEHD main web page.

Publicly disclosing a single state’s data in identifiable form requires the permission of the state’s LMI officer. When reporting results from studies that include multiple states, the results should be pooled across the states. The identity of the LED member states is obviously not confidential. You may say which states were used in your analysis. The chief disclosure officer for the RDC network will review compliance with this requirement in consultation with the Assistant Division Chief for LEHD.

## 8 BRB

**Name:** LEHD Business Register Bridge

**Code:** brb

**Group:** brb

**Origin:** Census Bureau constructed data

**Sector:** Business

**Period:** Quarterly

**Reference Period:** Three (3) Months

**Unit of Enumeration:** Combination of Establishments, Companies and/or Enterprises

**Source:** Bureau of the Census

**Administrative Data:** Partially

**Legal authority:** Title 13, Title 26 (contains FTI)

**Observations:** See Table 3.

File	Number of datafiles	Records
brb.us.xwalk	1	127,851,774
brb.us.ecflist	1	179,201,161
brb.us.brlist	1	143,701,504

Table 3: Number of observations for BRB

**Industry:** All

**NAICS coverage:** SIC only

**Documentation:** <http://www.census.gov/econ/overview/mu0600.html> and [http://lehd.dsd.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.dsd.census.gov/led/library/tech_user_guides.html)

**Short Description:** The LEHD employer-level data are stored at both an establishment and firm level and can be linked to Census Bureau establishment and firm level microdata (e.g. Economic Censuses and Surveys) via Business Register Bridge (BRB). The bridge provides a cross walk at various levels of business-unit aggregation. The most detailed exact crosswalk is at an EIN, State, 4-digit industry, county level. The bridge includes the full list of establishments in the LEHD data and in the Business Register that are associated with the business units (e.g., EIN/4-digit SIC/State/County) in the crosswalk and measures of activity (e.g., employment, sales).

**Long Description:** One of the many advantages of the LEHD database is the ability it offers to link information from Census Bureau business-level surveys (such as the Annual Survey of Manufacturers (ASM), or Business Expenditure Survey (BES)) to the set of workers employed by these units. The LEHD Business Register Bridge (BRB) provides researchers with the tool to make this linkage. The primary bridge between the LEHD data and the business data is the U.S. Census Bureau's annual Business Register (BR), a list of establishments the Bureau uses to develop the initial mailing list for the economic censuses and surveys. The BR contains data from several different sources. Primarily, however, the BR contains very reliable information on business identifiers, business organizational structure, and business location. All of this detail helps us to form links to the LEHD data. Unfortunately, we are not able to directly form establishment-to-establishment linkages between LEHD data and the Business Register. The establishment identification system for the BR is the same as for all other Census Bureau business data products but different from the LEHD establishment identifier. Unfortunately, there is no one best way to form linkages between these data sources. There are many alternatives, and the optimal linking strategy depends on the research objective. To provide researchers with as much flexibility as possible, the BRB has been constructed as a crosswalk that allows for a number of different ways to integrate these data sources.

**Limitations:** The Business Register and the EHF/ECF represent two different attempts to build a universe of active firms. Due to timing differences, the coverage of the two data sources will differ. Also, collection and reporting methods also play a role in the coverage available at any point in time. Differences in the two universes should therefore be expected and the implications of these differences should be taken into account when using the BRB.

**Coverage:** All sectors covered on the state's ECF and the Business Register.

**FTI:** All BRB data is considered FTI-commingled data.

File name	Variable	Field name
brb_us_ecflist.sas7bdat	EIN	Federal Employer ID Number (commingled)
brb_us_brlist.sas7bdat	All	All data sourced from BR
brb_us_xwalk.sas7bdat	All	All data links to BR

\$Id: brb\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 9 ECF

**Name:** LEHD Employer Characteristic File

**Code:** ecf/ect (for FTI)

**Group:** ecf/ect (for FTI)

**Origin:** Multiple

**Sector:** Business

**Period:** Quarterly

**Reference Period:** Three (3) Months

**Unit of Enumeration:** Combination of Establishments, Companies and/or Enterprises

**Source:** State Employment Security Agencies, IRS, other

**Administrative Data:** Yes

**Legal authority:** Title 13, U.S.C. except the variable 'EIN' on California files, which is Title 26, U.S.C.

**Observations:** See Table 4.

Table 4: Number of observations, by state, by group, for ECF

State	Group		Number of datafiles	Records	Filesize (MB)
Covered states			64	587,237,341	240,012.05
Alabama	AL	ecf	2	3,426,165	1,685.48
Arkansas	AR	ecf	2	954,157	452.49
California	CA	ecf	2	99,129,182	46,282.81
California	CA	ecft26	2	99,129,182	7,009.10
Colorado	CO	ecf	2	14,146,330	6,879.70
Delaware	DE	ecf	2	1,416,157	652.81
Florida	FL	ecf	2	47,026,785	22,337.79
Iowa	IA	ecf	2	9,009,472	4,238.50
Idaho	ID	ecf	2	4,144,272	2,002.11
Illinois	IL	ecf	2	34,401,303	16,368.45
Indiana	IN	ecf	2	7,052,084	3,441.49
Kansas	KS	ecf	2	8,318,828	3,989.95
Kentucky	KY	ecf	2	2,601,098	1,250.75
Maryland	MD	ecf	2	15,154,168	7,418.82
Maine	ME	ecf	2	3,224,812	1,520.61
Minnesota	MN	ecf	2	12,615,016	6,443.17
Missouri	MO	ecf	2	18,680,950	8,929.59
Montana	MT	ecf	2	3,129,539	1,470.77
North Carolina	NC	ecf	2	21,788,368	10,261.29

( cont)

Table 4 – Continued

State	Group		Number of datafiles	Records	Filesize (MB)
North Dakota	ND	ecf	2	1,185,496	563.73
New Jersey	NJ	ecf	2	20,480,777	9,808.47
New Mexico	NM	ecf	2	4,869,480	2,209.05
Oklahoma	OK	ecf	2	3,836,973	1,809.45
Oregon	OR	ecf	2	11,116,430	5,359.66
Pennsylvania	PA	ecf	2	30,066,151	14,616.05
South Carolina	SC	ecf	2	5,673,122	2,835.70
Texas	TX	ecf	2	51,261,545	24,614.88
Virginia	VA	ecf	2	11,760,264	5,676.17
Vermont	VT	ecf	2	767,752	361.20
Washington	WA	ecf	2	20,776,265	9,895.98
Wisconsin	WI	ecf	2	15,102,565	7,288.20
West Virginia	WV	ecf	2	4,992,653	2,337.84

**Industry:** Combination (industries)

**NAICS coverage:** All

**Documentation:** [http://lehd.dsd.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.dsd.census.gov/led/library/tech_user_guides.html)

**Short Description:** The Employer Characteristics File (ECF) consolidates most firm level information (size, location, industry, etc.) into two easily accessible files. The firm or SEIN level file contains one record for every YEAR QUARTER a firm is present on either the ES-202 or the UI, with more detailed information available for the establishments of multi-unit firms in the SEIN SEINUNIT file. The SEIN file is built up from the SEINUNIT file and contains no additional information, but should be viewed merely as an easier and/or more efficient way to access firm level data.

