LEHD Public Use Data Schema for J2J Explorer (beta) V4.2-rc1
LEHD Public Use Data Schema for J2J Explorer (beta) V4.2-rc1 LEHD Public Use Data Schema for J2J Explorer (beta) V4.2-rc1

Contents

1	Purp	pose	1
2	Add	itional information	1
3	Exte	ends	1
4	Supo	ersedes	1
5	Basi	c Schema	1
	5.1	Generic structure	1
	5.2	Identifiers	3
		5.2.1 Identifiers for j2j	3
		5.2.2 Identifiers for j2jod	4
	5.3	Indicators	5
6	Cate	egorical Variables	6
	6.1	agegrp	6
	6.2	concept_draft	6
	6.3	education	6
	6.4	ethnicity	6
	6.5	firmage	7
	6.6	firmsize	7
	6.7	ownercode	7
	6.8	periodicity	7
	6.9	quarter	7
	6.10	race	8
	6.11	seasonadj	8
	6.12	sex	8
	6.13	stusps	8
	6.14	Industry	10
		6.14.1 Industry levels	10
		6.14.2 Industry	10
	6.15	Geography	11
		6.15.1 Geographic levels	11
		6.15.2 National and state-level values	11
		6.15.3 State postal codes	12
		6.15.4 Detailed state and substate level values	13
	6.16	Aggregation level	15

7	7 Status flags		
8	Met	adata	18
	8.1	Version Metadata for J2J Files (version.txt)	18
	8.2	Additional Metadata for J2JOD Files (avail.csv)	18
	8.3	Metadata on Indicator Availability	19

(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.

1 Purpose

The public-use Job-to-Job Flows (J2J) data provided by the Longitudinal Employer-Household Dynamics Program are accessible through the J2J Explorer (beta). This document provides information on the schema used to format files downloaded through that application.

2 Additional information

The complete LEHD schema is documented in lehd_public_use_schema.pdf. LEHD-provided SHP files are separately described in lehd_shapefiles.pdf. The naming conventions of the data files is documented in lehd_csv_naming.pdf.

3 Extends

This is the first version of the schema for the J2J Explorer (beta) application.

4 Supersedes

No prior version.

5 Basic Schema

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

5.1 Generic structure

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

Note: The J2J Explorer (beta) provides	the full set of J2J indicate	ators in addition to two con	nposite Origin-Destination indica-
tors. Files downloadable through other lehd_public_use_schema.pdf.	means may be structure	ed differently, please consul	It the complete LEHD schema in

5.2 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.

5.2.1 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

5.2.2 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 of destination job
geography	Char(8)	Group: Geography code of destination job
ind_level	Char(1)	Group: Industry level of aggregation of destination job
industry	Char(5)	Group: Industry code of destination job
ownercode	Char(3)	Group: Ownership group code of destination job
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

5.3 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 name for the indicator. The 'Description' provides a complete description of the indicator. 'Units' identify the type of variable: counts, rates, monetary amounts. 'Concept' classifies each indicator in a descriptive category: employment, hire, separation, earnings, or flow. The 'Base' indicates the denominator used to compute the statistic, and may be 1.

(variables_j2japp.csv)

Indicator Vari-	Alternate Name	Status Flag	Indicator Name	Description	Units	Concept	Base
able		8					
Ј2Ј	j2j_doma2	sJ2J	Job-to-Job Flows	Job flows with a short or no observed nonemployment spell	Count	Hire	1
J2JS	fj2j_doma	2 sJ2JS	Stable Job-to-Job Flows	Job flows from stable employment into stable employment with a short or no observed nonemployment spell	Count	Hire	1
J2JSEarn_	Ofijiஐ _doma	2 <u>sji/2,EsiFa</u> rn	_ Orvig rage Earnings prior to Job-to-Job Flows	Average quarterly earnings prior to job flows with a short or no observed nonemployment spell	Dollars	Earnings	J2JS
J2JSEarn_	D ēj Ωij_doma.	2 <u>s</u> lf £q&Ea rn	Deserrage Earnings following Job-to-Job Flows	Average quarterly earnings following job flows with a short or no observed nonemployment spell	Dollars	Earnings	J2JS

6 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.

