

# LEHD Public Use Schema documents - File and Directory Naming Convention

Lars Vilhuber

28 February 2019

# Table of Contents

- 1. Purpose ..... 1
- 2. Schema for Data File Contents..... 1
- 3. Extends ..... 1
- 4. Supersedes ..... 1
- 5. Basic Filename Schema ..... 1
  - 5.1. QWIPU from the LED Extraction Tool ..... 2
  - 5.2. Metadata for QWI and J2J data files (version.txt) ..... 2
  - 5.3. Metadata for J2JOD files ..... 3
- 6. Directory Paths ..... 3
- 7. Description of Naming Components ..... 3
  - 7.1. Release ..... 3
  - 7.2. Types and Products..... 3
  - 7.3. geohi..... 5
  - 7.4. demo..... 5
  - 7.5. fas ..... 7
  - 7.6. geocat ..... 8
  - 7.7. indcat ..... 9
  - 7.8. owncat ..... 10
  - 7.9. sa..... 11
  - 7.10. ext..... 12
- 8. Changes ..... 13
  - 8.1. This version (revisions)..... 13
  - 8.2. Version 4.3.1 from 4.3.0..... 13
  - 8.3. Version 4.3.0 from 4.2.0..... 13

( [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](mailto:lars.vilhuber@cornell.edu).

## 1. Purpose

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/> and through LED Extraction Tool at <http://ledextract.ces.census.gov/>.

This document describes the file and directory naming schema for LEHD-provided CSV files.

## 2. Schema for Data File Contents

The contents (schema) for LEHD data files are described in [lehd\\_public\\_use\\_schema.pdf](#). The contents (schema) for shapefiles are described in [lehd\\_shapefiles.pdf](#).

## 3. Extends

This version modifies a portion of the structure of the metadata. Many files compliant with LEHD or QWI Schema v4.2 or lower will also be compliant with this schema, but compatibility is not guaranteed.

## 4. Supersedes

This version supersedes V4.2, for files released as of R2018Q3

## 5. Basic Filename Schema

All files are preceded by a file type definition, followed by additional information on the aggregation level of the file, or some other identifier.

```
[TYPE]_[DETAILS].[EXT]
```

( [naming\\_convention.csv](#) )

| type   | Used By                            | Description   |
|--|------------------------------------|---|
| [type]_[id]  | qwix                               | CSV files queried from the LED Extraction Tool at <a href="http://ledextract.ces.census.gov/">http://ledextract.ces.census.gov/</a> . |
| [type]_[geohi]_[d]_[emo]_[fas]_[geocat]_[indcat]_[ownercat]_[sa] | qwi qwir qwiv qwivr j2j j2jr j2jod | Common convention for currently provided files.   |
| lehd_shp_[geocat].zip  | shapefiles                         | ZIP files containing data for shapefiles  |
| version_[type].txt   | metadata                           | Machine-readable text files with metadata on geographic and temporal coverage   |
| j2jod_[geography]_[avail].csv                                    | metadata                           | Complex metadata (used only in J2JOD) - see below   |

## 5.1. QWIPU from the LED Extraction Tool

Files downloaded through the LED Extraction Tool at <http://ledextract.ces.census.gov/> follow the following naming convention:

```
[type]_[id].[EXT]
```

where [id] is the Request ID (a unique string of characters) generated every time "Submit Request" is clicked. The ID references each query submission made to the database.

## 5.2. Metadata for QWI and J2J data files (version.txt)

Metadata accompanies the data files, identifying provenance, geographic and temporal coverage. These files follow the following naming convention:

```
version_[type].txt
```

where each name component is described in more detail [below](#).

## 5.3. Metadata for J2JOD files

Because the origin-destination (J2JOD) data link two regions, we provide an auxiliary file with the time range for which cells containing data for each geographic pairing may appear in a data release. The reference region will always be either the origin or the destination. These files follow the following naming convention:

```
j2jod_[geography]_avail.csv
```

where each name component is described in more detail [below](#).

## 6. Directory Paths

Downloadable data files are organized on the download server (at the time of this writing: <https://lehd.ces.census.gov/data/>) in directories organized as follows:

```
SERVER/data/[PRODUCT]/[RELEASE]/[GEOHI][ /[TYPE] ]/{files}
```

where [\[PRODUCT\]](#), [\[RELEASE\]](#), [\[GEOHI\]](#), and [\[TYPE\]](#) are each described below. [\[TYPE\]](#) is optional.

## 7. Description of Naming Components

### 7.1. Release

A **release** is defined by the calendar year quarter when data production occurs. It is thus generically constructed as

```
R[YEAR]Q[QUARTER]
```

where [\[YEAR\]](#) is the 4-digit year and [\[QUARTER\]](#) the single-digit calendar year quarter (1-4).

### 7.2. Types and Products

( [naming\\_type.csv](#) )