**Long Description:** The primary data source for the Employer Characteristics File (ECF) are the ES-202 files. The ES-202 files contain information at both the firm and establishment level and the ECF continues this structure, but the "master" records are removed and the firm level information is appended to the establishment data. The next major source of data comes from the UI/EHF. Person level job information is aggregated to the firm level and is merged with the ES-202 information. The union of these two data sources determines the set of records on the ECF. The ES-202 data contains among other variables, information on establishment location, Federal EIN, and industry. Any missing establishment characteristics are filled with information from the closest quarter, the firm level mode, and eventually imputed, if no firm specific information is available. The UI/EHF and the ES-202 both contribute information on payroll and employment. Various algorithms are used to combine this information in the most sensible way possible. Finally, the address information available on the ES-202 is processed through an address standardization and Geographic Information System (GIS) processing system to create the Geocoded Address List (GAL). The resulting establishment level geocode information from the GAL is then merged back onto the ECF.

**Limitations:** The quality of the ECF depends primarily on the quality of the incoming UI and ES-202 data as received from the state. The quality of the data is generally high for 2000 and beyond, but earlier years' historical data files may be incomplete. Although the data undergoes various quality control checks and imputations of missing values once received at LEHD, in some cases this is only a partial solution. States, also vary in the quality of the establishment level data reported for multi-unit firms. The lack of consistent reporting of establishment level information can greatly affect the geographic dispersion of employment and

State		Start	End
Covered states		1989Q1	2004Q3
Alabama	AL	2001Q1	2004Q2
Arkansas	AR	2002Q3	2004Q1
California	CA	1991Q1	2003Q4
Colorado	CO	1990Q1	2004Q1
Delaware	DE	1997Q1	2004Q1
Florida	FL	1989Q1	2004Q1
Iowa	IA	1990Q1	2004Q1
Idaho	ID	1991Q1	2004Q1
Illinois	IL	1990Q1	2004Q1
Indiana	IN	1998Q1	2004Q1
Kansas	KS	1990Q1	2004Q1
Kentucky	KY	2001Q1	2004Q1
Maryland	MD	1990Q1	2004Q1
Maine	ME	1996Q1	2004Q1
Minnesota	MN	1994Q3	2004Q3
Missouri	MO	1990Q1	2004Q1
Montana	MT	1993Q1	2004Q1
North Carolina	NC	1990Q1	2003Q4
North Dakota	ND	1998Q1	2004Q2
New Jersey	NJ	1995Q1	2004Q1
New Mexico	NM	1990Q1	2004Q1
Oklahoma	OK	1999Q1	2004Q1
Oregon	OR	1990Q1	2004Q1
Pennsylvania	PA	1991Q1	2004Q1
South Carolina	SC	1998Q1	2004Q3
Texas	TX	1990Q1	2004Q2
Virginia	VA	1995Q3	2004Q1
Vermont	VT	2000Q1	2004Q1
Washington	WA	1990Q1	2004Q1
Wisconsin	WI	1990Q1	2004Q1
West Virginia	WV	1990Q1	2004Q1

Table 5: Coverage, by state, for ECF

payroll. In both 1992 and 1997 efforts were made to expand the collection of establishment level data. Expansions in the number of establishments in those years are likely not related to new economic activity, but reflect improved collection activities.

**Coverage:** Many states have only partial coverage of sectors such as agriculture, state government, and local municipalities. Federal government workers are also not typically covered. Due to limitations in the historical data, some years and quarters may not have complete coverage on both the EHF and ES-202 and this will be reflected in the ECF as well.

See Table 5.

**FTI:** The following files contain FTI:

File name	Variable	Field name
ecf_zz_sein_t26.sas7bdat	MODE_ES_EIN_EMP	Emp Mode Cleaned EIN
	QUARTER	Quarter QQ
	SEIN	State Employer ID Number
	MODE_ES_EIN	Unit Mode Cleaned EIN
ecf_zz_seinunit_t26.sas7bdat	ES_EIN_MISS	0=ok,1=not found,2+found off qtr
	ES_EIN	Cleaned EIN
	VALID_EIN	EIN in known IRD
	MODE_ES_EIN_EMP	Emp Mode Cleaned EIN
	EIN_BAD	Letters a-z,A-Z in EIN
	EIN	Original ES202 EIN
	EIN_DEFECT	Problem with EIN
	ES_EIN_FLAG	Quarters Away EIN data found
	MODE_ES_EIN	Unit Mode Cleaned EIN

\$Id: ecf\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 10 EHF

**Name:** LEHD Employment History Files (EHF)

**Code:** EHF

**Group:** ehf

**Origin:** Census Bureau constructed data from data provided by State agencies

**Sector:** Multiple

**Period:** Quarterly

**Reference Period:** Three (3) Months

**Unit of Enumeration:** Multiple (Enterprise (firm) Other Sub-Enterprise Reporting Unit, Individual Person)

**Source:** Employment Security Agencies and/or Bureau of Labor Statistics (BLS)

**Administrative Data:** Yes

**Legal authority:** The underlying data is provided to the Census Bureau based on Memoranda of Understanding with individual state agencies. It is protected under Title 13, U.S.C.

**Observations:** See data documentation See Table 6.

Table 6: Number of observations, by state, for EHF

State		Number of datafiles	Records	Filesize (MB)
Covered states		217	2,838,350,408	418,570.64
Alabama	AL	7	22,419,226	2,578.44
Arkansas	AR	7	9,519,686	1,044.09
California	CA	7	514,071,832	76,263.99
Colorado	CO	7	90,966,637	14,970.98
Delaware	DE	7	7,640,788	911.26
Florida	FL	7	256,230,604	39,224.97
Iowa	IA	7	25,761,052	3,200.69
Idaho	ID	7	23,835,178	3,816.40
Illinois	IL	7	221,251,965	33,438.76
Indiana	IN	7	114,988,266	17,429.23
Kansas	KS	7	52,731,691	8,172.11
Kentucky	KY	7	42,071,468	5,037.72
Maryland	MD	7	115,760,015	20,485.74
Maine	ME	7	14,680,364	1,938.88
Minnesota	MN	7	76,484,001	10,400.03
Missouri	MO	7	106,408,441	16,866.77
Montana	MT	7	13,651,082	2,006.10

( cont)

Table 6 – Continued

State		Number of datafiles	Records	Filesize (MB)
North Carolina	NC	7	140,709,238	21,001.80
North Dakota	ND	7	6,327,983	778.26
New Jersey	NJ	7	87,722,728	11,467.09
New Mexico	NM	7	21,480,296	2,925.88
Oklahoma	OK	7	21,756,391	2,573.13
Oregon	OR	7	61,857,523	9,536.13
Pennsylvania	PA	7	188,590,880	27,341.23
South Carolina	SC	7	39,151,294	4,754.95
Texas	TX	7	269,205,201	37,549.96
Virginia	VA	7	64,226,246	7,831.45
Vermont	VT	7	4,456,337	521.20
Washington	WA	7	112,085,555	18,187.27
Wisconsin	WI	7	97,269,196	14,389.28
West Virginia	WV	7	15,039,244	1,926.86

**Industry:** Combination (industries)

**NAICS coverage:** All sectors covered with the exception of agricultural workers.

**Documentation:** [http://lehd.did.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.did.census.gov/led/library/tech_user_guides.html)

**Short Description:** The Employment History File (EHF) provides a full time-series of earnings at all within-state jobs for all quarters covered by the LEHD system and provided by the state. It also provides activity calendars at a job, firm, and sub-firm reporting unit level.