6.1 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

6.2 concept_draft

(label_concept_draft.csv)

Concept	Definition
Hire	Measure derived from the the first quarter that a job is
	the main source of income (usually a hire)
Separation	Measure derived from the last quarter that a job is the
	main source of income (usually a separation)
Employment	Measure counting the stock of all jobs at a point in
	time
Earnings Measure of average earnings associated with a b	
	employment or job transition measure

6.3 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)

6.4 ethnicity

(label_ethnicity.csv)

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

6.5 firmage

(label_firmage.csv)

firmage	label	
0	All Firm Ages	
1	0-1 Years	
2	2-3 Years	
3	4-5 Years	
4	6-10 Years	
5	11+ Years	
N	Firm Age Not Available For Public-Sector Firms	

6.6 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 For Public-Sector Firms	

6.7 ownercode

(label_ownercode.csv)

ownercode	label
A00	State and local government plus private ownership
A01	Federal government
A05	All Private

6.8 periodicity

(label_periodicity.csv)

periodicity	label
A	Annual data
Q	Quarterly data

6.9 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)	

6.10 race

(label_race.csv)

race	label	
A0	All Races	
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	

6.11 seasonadj

(label_seasonadj.csv)

seasonadj	label
S	Seasonally adjusted
U	Not seasonally adjusted

6.12 sex

(label_sex.csv)

sex	label
0	All Sexes
1	Male
2	Female

6.13 stusps

(label_stusps.csv)

geography	stusps
00	US
01	AL
02	AK
04	AZ
05	AR
06	CA
08	CO
09	CT
10	DE
11	DC
12	FL

geography	stusps
13	GA
15	НІ
16	ID
17	IL
18	IN
19	IA
20	KS
21	KY
22	LA
23	ME
24	MD
25	MA
26	MI
27	MN
28	MS
29	MO
30	MT
31	NE
32	NV
33	NH
34	NJ
35	NM
36	NY
37	NC
38	ND
39	OH
40	OK
41	OR
42	PA
44	RI
45	SC
46	SD
47	TN
48	TX
49	UT
50	VT
51	VA
53	WA
54	WV
55	WI
56	WY
72	PR
78	VI

6.14 Industry

6.14.1 Industry levels

(label_ind_level.csv)

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

6.14.2 Industry

(label_industry.csv)

Only a small subset of available values shown. The 2017 NAICS (North American Industry Classification System) is used for all years. QWI releases prior to R2018Q1 used the 2012 NAICS classification (see Schema v4.1.3). For a full listing of all valid 2017 NAICS codes, see http://www.census.gov/cgi-bin/sssd/naicsrch?chart=2017.

industry	label	ind_level
00	All NAICS Sectors	A
000	All NAICS Subsectors	A
0000	All NAICS Industry Groups	A
11	Agriculture, Forestry, Fishing and Hunting	2
111	Crop Production	3
1111	Oilseed and Grain Farming	4
1112	Vegetable and Melon Farming	4
2382	Building Equipment Contractors	4
2383	Building Finishing Contractors	4
2389	Other Specialty Trade Contractors	4
31-33	Manufacturing	2
311	Food Manufacturing	3
3111	Animal Food Manufacturing	4
3112	Grain and Oilseed Milling	4
3113	Sugar and Confectionery Product Manufacturing	4

6.15 Geography

6.15.1 Geographic levels

Geography labels for data files are provided in separate files, by scope. Each file *label_geograpy_SCOPE.csv* may contain one or more types of records as flagged by geo_level. For convenience, a composite file containing all geocodes is available as label_geography.csv. The 2017 vintage of Census TIGER/Line geography is used for all tabulations as of the R2018Q1 release.

Shapefiles are described in a separate document.

