

LED: Data Visualization Plan

August 5, 2010

Data visualization tools have emerged since 2007 as important new ways to present statistical information to make it more understandable for data users. In 2006 the Local Employment Dynamics (LED) partnership pioneered a data visualization effort with the first release of its web-based OnTheMap tool, which graphically illustrates the relationship between jobs and workplaces on user-defined maps. Today, unlocking public access to government data is considered a critical component of the President's [Open Government Initiative](#).¹

Improving LED Data Visualization is one of 12 projects undertaken by the U.S. Census Bureau to increase government transparency by publishing high-value information. This project is part of the U.S. Department of Commerce's 2010 [Open Government Plan](#)² to improve understanding and use of Census Bureau data.

The LED data visualization plan describes the LED partnership, and outlines the basic LED data and products, operating principles, and LED's short-term plan of action for implementing the appropriate data visualization tools in a rapidly changing environment.

In particular, the 2010-2011 LED data visualization plan includes three primary components:

1. Implementing data visualization tools to illuminate the [Quarterly Workforce Indicator time series data](#),³ which describe the dynamics of local employment.
2. Enhancing [OnTheMap](#)⁴ with expanded content and features to describe the dynamics of job and worker locations, as well as developing derivative products for custom use.
3. Exploring existing or new tools for a third LED data line on the dynamics of job-to-job flows, and improving the interoperability between LED data lines.

This plan will be reviewed and updated periodically.

Comments about this plan should be sent to <CES.Local.Employment.Dynamics@census.gov>.

¹ Available at <http://www.whitehouse.gov/open> on July 25, 2010.

² Available at <http://open.commerce.gov/iii-transparency> on July 25, 2010.

³ Available at <http://lehd.did.census.gov/led/datatools/qwi-online.html> on July 25, 2010.

⁴ Available at <http://lehd.did.census.gov/led/datatools/onthemap4.html> on July 25, 2010.

THE LOCAL EMPLOYMENT DYNAMICS (LED) PARTNERSHIP

The Census Bureau has been building the voluntary federal-state LED partnership with state workforce agencies since 2000.

Through data use agreements, the partner states supply the Census Bureau with their historical and continuing quarterly Unemployment Insurance worker wage records and Quarterly Census of Earnings and Wages (QCEW, formerly known as ES-202) establishment-level records. According to [the Bureau of Labor Statistics](#),⁵ QCEW covers about 98 percent of all U.S. jobs. The partner's administrative records are then integrated with additional administrative records, decennial and economic censuses, surveys, and data sources that are available to the Census Bureau to create a longitudinal national frame of jobs.

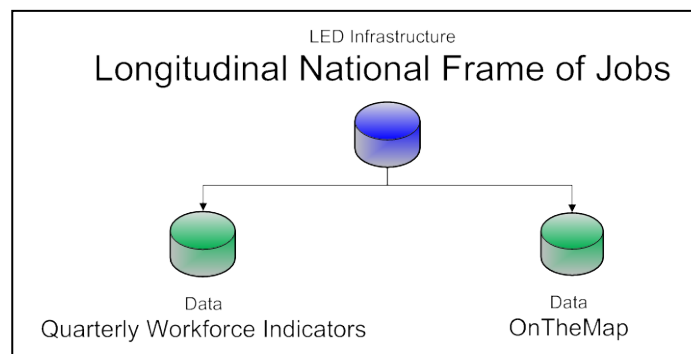
Today there are 51 LED partners, covering 48 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands. The two remaining states, Massachusetts and New Hampshire, are in the process of joining LED. Institutionalizing LED as a Census Bureau program is also part of the Department of Commerce's Open Government Plan.

LED DATA AND VISUALIZATION

There are over 140 million workers and about 8 million employers in the United States. The LED infrastructure does not currently include Federal civilian workers, the uniformed military, and the self-employed. Integration of data on Federal civilian workers and the self-employed is part of the plan to implement the LED program.

New data and products are in turn created from this rich and emerging statistical system to provide better understanding of the U.S. labor market and economy with unprecedented details, while still strictly protecting the confidentiality of the individual and business respondents.