**Long Description:** The Employment History Files (EHF) are designed to store the complete in-state work history for each individual that appears in the UI wage records. The EHF for each state contains one record for each employee-employer combination—a job—in that state in each year. Both annual and quarterly earnings variables are available in the EHF. Individuals who are employed, but never have strictly positive earnings at their employing SEIN (a theoretical possibility) in a given year do not have a record in the EHF for that year. To facilitate analysis, the EHF data are restructured into another file containing one observation per job (PIK-SEIN combination), with all quarterly earnings and activity information available on that record. The restructured file is called the Person History File (PHF). It should be noted that the actual file structure is at the PIK-SEIN-SEINUNIT-YEAR level for the EHF, and at the PIK-SEIN-SEINUNIT level for the PHF. Although only one state (Minnesota) has non-zero values for SEINUNIT, this allows the file structure to be homogeneous across states. An active job within a quarter, the primary job-level economic activity measure, is defined as having strictly positive quarterly earnings for the individual-employer pair that define the job.

A time-series similar to the aggregated job data, but based on observed activity (positive employment) in the ES-202 records, is available and computed at the SEINUNIT level (Unit History File, UHF) and the SEIN level (SEIN History File, SHF).

**Limitations:** No edits have been applied to these data. Under-reporting, data loss, etc. may affect diverse states and quarters. The data documentation has an up-to-date list of such data issues. The LEHD Program suggests users use Bayesian multiple imputation models to address such issues.

State		Start	End
Covered states		1985Q2	2004Q3
Alabama	AL	2001Q1	2004Q2
Arkansas	AR	2002Q3	2004Q2
California	CA	1991Q1	2003Q4
Colorado	CO	1990Q1	2004Q1
Delaware	DE	1997Q1	2004Q1
Florida	FL	1989Q1	2004Q1
Iowa	IA	1990Q1	2004Q1
Idaho	ID	1990Q1	2004Q1
Illinois	IL	1990Q1	2004Q1
Indiana	IN	1990Q1	2004Q1
Kansas	KS	1990Q1	2004Q1
Kentucky	KY	1996Q4	2004Q1
Maryland	MD	1985Q2	2004Q2
Maine	ME	1996Q1	2004Q1
Minnesota	MN	1994Q3	2004Q3
Missouri	MO	1990Q1	2004Q1
Montana	MT	1993Q1	2004Q1
North Carolina	NC	1990Q1	2003Q4
North Dakota	ND	1998Q1	2004Q2
New Jersey	NJ	1995Q1	2004Q1
New Mexico	NM	1990Q1	2004Q1
Oklahoma	OK	1999Q1	2004Q1
Oregon	OR	1990Q1	2004Q1
Pennsylvania	PA	1991Q1	2004Q1
South Carolina	SC	1998Q1	2004Q3
Texas	TX	1990Q1	2004Q2
Virginia	VA	1995Q3	2004Q1
Vermont	VT	2000Q1	2004Q1
Washington	WA	1990Q1	2004Q1
Wisconsin	WI	1990Q1	2004Q1
West Virginia	WV	1990Q1	2004Q1

Table 7: Coverage, by state, for EHF

**Coverage:** The earliest data requested by the LEHD system are 1990, quarter 1 (for historical reasons, Maryland goes back farther). Most states provided data beginning some time in the early 1990s. All partner states provide data beginning in 1997, quarter 1. The current research files are a snapshot of the live LEHD data infrastructure, and were selected to include 2004Q1 wherever possible. See Table 7 for state-by-state details.

**FTI:** The EHF contains no FTI.

\$Id: ehf\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 11 ES202

**Name:** LEHD - ES-202

**Code:** ES2

**Group:** ES2

**Origin:** State or Local Government Agency

**Sector:** Multiple

**Period:** Quarterly

**Reference Period:** Three (3) Months

**Unit of Enumeration:** Combination of Establishments, Companies and/or Enterprises

**Source:** State Employment Security Agencies

**Administrative Data:** Yes

**Legal authority:** The data is provided to the Census Bureau under Title 13, U.S.C., except the variable 'EIN' on California files, which is Title 26, U.S.C.

**Observations:** See Table 8.

Table 8: Number of observations, by state, by group, for ES202

State	Group	Number of datafiles	Records	Filesize (MB)
Covered states		1,598	353,912,649	77,519.58
Alabama	AL es202	18	2,493,534	698.63
Arkansas	AR es202	11	828,004	265.93
California	CA ecft26	1	58,571,345	3,141.77
California	CA es202	57	58,571,345	15,144.95
Colorado	CO es202	60	8,015,661	1,848.72
Delaware	DE es202	33	839,395	205.57
Florida	FL es202	66	27,653,497	6,907.89
Hawaii	HI es202	38	1,323,395	542.73
Iowa	IA es202	60	5,391,716	1,176.81
Idaho	ID es202	57	2,446,203	635.49
Illinois	IL es202	61	19,456,582	4,507.79
Indiana	IN es202	30	4,650,236	1,371.55
Kansas	KS es202	61	4,891,097	1,062.88
Kentucky	KY es202	18	1,971,121	538.63
Maryland	MD es202	62	8,777,361	2,087.92
Maine	ME es202	36	1,894,604	504.75
Minnesota	MN es202	43	7,472,023	2,081.76
Missouri	MO es202	61	10,055,160	2,006.02

( cont)

Table 8 – Continued

State	Group	Number of datafiles	Records	Filesize (MB)
Montana	MT es202	50	1,870,577	471.75
North Carolina	NC es202	60	12,958,104	2,519.75
North Dakota	ND es202	25	627,933	254.07
New Jersey	NJ es202	40	11,558,633	2,926.42
New Mexico	NM es202	61	2,796,209	649.57
Nevada	NV es202	29	1,589,192	643.21
Oklahoma	OK es202	25	2,489,118	680.87
Oregon	OR es202	61	6,326,082	1,660.29
Pennsylvania	PA es202	56	16,900,757	3,973.09
South Carolina	SC es202	29	3,229,024	1,346.80
Tennessee	TN es202	29	4,407,825	1,743.65
Texas	TX es202	60	29,554,042	6,960.08
Utah	UT es202	55	3,195,951	1,323.85
Virginia	VA es202	41	7,306,115	1,963.57
Vermont	VT es202	20	243,439	64.53
Washington	WA es202	61	11,676,973	2,774.80
Wisconsin	WI es202	62	8,909,541	2,105.86
West Virginia	WV es202	61	2,970,855	727.62

**Industry:** Combination (industries)

**NAICS coverage:** All NAICS sectors.

**Documentation:** [http://lehd.did.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.did.census.gov/led/library/tech_user_guides.html)

**Short Description:** The ES-202 data contains firm reports of business demographic information such as industry, number and type of establishments, location, etc. These files contain detailed name and address information that is not on the Employer Characteristics File (ECF). The files also contain monthly data on payroll and employment that is reported on a quarterly basis.