(label_geo_level.csv)

geo_l	ev ka lbel	description	sourceurl
В	Metropolitan	Identifies 5-digit CBSA code for	http://www.census.gov/-
	(complete)	metropolitan areas provided by the	population/metro/
		Census Bureau's Geography	
		Division. Balance of state	
		including micropolitan areas are	
		identified by custom codes as	
		[ST]999	
С	Counties	Identifies 5-digit FIPS/ANSI code	https://www.census.gov/geo/-
		for counties	reference/codes/cou.html
M	Metropolitan/N	Milchentioflies n7-digit code constructed	http://www.census.gov/-
	(state part)	from the 2-digit state FIPS code	population/metro/
		and the 5-digit CBSA code	
		provided by the Census Bureau's	
		Geography Division	
N	National (50	Custom code using 00 to denote	
	States + DC)	national scope	
S	States	Identifies 2-digit FIPS/ANSI codes	https://www.census.gov/geo/-
			reference/ansi_statetables.html
W	Workforce	2-digit state FIPS code and the	
	Investment	6-digit WIA identifier provided by	
	Areas	LED State Partners	

6.15.2 National and state-level values

(label_fipsnum.csv)

The file label_fipsnum.csv contains values and labels for all entities of geo_level *N* or *S*, and is a summary of separately available files.

geograph	geo_level	
00	National (50 States +	N
	DC)	
01	Alabama	S
02	Alaska	S
04	Arizona	S
05	Arkansas	S
06	California	S
08	Colorado	S
45	South Carolina	S
46	South Dakota	S
47	Tennessee	S
48	Texas	S
49	Utah	S

geograpl	geographylabel			
50	Vermont	S		
51	Virginia	S		
53	Washington	S		

6.15.3 State postal codes

Some parts of the schema use (lower or upper-case) state postal codes.

(label_stusps.csv)

geography	stusps
00	US
01	AL
02	AK
04	AZ
05	AR
06	CA
08	CO
09	CT
10	DE
11	DC
12	FL
13	GA
15	HI
16	ID
17	IL
18	IN
19	IA
20	KS
21	KY
22	LA
23	ME
24	MD
25	MA
26	MI
27	MN
28	MS
29	MO
30	MT
31	NE
32	NV
33	NH
34	NJ
35	NM
36	NY
37	NC
38	ND
39	ОН
40	OK
41	OR
42	PA
44	RI
45	SC
46	SD
47	TN
48	TX
I .	I

geography	stusps
49	UT
50	VT
51	VA
53	WA
54	WV
55	WI
56	WY
72	PR
78	VI

6.15.4 Detailed state and substate level values

Note: cross-state CBSA, in records of type $geo_level = M$, are present on files of type $label_geography_XX.csv$. A particular cross-state CBSA will appear on multiple files.

Scope	Format file
US	label_geography_us.csv
METRO	label_geography_metro.csv
States	
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
MA	label_geography_ma.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
ОН	label_geography_oh.csv

Scope	Format file
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
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

6.16 Aggregation level

(label_agg_level.csv)

Measures within the J2J and QWI data products are tabulated on many different dimensions, including demographic characteristics, geography, industry, and other firm characteristics. For Origin-Destination (O-D) tables, characteristics of the origin and destination firm can be tabulated separately. Every tabulation level is assigned a unique aggregation index, represented by the agg_level variable. This index starts from 1, representing a national level grand total (all industries, workers, etc.), and progresses through different combinations of characteristics. There are gaps in the progression to leave space for aggregation levels that may be included in future data releases.

agg_level is currently reported only for J2J data products.

The following variables are included in the label_agg_level.csv file:

Variable	Description
agg_level	index representing level of aggregation reported
	on a given record
worker_char	demographic (worker) characteristics reported
	on record
firm_char	firm characteristics reported on record. These
	will be the characteristics of the destination firm
	in O-D tabulations
firm_orig_char	characteristics of origin firm reported on record
	(O-D tabulations only)
j2j	Flag: Aggregation level available on J2J counts
	tables
j2jr	Flag: Aggregation level available on J2J rates
	tables
j2jod	Flag: Aggregation level available on J2J O-D
	tables
qwi	Flag: Aggregation level available on QWI

The characteristics available on an aggregation level are repeated using a series of flags following the standard schema:

- geo_level geographic level of table
- ind_level industry level of table
- by_variables flags indicating other dimensions reported, including ownership, demographics, firm age and size.