The two established LED data lines are structured as follows.



Quarterly Workforce Indicators (QWI). The QWI were first produced in 2003 to describe the dynamics of local employment. Thirty QWI have been updated quarterly as time series for each partner state by

- **worker age**, covering the 9 age groups of 14-18, 19-21, 22-24, 25-34, 35-44, 45-54, 55-64, 65-99, and 14-99
- **worker gender**, covering female, male, and both genders

⁵ Available at <http://www.bls.gov/cew/> on July 25, 2010.

- **industry**, covering 2-, 3-, and 4-digit [North America Industry Classification System](#)⁶ (NAICS) industries
- **job type**, covering private sector and all sectors
- **workplace geography**, covering states, counties, metropolitan areas, and workforce investment areas

The file size for the public-use QWI files of 48 state partners is large and growing, accounting for about 2 terabytes uncompressed for the second quarter of 2010. A web-based query system, [QWI Online](#),⁷ was introduced in 2004 to provide point-in-time tabular results for 8 QWIs based on user selections. In general, visualization of QWI, either as point-in-time data or as a time series, has been limited due to the size and complexity of the data.

OnTheMap. First produced in 2006 to describe the dynamics of location of jobs and workers, [OnTheMap data](#)⁸ includes three primary files that are updated and added annually, beginning with the year 2002:

- **residential area characteristics**, covering the number of jobs or workers living in each Census block (up to 8.2 million) by
 - job type (all jobs, primary jobs, all private jobs, and private primary jobs)
 - age group (age 29 or younger, 30 to 54, and age 55 or older)
 - earnings group (\$1,250 per month or less, \$1,251 to \$3,333 per month, more than \$3,333 per month)
 - 2-digit NAICS industry group
 - segment of the workforce (total, 3 age groups, 3 earnings groups, and 3 industry groups)
- **workplace area characteristics**, covering the number of jobs or workers working in each Census block by the same job type, age group, earnings group, industry group, and segment of the workforce
- **origin-destination**, covering the number of jobs or workers for each pair of residential Census block and workplace Census block by the same age group, earnings group, and reduced industry group (goods producing; trade, transportation, utilities; and all other)

The latest version of the public-use OnTheMap data, released in 2010, includes 47 states and 7 years of annual data from 2002 to 2008. The total uncompressed file size is over 250 gigabytes and will continue to grow with each annual release. To help with visualizing the data, a web-based mapping and reporting tool also known as [OnTheMap](#)⁹ is released simultaneously with the data. The creative approach to producing the OnTheMap data was recognized in the [2008 Economic Report of the President](#),¹⁰ and the OnTheMap visualization tool was featured as a major U.S. statistical innovation in the [United Nations Statistical Commission](#)¹¹ in 2009.

⁶ Available at <http://www.census.gov/eos/www/naics/> on July 25, 2010.

⁷ Available at <http://lehd.did.census.gov/led/datatools/qwiapp.html> on July 25, 2010.

⁸ Available at <http://lehd.did.census.gov/led/onthemap/OnTheMapDataTechDoc4.1.pdf> on July 28, 2010.

⁹ Available at <http://lehdmap4.did.census.gov/themap4/> on July 25, 2010.

¹⁰ Available at http://www.gpoaccess.gov/eop/2008/2008_erp.pdf on July 25, 2010.

¹¹ Available at http://unstats.un.org/unsd/statcom/statcom_09/seminars/innovation/innovations_seminar.htm on July 25, 2010.

OPERATING PRINCIPLES

The Census Bureau created LED to be an innovative program; it has taken a revolutionary new approach to analyzing and understanding socioeconomic data, as well as pioneering in the development of OnTheMap as a visualization and analytical tool.

