

On The Map

Data Available from the Cornell Institute for Social and Economic Research

Local Employment Dynamics

All About Jobs

On The Map (<http://lehd.dsd.census.gov/led/datatools/onthemap.html>) has been a very popular web-based application since its formal launch under the Local Employment Dynamics partnership in February 2006. Many inquiries have been made to access, use, and evaluate the underlying data, and this interest is growing. The Cornell Institute for Social and Economic Research (CISER) has agreed to provide access and download capability for the underlying **On The Map** public-use data (OTM data) at http://vrdc.ciser.cornell.edu/news/?page_id=205. This is a joint project between the Census Bureau, participating states, and Cornell University that is partly supported by the National Science Foundation.

Background on synthetic data

On The Map is the first synthetic data product released by the Census Bureau. Ten (10) distinct copies of the synthetic data, known as implicates, are produced; one of them is used for the current implementation of **On The Map**. The CISER site provides an Internet-accessible computing environment dedicated to the exploration and development of synthetic data. Background on the development and use of synthetic data may be found in, for example:

John Abowd. [Synthetic Data: A New Future for Public Use Micro-Data?](http://lehd.dsd.census.gov/led/library/presentations/Synthetic-Data-Census-20041207.pdf) PN-2004-01. (<http://lehd.dsd.census.gov/led/library/presentations/Synthetic-Data-Census-20041207.pdf>).

John Abowd and Julia Lane. [New Approaches to Confidentiality Protection: Synthetic Data, Remote Access and Research Data Centers](http://lehd.dsd.census.gov/led/library/techpapers/tp-2004-03.pdf), in Domingo-Ferrer, J. & Torra, V. (ed.), *Privacy in Statistical Databases*, Springer, Berlin-New York, 2004 (also available as LEHD Technical paper TP-2004-03. <http://lehd.dsd.census.gov/led/library/techpapers/tp-2004-03.pdf>)

John Abowd and Julia Lane. [Synthetic Data and Confidentiality Protection](http://lehd.dsd.census.gov/led/library/techpapers/tp-2003-10.pdf), TP-2003-10. (<http://lehd.dsd.census.gov/led/library/techpapers/tp-2003-10.pdf>)

What can be downloaded and accessed

OTM data can be accessed on the CISER site by logging on to a registered account and with a password. Registration is required, but is free and available to all. Once logged on, the data can be analyzed either in the CISER site, or downloaded in CSV format. Data are stored under */mixed/onthemap*; read-in programs for SAS and Stata can be found under */mixed/onthemap/programs* or on the web at <http://vrdc.ciser.cornell.edu/onthemap/>. The data package consists of Origin-Destination data (OD), Residence Area Characteristics data (RAC), Workplace Area Characteristics

data (WAC), and block-group level Quarterly Workforce Indicator data (QWI). Oregon and Texas have agreed to make all 10 implicates of their 2002-2003 data available at the launch of *OTM data* on the CISER site. Data for additional states and years will be added, pending on the voluntary participation of the state partners.

The place of residence counts are generated from a synthetic data model that conditions on disclosure-proofed place of work counts and other observable characteristics. Each of the 10 *implicate* files of the OD and the RAC represents an independent draw from the synthetic data model. Detailed information on the full *OTM data* and the synthetic data model can be found at http://vrdc.ciser.cornell.edu/onthemap/doc/otm_public_master.pdf.

The process of getting an account and access

The CISER site is open to all interested users subject to registration. There is no project approval process. In order to access the *OTM data*, interested data users must apply for an account by contacting the CISER administrators (virtualrdc@cornell.edu). Login is done using Secure Shell (SSH) tools (for a command line interface), and one of two optional graphical interfaces. Download of data using a web browser is NOT supported at this time. Details on access methods are provided at <http://vrdc.ciser.cornell.edu/onthemap/doc>.

What we ask the registered users to do

We ask the registered users to share their comments, observations, feedback, analysis, findings, and/or published papers by sending emails to 3 listservs:

led-qwi@lists.census.gov (QWI user community)
lehd-ltd@lists.census.gov (Local Transportation Dynamics community under development), and
ctpp_news@chrispy.net (Census Transportation Planning Package community)

The research and evaluation results will be used to enhance understanding and developmental efforts for future versions of *On The Map* and synthetic data in general.

Technical requirements to access and analyze data

Registered users may access and analyze data either (a) directly at the CISER site or (b) by downloading the data to their own computing systems. In either situation, the users must have

- an adequately powerful Internet-connected computer, and
- SSH client software.

VNC or NX client software may be optionally installed for graphical interface, and software capable of SFTP file transfer will be needed for data download. Depending on the actual use, the required memory, storage, and download speed may be significant due to the large size of the data files.

All required access software at the CISER site is open-source and free of charge, see the [online VirtualRDC guide](http://vrdc.ciser.cornell.edu/news/?p=13) (<http://vrdc.ciser.cornell.edu/news/?p=13>) for download links and detailed instructions. Windows and Linux clients are supported, but all software is known to work on Mac OS X as well. See the [Installed software page](http://vrdc.ciser.cornell.edu/news/?page_id=83) (http://vrdc.ciser.cornell.edu/news/?page_id=83) for a partial list of statistical software that is provided free-of-charge for research and evaluation use on the CISER site. Once logged on to the CISER site, you will find the data and some useful SAS and Stata programs under /mixed/onthemap/. Analysis software on the CISER site may only be used for non-proprietary scientific research. For all other purposes, users should download a copy of the *OTM data* to a local computer on which they have licensed the appropriate analysis software.

Where to get help

Users will be solely responsible for their hardware and software needs once they download the data from the CISER site. For further information and assistance at the CISER site, contact the [VirtualRDC administrators](mailto:virtualrdc@cornell.edu) (virtualrdc@cornell.edu).

What's Next?

Data for additional ***On The Map*** states may be added depending on the response and level of interest from the users and the voluntary participation of the states.

In addition to the CISER site, ***On The Map*** public use data may be obtained directly from the states or the Census Bureau.

Funding and disclaimers

The Cornell Institute for Social and Economic Research is not affiliated with the U.S. Census Bureau. All Census Bureau data made available at this facility are public-use data and have been reviewed and approved by the Census Bureau's Disclosure Review Board prior to release. The CISER Internet site is partially funded by commercial sources and the National Science Foundation through the following grants:

[#0427889](http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0427889) (<http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0427889>),
[#0339191](http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0339191) (<http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0339191>) and
[#9978093](http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=9978093) (<http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=9978093>).