LEHD Public Use Data Schema v4.1-draft

Contents

(Printable version)

> **!**   **Important**
> This document is not an official Census Bureau publication. It is compiled from publicly accessible information by Lars Vilhuber (Labor Dynamics Institute, Cornell University). Feedback is welcome. Please write us at lars.vilhuber@cornell.edu.

The public-use data from the Longitudinal Employer-Household Dynamics Program, including the Quarterly Workforce Indicators (QWI) and Job-to-Job Flows (J2J), are available for download with the following data schema. These data are available as Comma-Separated Value (CSV) files through the LEHD website's Data page at http://lehd.ces.census.gov/data/ .

This document describes the data schema for LEHD files. For each variable, a set of allowable values is defined. Definitions are provided as CSV files, with header variable definitions. The naming conventions of the data files is documented in lehd_csv_naming.html. Changes relative to the original v4.0 version are listed at the end.

1   Basic Schema

Each file is structured as a CSV file. The first columns contain [identifiers], subsequent columns contain [indicators], followed by status flags.

1.1   Generic structure

Column name
[Identifier1]
[Identifier2]
[Identifier3]
[…]
[Indicator 1]
[Indicator 2]
[Indicator 3]
[…]
[Status Flag 1]
[Status Flag 2]
[Status Flag 3]
[…]

Note: A full list of indicators for each type of file are shown below in the Indicators section. While all indicators are included in the CSV files, only the requested indicators will be included in data outputs from the LED Extraction Tool.

1.2 Identifiers

Records, unless otherwise noted, are parts of time-series data. Unique record identifiers are noted below, by file type. Identifiers without the year and quarter component can be considered a series identifier.

1.2.1 Mapping for Identifiers

(lehd_mapping_identifiers.csv)

Each of the released files has a set of variables uniquely identifying records (*Identifiers*). The table below relates the set of identifier specifications to the released files. The actual CSV files containing the identifiers for each set are listed after this table. Each identifier can take on a specified list of values, documented in the section on Categorical Variables.

identifiers	QWI	NQWI	J2J	J2JR	J2JOD	LODES
lehd_identifiers_qwi	1	1				
lehd_identifiers_j2j			1	1		
lehd_identifiers_j2jod					1	

1.2.2 Identifiers for j2j

(lehd_identifiers_j2j.csv)

Variable	Type	label
periodicity	Char(1)	Periodicity of report
seasonadj	Char(1)	Seasonal Adjustment Indicator
geo_level	Char(1)	Group: Geographic level of aggregation
geography	Char(8)	Group: Geography code
ind_level	Char(1)	Group: Industry level of aggregation
industry	Char(5)	Group: Industry code
ownercode	Char(3)	Group: Ownership group code
sex	Char(1)	Group: Gender code
agegrp	Char(3)	Group: Age group code (WIA)
race	Char(2)	Group: race
ethnicity	Char(2)	Group: ethnicity
education	Char(2)	Group: education
firmage	Char(1)	Group: Firm Age group
firmsize	Char(1)	Group: Firm Size group
year	Num	Time: Year
quarter	Num	Time: Quarter
agg_level	Num	Aggregation Level Indicator

1.2.3   Identifiers for j2jod

(lehd_identifiers_j2jod.csv)

Variable	Type	label
periodicity	Char(1)	Periodicity of report
seasonadj	Char(1)	Seasonal Adjustment Indicator
geo_level	Char(1)	Group: Geographic level of aggregation
geography	Char(8)	Group: Geography code
ind_level	Char(1)	Group: Industry level of aggregation
industry	Char(5)	Group: Industry code
ownercode	Char(3)	Group: Ownership group code
sex	Char(1)	Group: Gender code
agegrp	Char(3)	Group: Age group code (WIA)
race	Char(2)	Group: race
ethnicity	Char(2)	Group: ethnicity
education	Char(2)	Group: education
firmage	Char(1)	Group: Firm Age group
firmsize	Char(1)	Group: Firm Size group
year	Num	Time: Year
quarter	Num	Time: Quarter
agg_level	Num	Aggregation Level Indicator
geo_level_orig	Char(1)	Group: Geographic level of aggregation of origin job
geography_orig	Char(8)	Group: Geography code of origin job
ind_level_orig	Char(1)	Group: Industry level of aggregation of origin job
industry_orig	Char(5)	Group: Industry code of origin job
ownercode_orig	Char(3)	Group: Ownership group code of origin job
firmage_orig	Char(1)	Group: Firm Age group of origin job
firmsize_orig	Char(1)	Group: Firm Size group of origin job