Data visualization is a natural progression of advancing technology and an innovative mix of statistics, data mining, and computer science – all of which share the same purpose of making large datasets easy to explore, visualize, and understand. However, the field is still in its infancy, and it will likely evolve and grow very rapidly. Commercially or publicly available data visualization tools, such as [IBM's Many Eyes](#)¹² and [Google's Public Data](#),¹³ have grown rapidly since 2007. They excel in different ways, and most have no direct cost in their use. Some are more mature and have more sophisticated features, but others allow for larger file size or have unique technology such as motion. Many are experimental, lacking the stability that some users may desire. For example, technical support and problem resolution may not be available when issues surface. There is more commitment from custom developments such as the Census Bureau's OnTheMap, but development and maintenance of custom applications have higher costs.

Visualizing large amounts of data also means that imperfections of data files, such as outlying observations, can be more easily detected. The ability to detect problems and take corrective actions is a desirable feature from the product quality assurance perspective, but it can also inadvertently threaten the credibility of a program if many unexplained imperfections appear. LED has taken the latter possibility as impetus to build a top-quality product instead of retreating from implementing data visualization tools.

LED plans to continue to support and grow the OnTheMap tool. Substantial efforts will also be made to improve the visualization of QWI as time series data. Both OnTheMap and QWI will be upgraded by the integration of data on race, ethnicity, and federal workers by the end of 2010. As the program will be developing a third data line on the dynamics of job-to-job flows in the longer term, a balance of custom-designed and publicly available visualization tools will be explored for implementation.

Given the current conditions and understanding, the standing operating principles for LED on data visualization are:

1. Firmly commit to the continuing innovative development and expansion of OnTheMap with new features and creative use, as well as exploring custom development of visualization tools where none appear available commercially or publicly.
2. Actively pursue the exploration and implementation of commercially or publicly available data visualization tools, for the purpose of developing web-based applications, quality assurance practices, and improved training and presentation materials using LED data on an experimental basis.

¹² Available at <http://manyeyes.alphaworks.ibm.com/manyeyes/> on July 25, 2010.

¹³ Available at <http://www.google.com/publicdata/home> on July 25, 2010.

- Carefully consider the selection of LED data and commercially or publicly available data visualization tools according to their respective likelihood of success and their merits.
- Explicitly inform LED data users about the experimental nature of web-based data visualization products, as well as other limitations. For example, LED will not provide technical support or problem resolution with the experimental applications, and may withdraw any or all of them at will.