A shortened representation of the file is provided below, the complete file is available in the link above.

agg_level worker_char		firm_char	firm_orig_char	· j2j	j2jr	j2jod	qwi	geo_level
1				1	1	1	0	N
2	Sex			1	1	1	0	N
3	Age			1	1	1	0	N
4	Sex * Age			1	1	1	0	N
5	Race			1	1	1	0	N
9	Ethnicity			1	1	1	0	N
13	Race *			1	1	1	0	N
	Ethnicity							
129		Firm Size		1	1	1	0	N
257		NAICS		1	1	1	0	N
		Sector						
258	Sex	NAICS		1	1	1	0	N
		Sector						

agg_leve	l worker_char	firm_char	firm_orig_char	· j2j	j2jr	j2jod	qwi	geo_leve
1029	Race	State		1	1	1	0	S
1033	Ethnicity	State		1	1	1	0	S
1037	Race *	State		1	1	1	0	S
	Ethnicity							

7 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, 1, 5. 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			

8 Metadata

(variables_version.csv)

8.1 Version Metadata for J2J Files (version.txt)

Each data release is accompanied by one or more files with metadata on geographic and temporal coverage, in a compact notation. These files follow the following naming convention:

version_[type].txt

where each component is described in more detail in lehd_csv_naming.pdf.

The contents contains the following elements:

Component	Source	Description
product	[type]	Type as described in lehd_csv_naming.pdf
demo_fas [demo]_[fas] (optional - concatenated with product) us		(optional - concatenated with product) used for QWI to
		distinguish separate tabulations from the legal [demo]_[fas]
		combinations as described in lehd_csv_naming.pdf
geo	[stusps] or METRO	Covered [geography] (uppercase state postal code including
		entire nation or the word METRO)
geonum	[geography]	Numeric geography code
start	уууу:q	Start year and quarter
end	уууу:q	End year and quarter
schema	Vx.y.z	Version of the schema
release	RyyyyQq	Release quarter (identifies when the data was created)
internal	various	Internal identifier used for provenance tracking

For instance, the metadata for the R2017Q3 release of Missouri J2J tabulations (obtained from here) has the following content:

J2J MO 29 2000:2-2016:3 V4.2b-draft R2017Q3 j2jpu_mo_20171023_1412

Some J2J metadata may contain multiple lines, as necessary.

8.2 Additional Metadata for J2JOD Files (avail.csv)

(variables_avail.csv)

Because the origin-destination (J2JOD) data link two regions, we provide an auxiliary file with the time range that cells containing data for each geographic pairing may appear in a data release.

variable	type	label
geo_level	Char(1)	Geographic level of destination region
geography	Char(8)	Geography code of destination region
geo_level_orig	Char(1)	Geographic level of origin region
geography_orig	Char(8)	Geography code of origin region
start_year	Num	First year regional pair may be observed
start_quarter	Num	First quarter regional pair may be observed
end_year	Num	Last year regional pair may be observed
end_quarter	Num	Last quarter regional pair may be observed

The reference region will always be either the origin or the destination. National tabulations contain records where both origin and destination are geo_level=N; state tabulations contain records where geo_level in (N,S); metro tabulations contain records where geo_level in (N,S,B). Data may be suppressed for certain combinations of regions and quarters because the estimates do

not meet Census Bureau publication standards.

8.3 Metadata on Indicator Availability

(variables_lags.csv)

Each Indicator potentially requires leads and/or lags of data to be computed, and thus may not be available for certain time periods. The date range for J2J and J2JR can be found in version.txt; the date range for J2JOD can be found in avail.csv.

For each indicator, the following files contain the quarters of data required to be available relative to the overall date range described in the metadata for the release:

- lags_j2j.csv
- lags_j2japp.csv

The files are structured as follows:

variable	type	label
Indicator Variable	Num	Name of the Indicator
Quarters_Required_Prior Num		Number of quarters of data required to
		compute indicator relative to start quarter
Quarters_Required_Sub	se Nuem t	Number of quarters of data required to
		compute indicator relative to end quarter

This revision: Wed Dec 20 18:21:56 EST 2017