**Long Description:** The ES-202 is the common name for the Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW) program. These data are the main source of firm characteristics. The primary variables of interest are location, industry, and the number and type of establishments. The ES-202 data is the only source of detailed information about multi-unit firms in the LEHD Infrastructure, which play an important role in the production of both the QWI and OTM. Without this information all employment and payroll would be clustered at one geographic location, rather than at each establishment.

**Limitations:**

- The quality of the ES-202 files depend primarily on the quality of the incoming ES-202 data as received from the state. The quality of the data is generally high for 2000 and beyond, but earlier years' historical data files may be incomplete. Although the data undergoes various quality control checks once received at LEHD, in some cases portions of the original data cannot be restored from the state's archives. States, also vary in the quality of the establishment level data reported for multi-unit firms. The lack of consistent reporting of establishment level information can greatly affect the geographic dispersion of employment and payroll. In both 1992 and 1997 the states made a coordinated effort to expand the collection of establishment level data. Expansions in the number of establishments in those years are likely not related to new economic activity, but reflect improved collection activities.

- While the UI/EHF data can be used to obtain various measures of payroll and employment, these measures are not based on employment at a specific point in time, but rather the employees with positive earnings at some time during the quarter. The ES-202 data contains the level of employment at the firm and its establishments on the 13th of each month. For firms with relatively stable employment, the two sources of data will be similar, however for firms with large numbers of accessions and separations, the two sources can differ substantially.
- Many states have only partial coverage of sectors such as agriculture, state government, and local municipalities. Federal government workers are also not typically covered. Due to limitations in the historical data, some years and quarters may not have complete coverage on the ES-202.

**Coverage:** See Table 9.

State		Start	End
Covered states		1989Q1	2005Q1
Alabama	AL	2001Q1	2005Q1
Arkansas	AR	2002Q3	2005Q1
California	CA	1991Q1	2005Q1
Colorado	CO	1990Q1	2005Q1
Delaware	DE	1997Q1	2005Q1
Florida	FL	1989Q1	2004Q4
Hawaii	HI	1995Q4	2005Q1
Iowa	IA	1990Q1	2005Q1
Idaho	ID	1991Q1	2005Q1
Illinois	IL	1990Q1	2005Q1
Indiana	IN	1998Q1	2005Q1
Kansas	KS	1990Q1	2005Q1
Kentucky	KY	2001Q1	2005Q1
Maryland	MD	1990Q1	2005Q1
Maine	ME	1996Q1	2004Q4
Minnesota	MN	1994Q3	2005Q1
Missouri	MO	1990Q1	2005Q1
Montana	MT	1993Q1	2005Q1
North Carolina	NC	1990Q1	2005Q1
North Dakota	ND	1998Q1	2005Q1
New Jersey	NJ	1995Q1	2005Q1
New Mexico	NM	1990Q1	2005Q1
Nevada	NV	1998Q1	2005Q1
Oklahoma	OK	1999Q1	2004Q4
Oregon	OR	1990Q1	2005Q1
Pennsylvania	PA	1991Q1	2005Q1
South Carolina	SC	1998Q1	2005Q1
Tennessee	TN	1998Q1	2005Q1
Texas	TX	1990Q1	2005Q1
Utah	UT	1990Q1	2005Q1
Virginia	VA	1995Q3	2005Q1
Vermont	VT	2000Q1	2005Q1
Washington	WA	1990Q1	2005Q1
Wisconsin	WI	1990Q1	2005Q1
West Virginia	WV	1990Q1	2005Q1

Table 9: Coverage, by state, for ES202

**FTI:** The following files contain FTI:

File name	Variable	Field name
es202_zz_t26.sas7bdat	EIN	Federal Employer ID Number

\$Id: es202\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 12 GAL

**Name:** LEHD Geocoded Address List (LEHD-GAL)

**Code:** gal/gat

**Group:** gal/gat

**Origin:** Census Bureau constructed data

**Sector:** Multiple: contains residential and business address data

**Period:** Irregular

**Reference Period:** None

**Unit of Enumeration:** Combination (any units)

**Source:** Bureau of the Census, IRS (partial)

**Administrative Data:** Partially

**Legal authority:** Title 13. Data sourced from the Business Register is covered by Title 26

**Observations:** See Table 10.

Table 10: Number of observations, by state, by group, for GAL

State	Group	Number of datafiles	Records	Filesize (MB)
Covered states		1,596	930,860,552	82,461.99
Alaska	AK gal	9	3,462,374	122.75
Alaska	AK galt26	14	773,213	38.89
Alabama	AL gal	14	5,345,837	1,049.09
Alabama	AL galt26	14	5,263,978	243.27
Arkansas	AR gal	13	4,329,509	583.45
Arkansas	AR galt26	14	3,309,969	155.80
Arizona	AZ gal	8	6,230,191	1,009.88
Arizona	AZ galt26	14	5,697,348	257.78
California	CA gal	24	89,406,885	8,424.98
California	CA galt26	14	35,184,893	1,551.44
Colorado	CO gal	25	13,317,745	1,295.20
Colorado	CO galt26	14	5,703,358	258.56
Connecticut	CT gal	8	3,679,103	596.46
Connecticut	CT galt26	14	3,966,579	182.10
District of Columbia	DC gal	8	744,240	125.17
District of Columbia	DC galt26	14	881,848	41.24
Delaware	DE gal	18	1,693,512	195.20
Delaware	DE galt26	14	1,082,175	49.68
Florida	FL gal	25	45,934,781	4,848.70

( cont)