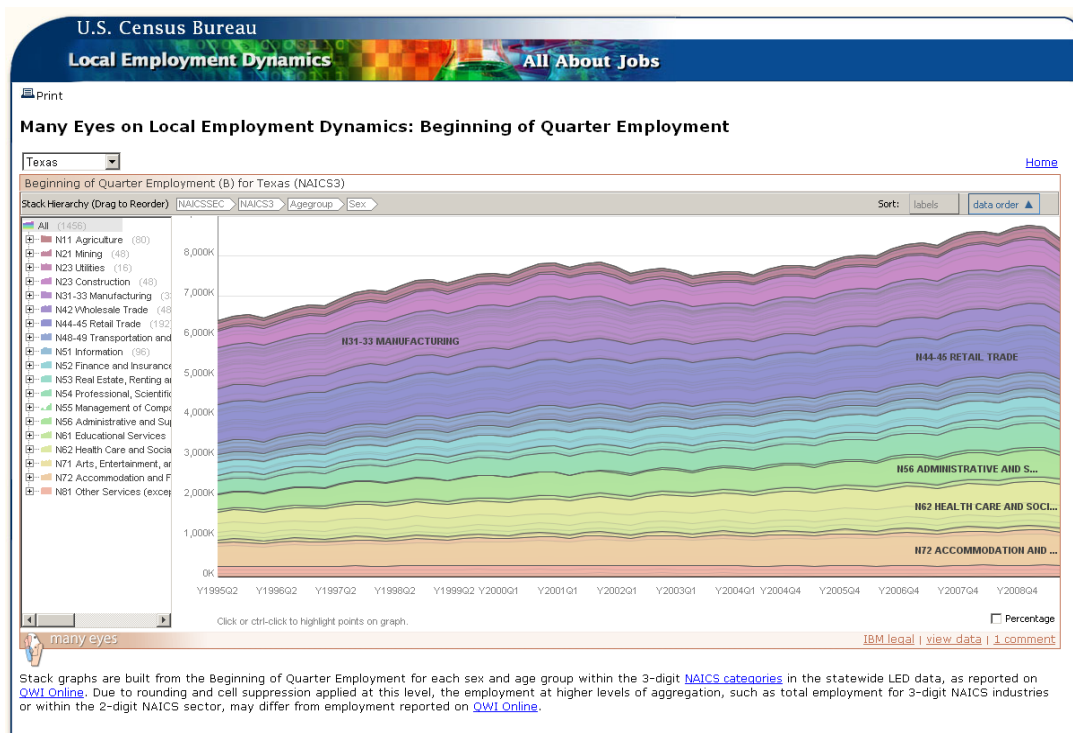
2010-2011 PLAN OF ACTION

QWI

- The following prototype has been developed applying the IBM Many Eyes Stack Graph for one of the QWIs, Beginning of Quarter Employment. The tentative experimental release date for this prototype is August 2010.

IBM Many Eyes Coverage Profile

Tool	Stack Graph
Indicator	Beginning of Quarter Employment
Geography	48 states
Gender	Female, Male
Age Groups	14-18,19-21,22-24,25-34,35-44,45-54,55-64,65-99
Industry	2-digit NAICS, 3-digit NAICS
Job Type	Private Sector Only
Time	Quarterly as far back as 1990, depending on state

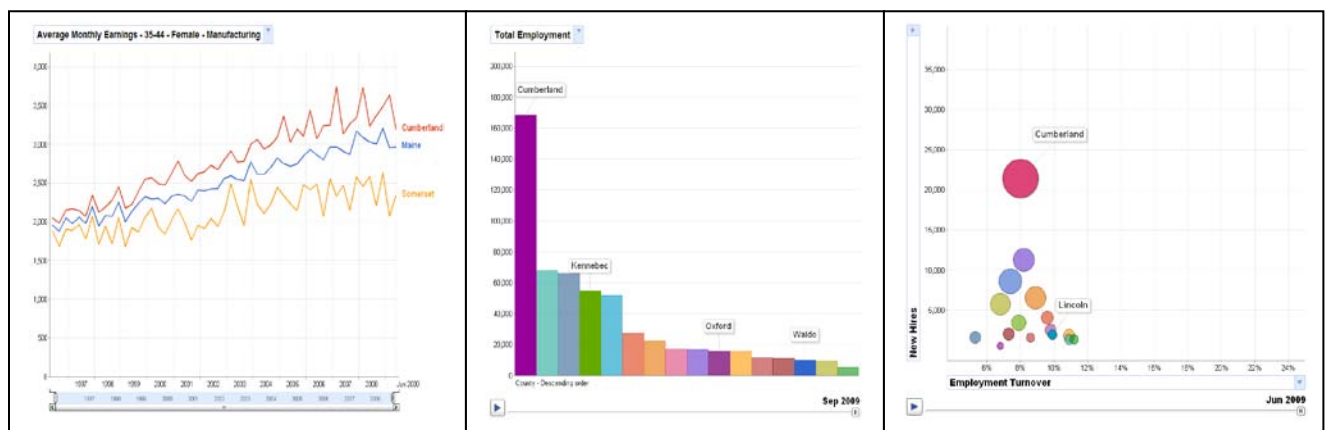


- Google launched [Public Data Explorer Labs](#)¹⁴ in 2010. LED has partnered with Google since March to develop two prototype postings. The first is for up to 5 individual pilot states with 9 QWI. The second is for all 48 state partners, but more limited in the number of QWI and features. Although the goal of the joint efforts is to post the complete set of QWI, it has become clear that data files at the 4-digit NAICS level are too large to post given the technology available at this time.

No tentative release dates have been set for these prototypes, but they are anticipated to be released in the fall of 2010.