1.2.4 Identifiers for qwi

(lehd_identifiers_qwi.csv)

Variable	Type	label
periodicity	Char(1)	Periodicity of report
seasonadj	Char(1)	Seasonal Adjustment Indicator
geo_level	Char(1)	Group: Geographic level of aggregation
geography	Char(8)	Group: Geography code
ind_level	Char(1)	Group: Industry level of aggregation
industry	Char(5)	Group: Industry code
ownercode	Char(3)	Group: Ownership group code
sex	Char(1)	Group: Gender code
agegrp	Char(3)	Group: Age group code (WIA)
race	Char(2)	Group: race
ethnicity	Char(2)	Group: ethnicity
education	Char(2)	Group: education
firmage	Char(1)	Group: Firm Age group
firmsize	Char(1)	Group: Firm Size group
year	Num	Time: Year
quarter	Num	Time: Quarter

1.3 Indicators

The following tables and associated mapping files list the indicators available on each file. The *'Indicator Variable'* is the short name of the variable on the CSV files, suitable for machine processing in a wide variety of statistical applications. When given, the *'Alternate name'* may appear in related documentation and articles. The *'Status Flag'* is used to indicate publication or data quality status (see Status Flags). The *'Indicator Name'* is a more verbose description of the indicator.

1.3.1 National QWI and state-level QWI

(variables_qwipu.csv)

Indicator Variable	Status Flag	Indicator Name
Emp	sEmp	Beginning-of-Quarter Employment: Counts
EmpEnd	sEmpEnd	End-of-Quarter Employment: Counts
EmpS	sEmpS	Full-Quarter Employment (Stable): Counts
EmpSpv	sEmpSpv	Full-Quarter Employment in the Previous Quarter: Counts
EmpTotal	sEmpTotal	Employment - Reference Quarter: Counts
HirA	sHirA	Hires All: Counts (Accessions)
HirN	sHirN	Hires New: Counts
HirR	sHirR	Hires Recalls: Counts
Sep	sSep	Separations: Counts
HirAEnd	sHirAEnd	End-of-Quarter Hires
HirAEndR	sHirAEndR	End-of-Quarter Hiring Rate
SepBeg	sSepBeg	Beginning-of-Quarter Separations
SepBegR	sSepBegR	Beginning-of-Quarter Separation Rate
HirAs	sHirAs	Hires All (Stable): Counts (Flows into Full-Quarter Employment)
HirNs	sHirNs	Hires New (Stable): Counts (New Hires to Full-Quarter Status)
SepS	sSepS	Separations (Stable): Counts (Flow out of Full-Quarter Employment)
SepSnx	sSepSnx	Separations (Stable): Next Quarter: Counts (Flow out of Full-Quarter Employment)
TurnOvrS	sTurnOvrS	Turnover (Stable)
FrmJbGn	sFrmJbGn	Firm Job Gains: Counts (Job Creation)
FrmJobLs	sFrmJobLs	Firm Job Loss: Counts (Job Destruction)
FrmJbC	sFrmJbC	Firm Job Change: Net Change
HirAEndRepl	sHirAEndRepl	Replacement Hires
HirAEndReplr	sHirAEndReplr	Replacement Hiring Rate
FrmJbGnS	sFrmJbGnS	Firm Job Gains (Stable): Counts
FrmJbLsS	sFrmJbLsS	Firm Job Loss (Stable): Counts
FrmJbCS	sFrmJbCS	Job Change (Stable): Net Change
EarnS	sEarnS	Full Quarter Employment (Stable): Average Monthly Earnings
EarnBeg	sEarnBeg	Beginning-of-Quarter Employment: Average Monthly Earnings
EarnHirAS	sEarnHirAS	Hires All (Stable): Average Monthly Earnings
EarnHireNS	sEarnHireNS	Hires New (Stable): Average Monthly Earnings
EarnSepS	sEarnSepS	Separations (Stable): Average Monthly Earnings
Payroll	sPayroll	Total Quarterly Payroll: Sum

1.3.2 Job-to-job flow counts (J2J)

(variables_j2j.csv)