| type  | product | Description   | Explanation   | url   |
|-------|---------|---|---|---|
| j2j   | j2j     | Job-to-Job Flows Counts (J2J)                                 | Counts of hires and separations, to and from employment and non-employment, summed across all establishments  | <a href="http://lehd.ces.census.gov/data/#j2j">http://lehd.ces.census.gov/data/#j2j</a>     |
| j2jod | j2j     | Job-to-Job Flows Origin-Destination                           | Origin and destination flows and earnings, tabulated by the destination establishment   | <a href="http://lehd.ces.census.gov/data/#j2j">http://lehd.ces.census.gov/data/#j2j</a>     |
| j2jr  | j2j     | Job-to-Job Flows Rates  | Hiring and separation rates, to and from employment and nonemployment, summed across all establishments   | <a href="http://lehd.ces.census.gov/data/#j2j">http://lehd.ces.census.gov/data/#j2j</a>     |
| lodes | lodes   | LEHD Origin-Destination Employment Statistics (LODES)         | LEHD Origin-Destination Employment Statistics (LODES) used by OnTheMap. Version 7 and higher of LODES was enumerated by 2010 census blocks. Version 6 and lower of LODES were enumerated with 2000 census blocks.   | <a href="http://lehd.ces.census.gov/data/#lodes">http://lehd.ces.census.gov/data/#lodes</a> |
| qwi   | qwi     | Quarterly Workforce Indicators Public Use Counts (QWIPU)      | The Quarterly Workforce Indicators (QWI) are a set of economic indicators including employment, job creation, earnings, and other measures of employment flows. The QWI are reported using detailed firm characteristics (geography, industry, age, size) and worker demographics information (sex, age, education, race, ethnicity). | <a href="http://lehd.ces.census.gov/data/#qwi">http://lehd.ces.census.gov/data/#qwi</a>     |
| qwir  | qwi     | Quarterly Workforce Indicators Public Use Rates               | Rates for flow variables in the Quarterly Workforce Indicators (QWI)  | tbd   |
| qwirv | qwi     | Quarterly Workforce Indicators Variability Measures for Rates | Variability measures for rates in the Quarterly Workforce Indicators (QWI)  | tbd   |

| type     | product | Description   | Explanation  | url                                  |
|----------|---------|---|--|--------------------------------------|
| qwiv     | qwi     | Quarterly Workforce Indicators Variability Measures | Variability measures for Quarterly Workforce Indicators (QWI)                                | tbd                                  |
| lehd_shp | all     | LEHD Shapefiles                                     | Shapefiles are used to provide mapping functionality in QWI Explorer and Job-to-Job Explorer | <a href="#">lehd_shapefiles.html</a> |

## 7.3. geohi

( [naming\\_geohi.csv](#) )

The geohi component in **filenames** is based on one of two possible code sets:

- the lower-case alphabetic FIPS or postal state code based on [FIPS PUB 5-2](#) (see also [label\\_stusps.csv](#) for upper-case variants).
- the numeric CBSA code corresponding to the metropolitan areas (see [label\\_geography\\_metro.csv](#))
- It is expanded to encompass additional codes denoting national coverage, or a collection of states.

For **directories**, files at the CBSA level are collected under a single **metro** directory.

| type    | Description   |
|---------|---|
| all     | Collection of all available states  |
| us      | National data (50 states + DC)  |
| metro   | Indicates collection of CBSA-level files ( <b>directory names only</b> )        |
| [st]    | Any legal 2-character lower-case state postal code                              |
| [NNNNN] | Any valid CBSA-derived code listed in <a href="#">label_geography_metro.csv</a> |

## 7.4. demo

( [naming\\_demo.csv](#) )

| type | Description                   |
|------|-------------------------------|
| d    | No demographic detail         |
| rh   | Race by Ethnicity tabulations |

| <b>type</b> | <b>Description</b>                 |
|-------------|------------------------------------|
| sa          | Sex by Age tabulations             |
| se          | Sex by Education tabulations       |
| sarhe       | Collection of sa-rh-se tabulations |



## 7.5. fas

( [naming\\_fas.csv](#) )

| <b>type</b> | <b>Description</b>         |
|-------------|----------------------------|
| f           | No firm size or age detail |
| fa          | Tabulations by firm age    |
| fs          | Tabulations by firm size   |

## 7.6. geocat

( [naming\\_geocat.csv](#) )

| <b>type</b> | <b>Description</b>                      |
|-------------|---|
| gb          | Metropolitan (complete)                 |
| gc          | Counties                                |
| gm          | Metropolitan/Micropolitan (state parts) |
| gn          | National (50 States + DC)               |
| gs          | States                                  |
| gw          | Workforce Investment Areas              |

## 7.7. indcat

( [naming\\_indcat.csv](#) )

| <b>type</b> | <b>Description</b>    |
|-------------|-----------------------|
| n           | all industries        |
| ns          | NAICS sectors         |
| n3          | NAICS subsectors      |
| n4          | NAICS industry groups |

## 7.8. owncat

( [naming\\_owncat.csv](#) )

| <b>type</b> | <b>Description</b>  |
|-------------|---|
| of          | Federal government (QWI Code A01)                             |
| op          | All Private (QWI Code A05)                                    |
| oslp        | State, local, and private ownership categories (QWI Code A00) |

## 7.9. sa

( [naming\\_sa.csv](#) )

| <b>type</b> | <b>Description</b>      |
|-------------|-------------------------|
| s           | Seasonally adjusted     |
| u           | Not seasonally adjusted |

## 7.10. ext

( [naming\\_ext.csv](#) )

| <b>type</b> | <b>Description</b>                          |
|-------------|---|
| csv         | Comma-Separated file                        |
| pdf         | PDF file (printable)                        |
| xls         | Excel file (only select output)             |
| xlsx        | Excel file (not available for larger files) |

# 8. Changes

For a description of how schema files are versioned, see [main directory](#).

## 8.1. This version (revisions)

- 2018-10-30: Initial release
- 2019-02-28: Switch to different rendering engine, only stylistic changes

## 8.2. Version 4.3.1 from 4.3.0

- 2018-10-30: Due to a change of the geographic boundaries for a WIB in Iowa, the shape files need to be updated.
- Adjust the "supersedes" definition to correctly reference the prior version.

## 8.3. Version 4.3.0 from 4.2.0

- Changed the format of the version.txt file. This does not affect data files, but it does affect the metadata, and thus requires an increase in the minor version.
- Changed the description of the download locations of the QWI files, as part of the "naming" schema.



*Important*

Some of the data products noted above do not exist yet.

This revision: Thu Feb 28 21:17:02 UTC 2019