Table 10 – Continued

State	Group	Number of datafiles	Records	Filesize (MB)	
Florida	FL	galt26	14	21,999,874	981.52
Hawaii	HI	gal	20	2,681,552	292.70
Hawaii	HI	galt26	14	1,430,334	66.38
Iowa	IA	gal	25	7,748,117	778.53
Iowa	IA	galt26	14	3,797,959	175.63
Idaho	ID	gal	24	3,817,233	372.04
Idaho	ID	galt26	14	1,700,756	76.90
Illinois	IL	gal	25	32,677,811	3,296.61
Illinois	IL	galt26	14	14,183,165	628.27
Indiana	IN	gal	17	11,320,005	1,385.59
Indiana	IN	galt26	14	6,879,048	309.66
Kansas	KS	gal	25	7,541,431	737.15
Kansas	KS	galt26	14	3,442,863	156.41
Kentucky	KY	gal	14	6,483,024	885.34
Kentucky	KY	galt26	14	4,605,123	209.54
Louisiana	LA	gal	8	5,198,166	830.05
Louisiana	LA	galt26	14	5,024,357	233.13
Massachusetts	MA	gal	8	6,903,936	1,117.38
Massachusetts	MA	galt26	14	7,319,221	336.63
Maryland	MD	gal	25	13,605,804	1,378.16
Maryland	MD	galt26	14	5,936,487	262.51
Maine	ME	gal	19	3,587,216	369.90
Maine	ME	galt26	14	1,800,195	86.39
Michigan	MI	gal	8	11,305,272	1,823.91
Michigan	MI	galt26	14	11,053,750	494.80
Minnesota	MN	gal	21	12,556,051	1,286.23
Minnesota	MN	galt26	14	5,954,055	272.38
Missouri	MO	gal	25	12,920,327	1,439.37
Missouri	MO	galt26	14	6,921,562	316.19
Mississippi	MS	gal	8	3,537,809	544.91
Mississippi	MS	galt26	14	3,234,169	152.98
Montana	MT	gal	22	2,882,903	270.67
Montana	MT	galt26	14	1,367,721	64.54
North Carolina	NC	gal	25	16,473,564	2,020.22
North Carolina	NC	galt26	14	9,794,912	452.47
North Dakota	ND	gal	17	1,465,941	160.56
North Dakota	ND	galt26	14	1,008,890	49.34
Nebraska	NE	gal	8	1,986,804	305.11
Nebraska	NE	galt26	14	2,263,077	108.97
New Hampshire	NH	gal	8	1,524,989	239.38
New Hampshire	NH	galt26	14	1,625,118	78.28
New Jersey	NJ	gal	20	19,940,688	2,070.72
New Jersey	NJ	galt26	14	10,187,033	461.27
New Mexico	NM	gal	25	4,726,812	479.07
New Mexico	NM	galt26	14	2,075,073	95.99
Nevada	NV	gal	17	3,764,752	470.90
Nevada	NV	galt26	14	2,320,012	104.98
New York	NY	gal	8	20,895,573	3,399.23

( cont)

Table 10 – Continued

State	Group	Number of datafiles	Records	Filesize (MB)	
New York	NY	galt26	14	22,209,741	1,020.67
Ohio	OH	gal	8	12,686,542	2,038.99
Ohio	OH	galt26	14	12,266,204	549.41
Oklahoma	OK	gal	16	6,415,794	760.61
Oklahoma	OK	galt26	14	4,127,741	192.52
Oregon	OR	gal	25	10,392,416	988.78
Oregon	OR	galt26	14	4,325,479	197.54
Pennsylvania	PA	gal	24	29,426,822	3,067.38
Pennsylvania	PA	galt26	14	13,417,762	598.33
Rhode Island	RI	gal	8	1,177,406	191.19
Rhode Island	RI	galt26	14	1,250,666	57.31
South Carolina	SC	gal	17	8,298,660	1,020.75
South Carolina	SC	galt26	14	4,905,010	224.89
South Dakota	SD	gal	8	886,997	136.88
South Dakota	SD	galt26	14	1,093,935	54.27
Tennessee	TN	gal	17	10,894,201	1,342.38
Tennessee	TN	galt26	14	6,526,381	293.77
Texas	TX	gal	25	51,771,463	5,468.65
Texas	TX	galt26	14	23,798,608	1,073.66
Utah	UT	gal	25	5,558,434	546.58
Utah	UT	galt26	14	2,363,850	106.98
Virginia	VA	gal	20	14,754,427	1,658.69
Virginia	VA	galt26	14	7,991,509	361.60
Vermont	VT	gal	15	1,416,619	164.51
Vermont	VT	galt26	14	970,783	48.47
Washington	WA	gal	25	18,224,318	1,751.94
Washington	WA	galt26	14	7,466,302	339.91
Wisconsin	WI	gal	25	14,751,335	1,511.81
Wisconsin	WI	galt26	14	6,538,919	291.65
West Virginia	WV	gal	25	4,368,446	422.19
West Virginia	WV	galt26	14	1,813,792	85.55
Wyoming	WY	gal	8	622,018	96.91
Wyoming	WY	galt26	14	745,554	36.35

**Industry:** Combination (industries)

**NAICS coverage:** All

**Documentation:** [http://lehd.did.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.did.census.gov/led/library/tech_user_guides.html)

**Short Description:**

**Long Description:**

**Limitations:** See data documentation. Not all data is available nationally.

**Coverage:** Atemporal. Input data can go back to 1990.

**FTI:** The following files contain FTI:

Table 11: Variables with FTI: gal\_zz\_2003\_t26.sas7bdat

Field name	Data dictionary reference name	Starting position	Field size	Data type
	HIT_BR	00016	8	N
	HIT_OTHERS	00024	8	N
3-digit Census MCD	A_MCD	00633	3	A/N
4-digit Census Place	A_PLACE	00641	4	A/N
5-digit FIPS MCD	A_FIPSMCD	00628	5	A/N
5-digit FIPS Place	A_FIPSPL	00636	5	A/N
6-character Traffic Analysis Zone (leading blanks)	A_TAZ	00617	6	A/N
Address on AHS 2002 =1; else=0	H2002	00582	1	A/N
Address on ES202 1991 =1; else=0	E1991	00567	1	A/N
Address on ES202 1992 =1; else=0	E1992	00568	1	A/N
Address on ES202 1993 =1; else=0	E1993	00569	1	A/N
Address on ES202 1994 =1; else=0	E1994	00570	1	A/N
Address on ES202 1995 =1; else=0	E1995	00571	1	A/N
Address on ES202 1996 =1; else=0	E1996	00572	1	A/N
Address on ES202 1997 =1; else=0	E1997	00573	1	A/N
Address on ES202 1998 =1; else=0	E1998	00574	1	A/N
Address on ES202 1999 =1; else=0	E1999	00575	1	A/N
Address on ES202 2000 =1; else=0	E2000	00576	1	A/N
Address on ES202 2001 =1; else=0	E2001	00577	1	A/N
Address on ES202 2002 =1; else=0	E2002	00578	1	A/N
Address on ES202 2003 =1; else=0	E2003	00579	1	A/N
Address on ES202 2004 =1; else=0	E2004	00580	1	A/N
Address on ES202 2005 =1; else=0	E2005	00581	1	A/N
Address on MAF 2004 =1; else=0	M2004	00583	1	A/N
Address on POW 2001 =1; else=0	P2001	00584	1	A/N
Address on POW 2002 =1; else=0	P2002	00585	1	A/N
Address on POW 2003 =1; else=0	P2003	00586	1	A/N
Address on POW 2004 =1; else=0	P2004	00587	1	A/N
Address on POW 2005 =1; else=0	P2005	00588	1	A/N
Census Block suffix 1	A_BLOCK_SUF1	00606	1	A/N
Census Block suffix 2	A_BLOCK_SUF2	00607	1	A/N
Census block within tract	A_BLOCK	00602	4	A/N
Census tract within county	A_TRACT	00611	6	A/N
Code1 Census block id 3 digit	C1_BLOCK	00514	3	A/N
Code1 Census block id 4 digit	C1_BLOCK4	00057	4	A/N
Code1 Census geocode (tract)	C1_GEOCODE	00502	12	A/N
Code1 USPS record type	C1_USPSRECTYPE	00492	1	A/N
Code1 ZIP	C1_ZIP	00482	5	A/N
Code1 ZIP code status	C1_ZIP_STATUS	00496	1	A/N
Code1 ZIP return code	C1_ZIP_RC	00046	1	A/N
Code1 ZIP source	C1_ZIP_SRC	00491	1	A/N
Code1 ZIP+4 code	C1_ZIP4	00487	4	A/N
Code1 ZIP4 return code	C1_ZIP4_RC	00047	1	A/N