Google Public Data Coverage Profile for Individual States

Tools	Line Chart, Bar Chart, Motion Chart
Indicators ¹⁵	Beginning of Quarter Employment (Also Total Employment) Net Job Flows Job Creation New Hires Stable New Hires Separations Turnover Average Monthly Earnings Average New Hire Earnings
Geography	Up to 5 pilot states, County
Gender	Female, Male, Both Genders
Age Groups	14-18,19-21,22-24,25-34,35-44,45-54,55-64,65-99,14-99
Industry	2-digit NAICS, 3-digit NAICS
Job Type	Private Sector Only
Time	Quarterly as far back as 1990, depending on state

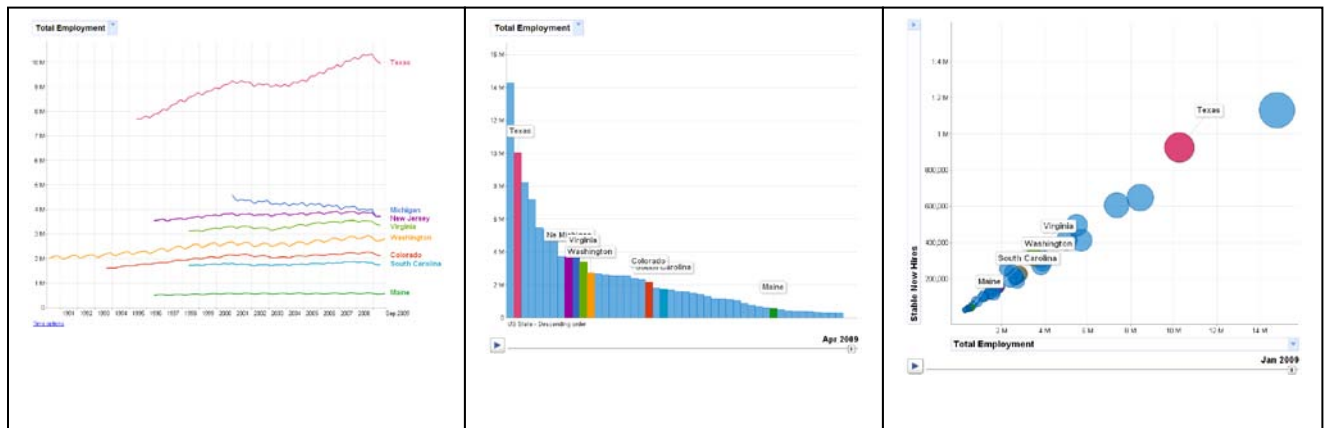


¹⁴ Available at <http://www.google.com/publicdata/home> on July 25, 2010.

¹⁵ They are the same 8 indicators as in QWI Online, plus Stable New Hires.

Google Public Data Coverage Profile for All State Partners

Tools	Line Chart, Bar Chart, Motion Chart
Indicators	Beginning of Quarter Employment (Also Total Employment) New Hires Stable New Hires
Geography	48 states, County
Gender	Female, Male, Both Genders
Age Groups	14-18,19-21,22-24,25-34,35-44,45-54,55-64,65-99, 14-99
Industry	2-digit NAICS
Job Type	Private Sector Only
Time	Quarterly as far back as 1990, depending on state



OnTheMap

1. Version 5 of OnTheMap is scheduled to be released in December 2010. In addition to including Connecticut as the 48th state and 2009 data, this new version will introduce visualization features on the direction and distance of potential commuting, which are being developed.