Indicator Variable	Alternate name	Status Flag	Indicator Name
MHire	all_doma2	sHire	Main Job Accessions - Counts
MSep	all_doms2	sSep	Main Job Separations - Counts
MJobStart	all_dest	sMJobStart	Main Job Starts - Counts (hires + main source of earnings changes)
MJobEnd	all_orig	sMJobEnd	Main Job Ends - Counts (separations + main source of earnings changes)
EEHire	ee_doma2	sEEHire	EE Accessions - Counts
EESep	ee_doms2	sEESep	EE Separations - Counts
AQHire	aq_doma2	sAQHire	AQ Accessions - Short Duration Nonemployment Spell - Counts
AQSep	aq_doms2	sAQSep	AQ Separations - Short Duration Nonemployment Spell - Counts
J2JHire	eeall_doma2	sJ2JHire	Job-to-Job Accessions - Counts
J2JSep	eeall_doms2	sJ2JSep	Job-to-Job Separations - Counts
NEHire	ne_doma2	sNEHire	Main Job Accessions - Any Nonemployment Spell - Counts
ENSep	en_doms2	sENSep	Main Job Separations - Any Nonemployment Spell - Counts
NEPersist	ne2_doma2	sNEPersist	Main Job Accessions - Persistent Nonemployment Spell - Counts
ENPersist	en2_doms2	sENPersist	Main Job Separations - Persistent Nonemployment Spell - Counts
NEFullQ	ne2p_doma2	sNEFullQ	Main Job Accessions - Full-qtr Nonemployment Spell - Counts
ENFullQ	en2p_doms2	sENFullQ	Main Job Separations - Full-qtr Nonemployment Spell - Counts
MainB	domB	sMainB	Main Job - Beginning of Quarter - Counts
MainE	domE	sMainE	Main Job - End of Quarter - Counts

1.3.3 Job-to-job flow rates (J2JR)

(variables_j2jr.csv)

Indicator Variable	Alternate name	Status Flag	Indicator Name
MHireR	all_doma2_rate	sMHireR	Main Job Accessions - Rate
MSepR	all_doms2_rate	sMSepR	Main Job Separations - Rate
MJobStartR	all_dest_rate	sMJobStartRate	Main Job Starts - Rate
MJobEndR	all_orgin_rate	sMJobEndRate	Main Job Ends - Rate
EEHireR	eea_rate	sEEHireRate	EE Accessions - Rate
EESepR	ees_rate	sEESepR	EE Separations - Rate
AQHireR	aq_doma2_rate	sAQHireR	AQ Accessions - Short Duration Nonemployment Spell - Rate
AQSepR	aq_doms2_rate	sAQSepR	AQ Separations - Short Duration Nonemployment Spell - Rate
J2JHireR	eeall_doma2_rate	sJ2JHireR	Job-to-Job Accessions - Rate
J2JSepR	eeall_doms2_rate	sJ2JSepR	Job-to-Job Separations - Rate
NEHireR	ne_doma2_rate	sNEHireR	Main Job Accessions - Any Nonemployment Spell - Rate
ENSepR	en_doms2_rate	sENSepR	Main Job Separations - Any Nonemployment Spell - Rate
NEPersistR	ne2_doma2_rate	sNEPersistR	Main Job Accessions - Persistent Nonemployment Spell - Rate
ENPersistR	en2_doms2_rate	sENPersistR	Main Job Separations - Persistent Nonemployment Spell - Rate
NEFullQR	ne2p_doma2_rate	sNEFullQR	Main Job Accessions - Full-qtr Nonemployment Spell - Rate
ENFullQR	en2p_doms2_rate	sENFullQR	Main Job Separations - Full-qtr Nonemployment Spell - Rate

1.3.4 Job-to-job flow Origin-Destination (J2JOD)

(variables_j2jod.csv)

Indicator Variable	Alternate name	Status Flag	Indicator Name
EE	ee	sEE	Direct Job Flows - Counts
AQHire	aq_doma2	sAQHire	Main Job Accessions - Short Nonemployment Spell - Counts
EEFullQ	fee	sEEFullQ	Stable Job to Stable Job Direct Flows - Counts
AQFullQHire	faq_doma2	sAQFullQHire	Stable Job Accessions From Stable Jobs - Short Nonemployment Spell Between Jobs - Counts
EEFullQEarn_dest	fee_kfqearn	sEEFullQEarn_dest	Stable Job to Stable Job Direct Flow - Average Full-Quarter Earnings in First Full Quarter of Job
EEFullQEarn_change	fee_dearn	sEEFullQEarn_change	Stable Job to Stable Job Direct Flow - Average Percentage Real Earnings Change
AQFullQEarn_dest	faq_kfqearn	sAQFullQEarn_dest	Stable Job to Stable Job Direct Flow - Short Nonemployment Spell Between Jobs - Average Full-Quarter Earnings in First Full Quarter of Job
AQFullQEarn_change	faq_dearn	sAQFullQEarn_change	Stable Job to Stable Job Direct Flow - Short Nonemployment Spell Between Jobs - Average Percentage Real Earnings Change

2   Categorical Variables

Categorical variable descriptions are displayed above each table, with the variable name shown in parentheses. Unless otherwise stated, every possible value/label combination for each categorical variable is listed. Please note that not all values will be available in every table.

