

## OnTheMap Data Structure Format Version 4.1

### Introduction

OnTheMap data is released both as part of the OnTheMap application and in its raw form as a set of comma separated variable (CSV) text files. This document describes the structure of those raw files and provides basic information for users who want to perform analytical work on the data outside of the OnTheMap application.

### File Organization

The CSV data files are released at the state level and they are organized into three groups within each state. The three groups of files are named as follows, according to their contents:

**OD** – Origin-Destination data, jobs totals are associated with both a home Census Block and a work Census Block

**RAC** – Residence Area Characteristic data, jobs are totaled by home Census Block

**WAC** – Workplace Area Characteristic data, jobs are totaled by work Census Block

At the root distribution level, each state has a directory named by its lowercase, 2-letter postal code. Within each state directory are three directories, one for each of the files groups described above. Also within the state directory are two text files. One is a version document, called “version.txt”.

The version file contains information about the state, the data vintage, and the format version. The “data vintage” is an eight character code based on the date the data was prepared for release; it is the form of YYYYMMDD.<sup>1</sup> New or corrected data may cause newer vintages of data to be released. Only files that have new or changed data will be included in future vintages. The “format version” is a number that refers to file organization and structure of the data release and is fully defined by this document. Restructuring of the existing data into different file formats or a different organization structure would result in a change in the format version.

The other file in each state directory is an MD5SUM file, called “otm\_[ST].md5sum” where [ST] is the respective state postal code. With the correct software tools, the MD5SUM file can be used to check that the individual data files were downloaded without any file corruption or loss of data. Further information on MD5SUM can be found at <http://en.wikipedia.org/wiki/Md5sum>.

A typical directory tree looks as follows:

```
\
|-- ak
|   |-- od
|   |   |-- (individual origin-destination data files)
|   |-- rac
|   |   |-- (individual residence area characteristics data files)
|   |-- wac
```

---

<sup>1</sup> Note: The data vintage is *not* the same as the createdate provided within the data files. The createdate refers to an internal processing point, while the data vintage is the code that refers to a specific release of public OnTheMap data.

```
| | `-- (individual workplace area characteristics data files)
| | |-- otm_ak.md5sum
| | `-- version.txt
|-- az
|
(more states)
```

The CSV files are compressed using a GZip algorithm, which appends a “.gz” to the end of the filename. Current compression/extraction software packages should be able to access these files.

### **Data Coverage and Availability**

Data is currently available for most states in most years. The LED Partnership consists of 51 State and Territorial Partners, of which 47 are producing data for OnTheMap. With the exceptions listed below all 47 states have OD, RAC, and WAC data for all 7 years of data (2002-2008).

States and territories which are not in “regular production” do have RAC datasets available, however these only include jobs for workers employed in the 47 OnTheMap states. The partners that are *not* in regular production are Connecticut, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. Massachusetts and New Hampshire are not yet part of the LED Partnership.

Due to historical availability or data quality, a handful of state-year combinations of data are *not* part of OnTheMap. These are Arizona (2002 and 2003), Arkansas (2002), Mississippi (2002 and 2003), and North Carolina (2008).

For a state-year with a full complement of data, there are 8 OD files, 40 RAC files, and 40 WAC files.

For more information on the LED partnership, please visit <http://lehd.did.census.gov/>.

### **Format Version**

With the release of data accompanying Version 4 of the OnTheMap application, the format version of the data release is being re-baselined. This document (format version 4.0) describes the structure for the data released along with the following OnTheMap Application versions:

Application Version 3.0 – Released September 4, 2008  
Application Version 3.1 – Released September 11, 2008  
Application Version 3.2 – Released December 15, 2008  
Application Version 4.0 – Released December 14, 2009  
Application Version 4.1 – Released May 27, 2010

For further information on the background and release history of OnTheMap data, please visit <http://lehd.did.census.gov/led/datatools/onthemap4.html>

### **File Naming Conventions and File Structure**

The structure of the three types of files are described on the following pages:

## **OD**

Filename of the OD datasets are described by the following templates

[ST]\_od\_[PART]\_[TYPE]\_[YEAR]\_1.csv.gz where

[ST] = lowercase, 2-letter postal code for a chosen state

[PART] = Part of the state file, can have a value of either “main” or “aux”. Complimentary parts of the state file, the main part includes jobs with both workplace and residence in the state and the aux part includes jobs with the workplace in the state and the residence outside of the state.

[TYPE] = Job Type, can have a value of “ja” for All Jobs, “jp” for Primary Jobs, “pa” for All Private Jobs, or “pp” for Private Primary Jobs.