(cont)

Table 11 – Continued

Field name	Data dictionary reference name	Starting position	Field size	Data type
Code1 address correctness score	C1_ADDRESS_CS	00494	1	A/N
Code1 address return code	C1_ADDRESS_RC	00042	1	A/N
Code1 address w/apt	C1_ADDRESS	00382	70	A/N
Code1 alias/base return code	C1_ALIAS_RC	00043	1	A/N
Code1 apartment return code	C1_APT_RC	00051	1	A/N
Code1 carrier route return code	C1_CARRTE_RC	00048	1	A/N
Code1 city name	C1_CITY	00452	28	A/N
Code1 city/state return code	C1_CITYSTATE_RC	00045	1	A/N
Code1 directional return code	C1_DIRECTIONAL_RC	00049	1	A/N
Code1 dropped information code	C1_DROPPEDINFO_RC	00044	1	A/N
Code1 general return code	C1_GENERAL_RC	00041	1	A/N
Code1 geocode return code	C1_GEO_RC	00501	1	A/N
Code1 lat/long coordinate	C1_LATLONG	00517	20	A/N
Code1 lat/long level	C1_LATLONG_RC	00537	1	A/N
Code1 master file vintage	C1_VINTDATE	00497	4	A/N
Code1 overall correctness	C1_OVERALL_RC	00493	1	A/N
Code1 state abbrev	C1_STATE	00480	2	A/N
Code1 street name correctness score	C1_STREETNAME_CS	00495	1	A/N
Code1 suffix return code	C1_SUFFIX_RC	00050	1	A/N
Code1; Maf; Equals maf; btWn maf; cOmmercial; mIxed; reSidential; all aDdresses	A_BLOCK_SRC	00608	1	A/N
Commercial, Mixed, or Residential	OCCUPANT_TYPE	00589	1	A/N
Contains only T26-sourced information	FLAG_T26	00038	3	N
FIPS county within state	A_CTY	00625	3	A/N
FIPS state	A_ST	00623	2	A/N
FIPS state—FIPS county	A_SSCCC	00659	5	A/N
FIPS state—FIPS county—Census tract	A_GEOCODE	00590	11	A/N
Latitude, 6 implied decimal places	A_LATITUDE	00000	8	N
Longitude, 6 implied decimal places	A_LONGITUDE	00008	8	N
MSA—PMSA	A_MSAPMSA	00645	8	A/N
Maf, Code1, Derived, Block (or part) internal point	A_LATLONG_SRC	00601	1	A/N
Maf; Assigned by distance	A_BLOCK_SUF_SRC	00610	1	A/N
Quality of lat/long	A_GEOQUAL	00035	3	N
Source file of this address	SRCMAST	00052	5	A/N
Tracking ID	BIGSRCID	00356	26	A/N
Unique GAL address ID	GALID	00538	29	A/N
A—YYYYMMDD_HHMM_ST—nnnnnnnnnn				
Vality additional address info	V_ADDADDR	00245	35	A/N
Vality address type	V_ADDRITYP	00280	1	A/N
Vality box type	V_BTYPE	00153	7	A/N
Vality box value	V_BVAL	00160	10	A/N
Vality building name	V_BUILDN	00215	30	A/N
Vality floor type	V_FTYPE	00170	5	A/N
Vality floor value	V_FVAL	00175	10	A/N
Vality house number	V_HNUM	00061	10	A/N

(cont)

Table 11 – Continued

Field name	Data dictionary reference name	Starting position	Field size	Data type
Vality house number suffix	V_HNUMS	00071	10	A/N
Vality input pattern	V_INPATT	00336	20	A/N
Vality multi-unit type	V_MUTYPE	00200	5	A/N
Vality multi-unit value	V_MUVAL	00205	10	A/N
Vality rural route type	V_RTYPE	00140	3	A/N
Vality rural route value	V_RVAL	00143	10	A/N
Vality street name	V_SNAME	00105	25	A/N
Vality street prefix - directional	V_SPRED	00081	2	A/N
Vality street prefix - type	V_SPRET	00085	20	A/N
Vality street suffix - directional	V_SSUFD	00083	2	A/N
Vality street suffix - qualifier	V_SSUFQ	00135	5	A/N
Vality street suffix - type	V_SSUFT	00130	5	A/N
Vality unhandled data	V_UNDATA	00301	35	A/N
Vality unhandled pattern	V_UNPATT	00281	20	A/N
Vality unit type	V_UTYPE	00185	5	A/N
Vality unit value	V_UVAL	00190	10	A/N
Vintage of Census geography (GRF)	A_VINTAGE	00032	3	N
Workforce Investment Board area	A_WIB	00653	6	A/N
only addr on blockFace; extraPolation	A_LATLONG_DRV	00609	1	A/N

\$Id: gal\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 13 ICF

**Name:** LEHD Individual Characteristic Files

**Code:** ICF and ICT (for FTI)

**Group:** ICF and ICT (for FTI)

**Origin:** State or Local Government Agency

**Sector:** Persons

**Period:** None

**Reference Period:** None

**Unit of Enumeration:** Individual Person

**Source:** Bureau of the Census, SSA

**Administrative Data:** Yes

**Legal authority:** The data were compiled from sources available within the Census Bureau under Title 13, U.S.C. Some items were integrated from sources covered by Title 26, U.S.C.