2.1   agegrp

(label_agegrp.csv)

agegrp	label
A00	All Ages (14-99)
A01	14-18
A02	19-21
A03	22-24
A04	25-34
A05	35-44
A06	45-54
A07	55-64
A08	65-99

2.2   education

(label_education.csv)

education	label
E0	All Education Categories
E1	Less than high school
E2	High school or equivalent, no college
E3	Some college or Associate degree
E4	Bachelor's degree or advanced degree
E5	Educational attainment not available (workers aged 24 or younger)

2.3   ethnicity

(label_ethnicity.csv)

ethnicity	label
A0	All Ethnicities
A1	Not Hispanic or Latino
A2	Hispanic or Latino

2.4   firmage

(label_firmage.csv)

firmage	label
0	All Firm Ages
1	0-1 Years
2	2-3 Years
3	4-5 Years

firmage	label
4	6-10 Years
5	11+ Years
N	Firm Age Not Available

2.5 firmsize

(label_firmsize.csv)

firmsize	label
0	All Firm Sizes
1	0-19 Employees
2	20-49 Employees
3	50-249 Employees
4	250-499 Employees
5	500+ Employees
N	Firm Size Not Available

2.6 ownercode

(label_ownercode.csv)

ownercode	label
A00	All (1-5)
A01	Federal government
A05	All Private (5)

2.7 periodicity

(label_periodicity.csv)

periodicity	label
A	Annual data
Q	Quarterly data

2.8 quarter

(label_quarter.csv)

quarter	label
1	1st Quarter of the Year (January-March)
2	2nd Quarter of the Year (April-June)
3	3rd Quarter of the Year (July-September)
4	4th Quarter of the Year (October-December)

2.9 race

(label_race.csv)

race	label
A0	All Races

race	label
A1	White Alone
A2	Black or African American Alone
A3	American Indian or Alaska Native Alone
A4	Asian Alone
A5	Native Hawaiian or Other Pacific Islander Alone
A6	Some Other Race Alone (Not Used)
A7	Two or More Race Groups

2.10   seasonadj

(label_seasonadj.csv)

seasonadj	label
S	Seasonally adjusted
U	Not seasonally adjusted

2.11   sex

(label_sex.csv)

sex	label
0	All Sexes
1	Male
2	Female

2.12 Industry

2.12.1 Industry levels

(label_ind_level.csv)

ind_level	label
A	All Industries
S	NAICS Sectors
3	NAICS Subsectors
4	NAICS Industry Groups

2.12.2 Industry

(label_industry.csv)

Only a small subset of available values shown. The 2007 NAICS (North American Industry Classification System) is used for all years. For a full listing of all valid NAICS codes, see http://www.census.gov/eos/www/naics/.

industry	label
00	All NAICS Sectors
000	All NAICS Subsectors
0000	All NAICS Industry Groups
11	Agriculture, Forestry, Fishing and Hunting
111	Crop Production
1111	Oilseed and Grain Farming
1112	Vegetable and Melon Farming
. . .	
2382	Building Equipment Contractors
2383	Building Finishing Contractors
2389	Other Specialty Trade Contractors
31-33	Manufacturing
311	Food Manufacturing
3111	Animal Food Manufacturing
3112	Grain and Oilseed Milling
3113	Sugar and Confectionery Product Manufacturing
. . .	

2.13   Geography

2.13.1   Geographic levels

(label_geo_level.csv)

geo_level	label
C	Counties
M	Metropolitan/Micropolitan
N	National (50 States + DC)
S	States
W	Workforce Investment Areas

Geography labels are provided in separate files by state. Note that cross-state CBSA will have state-specific parts, and thus will appear in multiple files. A separate label_fipsnum.csv contains values and labels for all entities of geo_level *n* or *s*, and is a summary of separately available files.