[YEAR] = Year of job data. Can have the value of 2002-2008 for most states.

As an example the main OD file of Primary Jobs in 2007 for California would be the file:

ca\_od\_main\_jp\_2007\_1.csv.gz

The structure of the OD files is as follows:

<b>Origin-Destination (OD) File Structure</b>				
<b>Pos</b>	<b>Variable</b>	<b>Type</b>	<b>Length</b>	<b>Explanation</b>
1	w_geocode	Char	15	Workplace Census Block Code
2	h_geocode	Char	15	Residence Census Block Code
3	total	Num	8	Total number of jobs
4	age1	Num	8	Number of jobs of workers age 29 or younger
5	age2	Num	8	Number of jobs for workers age 30 to 54
6	age3	Num	8	Number of jobs for workers age 55 or older
7	earn1	Num	8	Number of jobs with earnings \$1250/month or less
8	earn2	Num	8	Number of jobs with earnings \$1251/month to \$3333/month
9	earn3	Num	8	Number of jobs with earnings greater than \$3333/month
10	ind1	Num	8	Number of jobs in Goods Producing industry sectors
11	ind2	Num	8	Number of jobs in Trade, Transportation, and Utilities industry sectors
12	ind3	Num	8	Number of jobs in All Other Services industry sectors
13	createdate	Char	8	Date on which data was created, formatted as YYYYMMDD

## RAC

Filenames of the RAC datasets are described by the following template:

[ST]\_rac\_[SEG]\_[TYPE]\_[YEAR]\_1.csv.gz where

[ST] = lowercase, 2-letter postal code for a chosen state

[SEG] = Segment of the workforce, can have the values of “total”, “age1”, “age2”, “age3”, “earn1”, “earn2”, “earn3”, “ind1”, “ind2”, or “ind3”. These correspond to the same segments of the workforce as are listed in the OD file structure above.

[TYPE] = Job Type, can have a value of “ja” for All Jobs, “jp” for Primary Jobs, “pa” for All Private Jobs, or “pp” for Private Primary Jobs.

[YEAR] = Year of job data. Can have the value of 2002-2008 for most states.

As an example the RAC file of All Jobs for workers age 29 or younger in 2004 for New Jersey would be the file: nj\_rac\_age1\_ja\_2004\_1.csv.gz

Residence Area Characteristics (RAC) File Structure				
Pos	Variable	Type	Length	Explanation
1	h_geocode	Char	15	Residence Census Block Code
2	total	Num	8	Total number of jobs
3	age1	Num	8	Number of jobs of workers age 29 or younger
4	age2	Num	8	Number of jobs for workers age 30 to 54
5	age3	Num	8	Number of jobs for workers age 55 or older
6	earn1	Num	8	Number of jobs with earnings \$1250/month or less
7	earn2	Num	8	Number of jobs with earnings \$1251/month to \$3333/month
8	earn3	Num	8	Number of jobs with earnings greater than \$3333/month
9	naics_s01	Num	8	Number of jobs in NAICS sector 11 (Agriculture, Forestry, Fishing and Hunting)
10	naics_s02	Num	8	Number of jobs in NAICS sector 21 (Mining, Quarrying, and Oil and Gas Extraction)
11	naics_s03	Num	8	Number of jobs in NAICS sector 22 (Utilities)
12	naics_s04	Num	8	Number of jobs in NAICS sector 23 (Construction)
13	naics_s05	Num	8	Number of jobs in NAICS sector 31-33 (Manufacturing)
14	naics_s06	Num	8	Number of jobs in NAICS sector 42 (Wholesale Trade)
15	naics_s07	Num	8	Number of jobs in NAICS sector 44-45 (Retail Trade)
16	naics_s08	Num	8	Number of jobs in NAICS sector 48-49 (Transportation and Warehousing)
17	naics_s09	Num	8	Number of jobs in NAICS sector 51 (Information)
18	naics_s10	Num	8	Number of jobs in NAICS sector 52 (Finance and Insurance)
19	naics_s11	Num	8	Number of jobs in NAICS sector 53 (Real Estate and Rental and Leasing)
20	naics_s12	Num	8	Number of jobs in NAICS sector 54 (Professional, Scientific, and Technical Services)
21	naics_s13	Num	8	Number of jobs in NAICS sector 55 (Management of Companies and Enterprises)
22	naics_s14	Num	8	Number of jobs in NAICS sector 56 (Administrative and Support and Waste Management and Remediation Services)
23	naics_s15	Num	8	Number of jobs in NAICS sector 61 (Educational Services)
24	naics_s16	Num	8	Number of jobs in NAICS sector 62 (Health Care and Social Assistance)
25	naics_s17	Num	8	Number of jobs in NAICS sector 71 (Arts, Entertainment, and Recreation)
26	naics_s18	Num	8	Number of jobs in NAICS sector 72 (Accommodation and Food Services)
27	naics_s19	Num	8	Number of jobs in NAICS sector 81 (Other Services [except Public Administration])
28	naics_s20	Num	8	Number of jobs in NAICS sector 92 (Public Administration)
29	createdate	Char	8	Date on which data was created, formatted as YYYYMMDD