**Observations:** See Table 12.

Table 12: Number of observations, by state, by group, for ICF

State	Group	Number of datafiles	Records	Filesize (MB)
Covered states		62	402,634,528	42,070.20
Alabama	AL icf	1	2,921,217	374.87
Alabama	AL icft26	1	2,921,217	241.36
Arkansas	AR icf	1	1,590,571	204.77
Arkansas	AR icft26	1	1,590,571	131.44
California	CA icf	1	34,012,258	4,336.96
California	CA icft26	1	34,012,258	2,676.46
Colorado	CO icf	1	6,729,548	858.27
Colorado	CO icft26	1	6,729,548	541.36
Delaware	DE icf	1	806,071	103.29
Delaware	DE icft26	1	806,071	66.38
Florida	FL icf	1	18,252,091	2,320.04
Florida	FL icft26	1	18,252,091	1,461.77
Iowa	IA icf	1	2,505,678	321.18
Iowa	IA icft26	1	2,505,678	208.41
Idaho	ID icf	1	1,730,508	221.52
Idaho	ID icft26	1	1,730,508	140.14
Illinois	IL icf	1	13,875,529	1,767.07
Illinois	IL icft26	1	13,875,529	1,121.48

( cont)

Table 12 – Continued

State	Group	Number of datafiles	Records	Filesize (MB)	
Indiana	IN	icf	1	6,810,867	873.73
Indiana	IN	icft26	1	6,810,867	563.23
Kansas	KS	icf	1	3,897,988	498.09
Kansas	KS	icft26	1	3,897,988	315.27
Kentucky	KY	icf	1	3,587,680	460.12
Kentucky	KY	icft26	1	3,587,680	295.99
Maryland	MD	icf	1	7,235,448	925.88
Maryland	MD	icft26	1	7,235,448	587.50
Maine	ME	icf	1	1,206,021	154.18
Maine	ME	icft26	1	1,206,021	98.00
Minnesota	MN	icf	1	5,019,721	642.09
Minnesota	MN	icft26	1	5,019,721	412.84
Missouri	MO	icf	1	6,750,881	865.16
Missouri	MO	icft26	1	6,750,881	554.58
Montana	MT	icf	1	1,065,111	136.60
Montana	MT	icft26	1	1,065,111	87.54
North Carolina	NC	icf	1	9,630,526	1,241.21
North Carolina	NC	icft26	1	9,630,526	784.52
North Dakota	ND	icf	1	608,486	77.91
North Dakota	ND	icft26	1	608,486	50.66
New Jersey	NJ	icf	1	7,895,006	1,006.49
New Jersey	NJ	icft26	1	7,895,006	636.09
New Mexico	NM	icf	1	1,902,184	245.76
New Mexico	NM	icft26	1	1,902,184	157.04
Oklahoma	OK	icf	1	2,429,028	310.80
Oklahoma	OK	icft26	1	2,429,028	201.34
Oregon	OR	icf	1	4,374,229	563.52
Oregon	OR	icft26	1	4,374,229	357.90
Pennsylvania	PA	icf	1	11,983,407	1,534.79
Pennsylvania	PA	icft26	1	11,983,407	982.91
South Carolina	SC	icf	1	3,799,229	490.79
South Carolina	SC	icft26	1	3,799,229	314.05
Texas	TX	icf	1	19,411,438	2,480.27
Texas	TX	icft26	1	19,411,438	1,560.38
Virginia	VA	icf	1	6,452,661	831.88
Virginia	VA	icft26	1	6,452,661	535.44
Vermont	VT	icf	1	529,721	67.96
Vermont	VT	icft26	1	529,721	43.27
Washington	WA	icf	1	7,438,184	957.54
Washington	WA	icft26	1	7,438,184	605.27
Wisconsin	WI	icf	1	5,437,229	702.99
Wisconsin	WI	icft26	1	5,437,229	460.19
West Virginia	WV	icf	1	1,428,748	183.60
West Virginia	WV	icft26	1	1,428,748	118.07

**Industry:** n.a.

**NAICS coverage:** n.a.

**Documentation:** [http://lehd.did.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.did.census.gov/led/library/tech_user_guides.html)

**Short Description:** The Individual Characteristics File (**ICF**) for each state contains one record for every person who is ever employed in that state over the time period spanned by the state's unemployment insurance records.

**Long Description:** The Individual Characteristics File (**ICF**) is constructed in the following manner. First, the universe of individuals is defined by compiling the list of unique **PIKs** from the Employment History Files (**EHF**). Demographic information from the Personal Characteristics File (**PCF**) is then merged on by **PIK**, and records without a valid match flagged. **PIK**-survey identifier crosswalks link the **CPS** and **SIPP** ID variables into the **ICF**, and sex and age information from the **CPS** is used to complement and verify the **PCF**-provided information.

**Limitations:**

**Coverage:** See Table 13.

State		Start	End
Covered states		1985Q2	2004Q3
Alabama	AL	2001Q1	2004Q2
Arkansas	AR	2002Q3	2004Q2
California	CA	1991Q3	2003Q4
Colorado	CO	1990Q1	2004Q1
Delaware	DE	1998Q3	2004Q1
Florida	FL	1992Q4	2004Q1
Iowa	IA	1998Q4	2004Q1
Idaho	ID	1990Q1	2004Q1
Illinois	IL	1990Q1	2004Q1
Indiana	IN	1990Q1	2004Q1
Kansas	KS	1990Q1	2004Q1
Kentucky	KY	1996Q4	2004Q1
Maryland	MD	1985Q2	2004Q2
Maine	ME	1996Q1	2004Q1
Minnesota	MN	1994Q3	2004Q3
Missouri	MO	1990Q1	2004Q1
Montana	MT	1993Q1	2004Q1
North Carolina	NC	1991Q1	2003Q4
North Dakota	ND	1998Q1	2004Q2
New Jersey	NJ	1996Q1	2004Q1
New Mexico	NM	1995Q3	2004Q1
Oklahoma	OK	2000Q1	2004Q1
Oregon	OR	1991Q1	2004Q1
Pennsylvania	PA	1991Q1	2004Q1
South Carolina	SC	1998Q1	2004Q3
Texas	TX	1995Q1	2004Q2
Virginia	VA	1998Q1	2004Q1
Vermont	VT	2000Q1	2004Q1
Washington	WA	1990Q1	2004Q1
Wisconsin	WI	1990Q1	2004Q1
West Virginia	WV	1997Q1	2004Q1

Table 13: Coverage, by state, for ICF

**FTI:** The following files contain FTI:

File name	Variable	Field name
icf_zz_t26.sas7bdat	HUID_1999	Admin record huid
	HUIDSRC_1999	Admin record source of huid
	HSRC23_1999	Basic street address conflict flag
	FLAG_LATLONG	Flag quality of latitude/longitude of residence
	LATITUDE_LIVE	Latitude of residence, 6 implied decimal places
	LONGITUDE_LIVE	Longitude of residence, 6 implied decimal places
	COUNTY_LIVE	County of Residence as of year 1999:2
	COUNTYLIVEIMPUTEDCOUNTRY OF RESIDENCE IMPUTATION FLAG	

\$Id: icf\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 14 QWI

**Name:** LEHD-QWI (SEINUNIT file)  
**Code:** QWI  
**Group:** qwi  
**Origin:** Census Bureau constructed data  
**Sector:** Business  
**Period:** Quarterly  
**Reference Period:** Three (3) Months  
**Unit of Enumeration:** Establishment  
**Source:** Bureau of the Census  
**Administrative Data:** No  
**Legal authority:** Title 13, U.S.C.  
**Observations:** See Table 14.