2.13.2   National and state-level values

(label_fipsnum.csv)

geography	label
00	National (50 States + DC)
02	Alaska
01	Alabama
05	Arkansas
04	Arizona
06	California
08	Colorado
. . .	
46	South Dakota
47	Tennessee
48	Texas
49	Utah
51	Virginia
50	Vermont
53	Washington
55	Wisconsin

2.13.3   Detailed state and substate level values

For a full listing of all valid geography codes (except for WIA codes), see http://www.census.gov/geo/maps-data/data/tiger.html. Note about geography codes: Four types of geography codes are represented with this field. Each geography has its own code structure.

- State is the 2-digit FIPS code.

- County is the 5-digit FIPS code.

- Metropolitan/Micropolitan codes are constructed from the 2-digit state FIPS code and the 5-digit CBSA code provided by the Census Bureau's Geography Division.

  - In the QWI, the metropolitan/micropolitan areas are the state parts of the full CBSA areas.

  - In J2J, tabulations are based on the complete metropolitan/micropolitan area.

- The WIA code is constructed from the 2-digit state FIPS code and the 6-digit WIA identifier provided by LED State Partners.

The 2014 vintage of Census TIGER geography is used for all tabulations as of the 2014Q3 release.

! **Important**
The above section should include hyperlinks to the appropriate reference.

State	Format file
AK	label_geography_ak.csv
AL	label_geography_al.csv
AR	label_geography_ar.csv
AZ	label_geography_az.csv
CA	label_geography_ca.csv
CO	label_geography_co.csv
CT	label_geography_ct.csv
DC	label_geography_dc.csv
DE	label_geography_de.csv
FL	label_geography_fl.csv
GA	label_geography_ga.csv
HI	label_geography_hi.csv
IA	label_geography_ia.csv
ID	label_geography_id.csv
IL	label_geography_il.csv
IN	label_geography_in.csv
KS	label_geography_ks.csv
KY	label_geography_ky.csv
LA	label_geography_la.csv
MD	label_geography_md.csv
ME	label_geography_me.csv
MI	label_geography_mi.csv
MN	label_geography_mn.csv
MO	label_geography_mo.csv
MS	label_geography_ms.csv
MT	label_geography_mt.csv
NC	label_geography_nc.csv
ND	label_geography_nd.csv
NE	label_geography_ne.csv
NH	label_geography_nh.csv
NJ	label_geography_nj.csv
NM	label_geography_nm.csv
NV	label_geography_nv.csv
NY	label_geography_ny.csv
OH	label_geography_oh.csv
OK	label_geography_ok.csv
OR	label_geography_or.csv
PA	label_geography_pa.csv
RI	label_geography_ri.csv
SC	label_geography_sc.csv
SD	label_geography_sd.csv
TN	label_geography_tn.csv
TX	label_geography_tx.csv
UT	label_geography_ut.csv
VA	label_geography_va.csv

State	Format file
VT	label_geography_vt.csv
WA	label_geography_wa.csv
WI	label_geography_wi.csv
WV	label_geography_wv.csv
WY	label_geography_wy.csv

2.14 Aggregation level

(label_agg_level.csv)

The linked file (label_agg_level.csv) has columns indicating exactly which detailed variabes are included. Columns geo_level and ind_level are explained above.

agg_level	worker_char	firm_char	firm_orig_char	j2j	j2jr
1				0	1
2	Ethnicity			0	1
3	Race			0	1
4	Race * Ethnicity			0	1
5	Education			0	1
6	Age			0	1
7	Sex			0	1
8	Sex * Education			0	1
9	Sex * Age			0	1
10		Firm Size		0	1

3 Status flags

(label_flags.csv)

Each status flag in the tables above contains one of the following valid values. The values and their interpretation are listed in the table below.

! **Important**

Note: Currently, the J2J tables only contain status flags *-1* and *1.* Status flags with values 10 or above only appear in online applications, not in CSV files.

---

flag	label
-2	no data available in this category for this quarter
-1	data not available to compute this estimate
1	OK
5	Value suppressed because it does not meet US Census Bureau publication standards.
6	Value calculated from other released measures - no significant distortion
7	Value calculated from other released measures - some of which have significantly distorted data
9	Data significantly distorted - fuzzed value released
10	Aggregate of cells - no significant distortion
11	Aggregate of cells not released because component cells do not meet U.S. Census Bureau publication standards
12	Aggregate of cells - some of which have significantly distorted data

4 Changes

4.1 This version from previous releases of this document

- 2015-02-25: corrected flag values

- 2015-02-25: documents are now identified by date, not revision

- 2015-03-10: Correction of the TIGER vintage htat is used for geographic references

4.2 Version 4.1-draft from 4.0

- added J2J, National QWI spec

This revision: Wed Mar 11 22:10:17 EDT 2015