## WAC

Filenames of the WAC datasets are described by the following template:

[ST]\_wac\_[SEG]\_[TYPE]\_[YEAR]\_1.csv.gz where

[ST] = lowercase, 2-letter postal code for a chosen state

[SEG] = Segment of the workforce, can have the values of “total”, “age1”, “age2”, “age3”, “earn1”, “earn2”, “earn3”, “ind1”, “ind2”, or “ind3”. These correspond to the same segments of the workforce as are listed in the OD file structure above.

[TYPE] = Job Type, can have a value of “ja” for All Jobs, “jp” for Primary Jobs, “pa” for All Private Jobs, or “pp” for Private Primary Jobs.

[YEAR] = Year of job data. Can have the value of 2002-2008 for most states.

As an example the WAC file of Private Primary Jobs for workers earning more that \$3333/mo in 2008 for Iowa would be the file: ia\_wac\_earn3\_pp\_2008\_1.csv.gz

Workplace Area Characteristics (WAC) File Structure				
Pos	Variable	Type	Length	Explanation
1	w_geocode	Char	15	Workplace Census Block Code
2	total	Num	8	Total number of jobs
3	age1	Num	8	Number of jobs of workers age 29 or younger
4	age2	Num	8	Number of jobs for workers age 30 to 54
5	age3	Num	8	Number of jobs for workers age 55 or older
6	earn1	Num	8	Number of jobs with earnings \$1250/month or less
7	earn2	Num	8	Number of jobs with earnings \$1251/month to \$3333/month
8	earn3	Num	8	Number of jobs with earnings greater than \$3333/month
9	naics_s01	Num	8	Number of jobs in NAICS sector 11 (Agriculture, Forestry, Fishing and Hunting)
10	naics_s02	Num	8	Number of jobs in NAICS sector 21 (Mining, Quarrying, and Oil and Gas Extraction)
11	naics_s03	Num	8	Number of jobs in NAICS sector 22 (Utilities)
12	naics_s04	Num	8	Number of jobs in NAICS sector 23 (Construction)
13	naics_s05	Num	8	Number of jobs in NAICS sector 31-33 (Manufacturing)
14	naics_s06	Num	8	Number of jobs in NAICS sector 42 (Wholesale Trade)
15	naics_s07	Num	8	Number of jobs in NAICS sector 44-45 (Retail Trade)
16	naics_s08	Num	8	Number of jobs in NAICS sector 48-49 (Transportation and Warehousing)
17	naics_s09	Num	8	Number of jobs in NAICS sector 51 (Information)
18	naics_s10	Num	8	Number of jobs in NAICS sector 52 (Finance and Insurance)
19	naics_s11	Num	8	Number of jobs in NAICS sector 53 (Real Estate and Rental and Leasing)
20	naics_s12	Num	8	Number of jobs in NAICS sector 54 (Professional, Scientific, and Technical Services)
21	naics_s13	Num	8	Number of jobs in NAICS sector 55 (Management of Companies and Enterprises)
22	naics_s14	Num	8	Number of jobs in NAICS sector 56 (Administrative and Support and Waste Management and Remediation Services)
23	naics_s15	Num	8	Number of jobs in NAICS sector 61 (Educational Services)
24	naics_s16	Num	8	Number of jobs in NAICS sector 62 (Health Care and Social Assistance)
25	naics_s17	Num	8	Number of jobs in NAICS sector 71 (Arts, Entertainment, and Recreation)
26	naics_s18	Num	8	Number of jobs in NAICS sector 72 (Accommodation and Food Services)
27	naics_s19	Num	8	Number of jobs in NAICS sector 81 (Other Services [except Public Administration])
28	naics_s20	Num	8	Number of jobs in NAICS sector 92 (Public Administration)
29	createdate	Char	8	Date on which data was created, formatted as YYYYMMDD