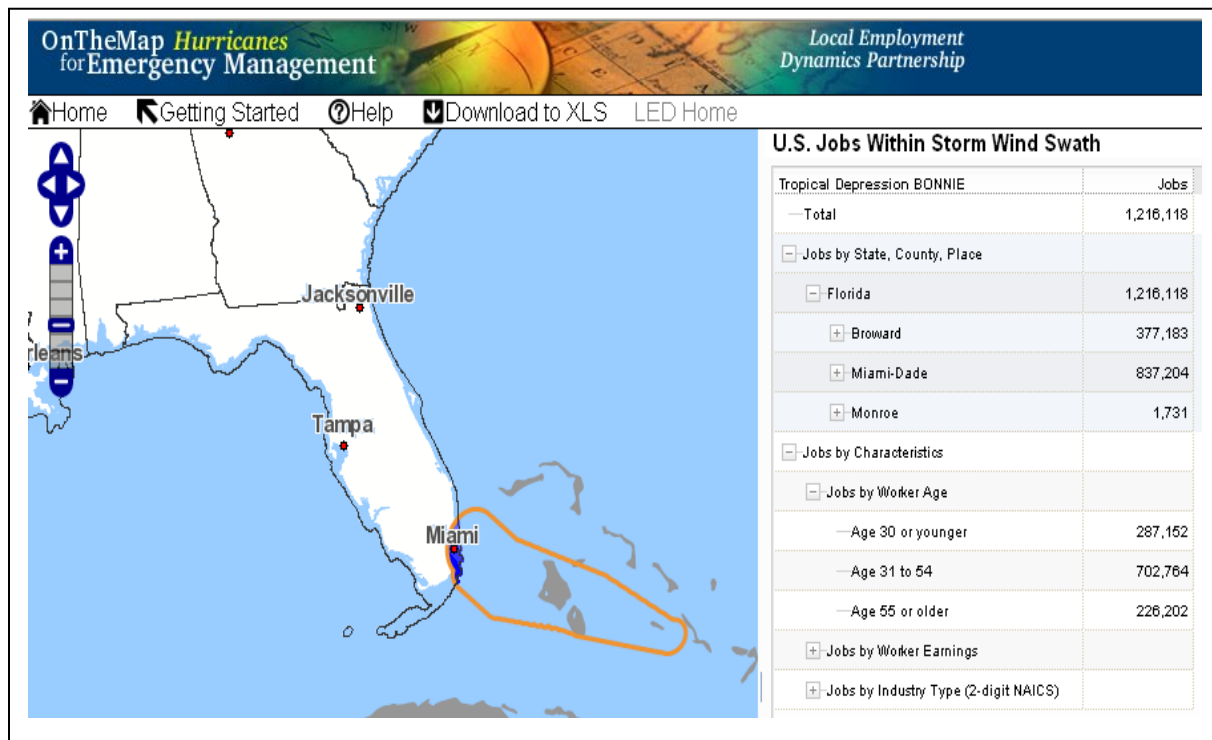
OnTheMap Version 5 Coverage Profile

Indicators	Beginning of Quarter Employment
Geography	48 states with Census block as base unit; reports for over 15 possible geographical levels
Age	30 or younger; 31 to 54; 55 or older
Gender	Female, Male, Both Genders
Earnings	\$1,250 per month or less; \$1,251 to \$3,333 per month; more than \$3,333 per month
Industry	2-digit NAICS
Job Type	All Jobs, Primary Jobs, All Private Jobs, Private Primary Jobs
Time	2002-2009 with second quarter as reference

2. The inaugural version of [OnTheMap for Emergency Management: Hurricanes](#)¹⁶ was released¹⁷ on July 12, 2010. The web-based spin-off application from OnTheMap uses live feeds from the National Hurricane Center about the location of a storm and its wind swath to provide real-time labor market information when the storm strikes land in the United States. Efforts are ongoing to improve the map layer and related look and feel. The tentative release date for these updates is September 2010.

OnTheMap for Emergency Management: Hurricanes Coverage Profile

Indicators	Beginning of Quarter Employment
Geography	47 states at Census block level; reports for Counties and Places
Age	30 or younger; 31 to 54; 55 or older
Earnings	\$1,250 per month or less; \$1,251 to \$3,333 per month; more than \$3,333 per month
Industry	2-digit NAICS
Job Type	All Private Jobs
Time	2008 with second quarter as reference
Note	<ul style="list-style-type: none"> • Activates with live feeds from the National Hurricane Center when a tropical storm or hurricane is formed • Generates and updates local OnTheMap jobs data when the storm's wind swath reaches U.S. land



¹⁶ Available at <http://lehdmap4.did.census.gov/themap4/em/index.html> on July 25, 2010.

¹⁷ Available at http://www.census.gov/newsroom/releases/archives/hurricanes_tropical_storms/cb10-106.html on July 25, 2010.