Table 14: Number of observations, by state, for QWI

State		Number of datafiles	Records	Filesize (MB)
Covered states		26	166,312,551	583,715.02
Arkansas	AR	1	534,484	4,176.01
California	CA	1	43,638,767	89,783.13
Colorado	CO	1	5,605,248	43,791.32
Delaware	DE	1	535,373	1,078.45
Florida	FL	1	17,786,826	36,987.38
Iowa	IA	1	1,962,014	15,328.57
Idaho	ID	1	2,019,670	15,779.01
Indiana	IN	1	3,605,097	8,238.45
Kentucky	KY	1	1,250,675	2,841.88
Maine	ME	1	1,342,283	10,486.88
Minnesota	MN	1	4,748,456	9,316.32
Missouri	MO	1	5,669,283	44,291.57
Montana	MT	1	1,412,306	11,033.95
North Carolina	NC	1	10,150,481	22,599.01
North Dakota	ND	1	546,389	4,269.01
New Jersey	NJ	1	6,707,798	52,405.01
New Mexico	NM	1	1,578,333	12,331.07
Oklahoma	OK	1	1,511,516	11,809.07
Oregon	OR	1	5,424,000	42,375.32

( cont)

Table 14 – Continued

State		Number of datafiles	Records	Filesize (MB)
Pennsylvania	PA	1	8,660,512	19,016.32
Texas	TX	1	18,074,151	40,863.13
Virginia	VA	1	4,572,014	35,719.20
Vermont	VT	1	397,798	3,108.13
Washington	WA	1	9,536,038	19,100.45
Wisconsin	WI	1	7,712,727	16,593.01
West Virginia	WV	1	1,330,312	10,393.38

**Industry:** Combination (industries)

**NAICS coverage:** All

**Documentation:** [http://lehd.did.census.gov/led/library/tech\\_user\\_guides.html](http://lehd.did.census.gov/led/library/tech_user_guides.html)

**Short Description:** The [QWI](#) are a set of economic indicators – including employment, job creation, wages, and worker turnover – that can be queried by different levels of geography – state, county, metro, and workforce investment area – as well as by detailed industry, gender, and age of workers. The SEINUNIT file is the core dataset from which publication tables are tabulated. Public-use tabulations are not available in the [RDC](#). Internal-use confidential tabulations are considered “Custom Extracts”.

**Long Description:**

**Limitations:** These are very large filesizes. Users need to ensure that appropriate space is available for their project.

**Permission:** Bureau of the Census

**Coverage:** See Table 15.

State		Start	End
Covered states		1990Q1	2004Q3
Alabama	AL	2001Q1	2004Q2
Arkansas	AR	2002Q3	2004Q1
California	CA	1991Q3	2003Q4
Colorado	CO	1993Q2	2004Q1
Delaware	DE	1998Q3	2004Q1
Florida	FL	1992Q4	2004Q1
Iowa	IA	1998Q4	2004Q1
Idaho	ID	1991Q1	2004Q1
Illinois	IL	1990Q1	2004Q1
Indiana	IN	1998Q1	2004Q1
Kansas	KS	1993Q1	2004Q1
Kentucky	KY	2001Q1	2004Q1
Maryland	MD	1990Q1	2004Q1
Maine	ME	1996Q1	2004Q1
Minnesota	MN	1994Q3	2004Q3
Missouri	MO	1995Q1	2004Q1
Montana	MT	1993Q1	2004Q1
North Carolina	NC	1991Q1	2003Q4
North Dakota	ND	1998Q1	2004Q2
New Jersey	NJ	1996Q1	2004Q1
New Mexico	NM	1995Q3	2004Q1
Oklahoma	OK	1999Q1	2004Q1
Oregon	OR	1991Q1	2004Q1
Pennsylvania	PA	1997Q1	2004Q1
South Carolina	SC	1998Q1	2004Q3
Texas	TX	1995Q1	2004Q2
Virginia	VA	1995Q3	2004Q1
Vermont	VT	2000Q1	2004Q1
Washington	WA	1990Q1	2004Q1
Wisconsin	WI	1990Q1	2004Q1
West Virginia	WV	1997Q1	2004Q1

Table 15: Coverage, by state, for QWI

**FTI:** No files contain FTI.

\$Id: qwi\_ces\_description.tex 181 2007-06-08 02:55:26Z vilhu001 \$

## 15 Acronyms used

**ASM** Annual Survey of Manufacturers

**BES** Business Expenditure Survey

**BLS** Bureau of Labor Statistics

**BR** Business Register

**BRB** Business Register Bridge

**CPS** Current Population Survey

**DRB** Disclosure Review Board

**ECF** Employer Characteristics File

**ES-202** ES-202. An older name for the [QCEW](#) program

**EHF** Employment History Files

**EIN** (federal) Employer Identification Number

**FIPS** Federal information processing standards codes issued by National Institute of Standards and Technology ([NIST](#))

**FTI** Federal Tax Information, typically covered under Title 26, U.S.C.

**GAL** Geocoded Address List

**GIS** Geographic Information System

**ICF** Individual Characteristics File

**LED** Local Employment Dynamics

**LEHD** Longitudinal Employer-Household Dynamics

**LMI** Labor Market Information

**NIST** National Institute of Standards and Technology

**OTM** OnTheMap

**PCF** Personal Characteristics File

**PHF** Person History File

**PIK** Protected Identity Key

**QCEW** Quarterly Census of Employment and Wages, managed by the Bureau of Labor Statistics ([BLS](#))

**QWI** Quarterly Workforce Indicators

**RDC** Research Data Center

**SEIN** State employer identification number. It is constructed from the state Federal information processing standards codes ([FIPS](#)) code and the UI account number. The BLS refers to the UI account number in combination with the reporting unit number as SESA-ID

**SEINUNIT** SEIN reporting unit

**SIC** Standard Industry Classification

**SIPP** Survey of Income and Program Participation

**SPF** Successor-Predecessor File

**SSN** Social Security Number

**U2W** Unit-to-Worker Impute

**UI** unemployment insurance

## References

- Abowd, John M., Bryce E. Stephens and Lars Vilhuber. 2006. Confidentiality Protection in the Census Bureau's Quarterly Workforce Indicators. Technical Paper TP-2006-02 LEHD, U.S. Census Bureau.
- Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer and Simon D. Woodcock. 2006. The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. Technical paper TP-2006-01 U.S. Census Bureau, LEHD and Cornell University.

\$Id: snapshot\_ces\_description\_all.tex 181 2007-06-08 02:55:26Z vilhu001 \$