Making LED Data and Applications Available

Jeremy S. Wu

Data dissemination refers to the means of making data available to the users. While producing quality data in a timely manner is a core mission of the Local Employment Dynamics (LED) partnership, putting these data in the hands of the users, subject to confidentiality protection, is equally important.

This paper provides a brief summary of what LED has established and is establishing as its data dissemination methods, as well as an emerging strategy for future development.

THE DEFAULT METHODS

Upon completion of data processing each quarter, about 38 standard tables containing Quarterly Workforce Indicators (QWI) are produced for each LED data-supplying state partner. These tables have been approved for public use by the Census Bureau Disclosure Review Board. This complete set of data is made available in three ways:

- **DVD.** The tables are exported to a Comma Separated Value (CSV) file format, zipped with password protection, and placed in a DVD. The DVD is sent by FedEx to the sponsoring state labor market information agency.

- **HTTP.** The same files are stored in the Census Bureau's Bowie Computer Center for secured HyperText Transfer Protocol (HTTP) download. Each LED state partner is assigned an access code so that it can "pull" its QWI data electronically whenever it is ready or desires.

- **CRADC.** The same files are also stored in the Cornell Restricted Access Data Center (CRADC), which is operated by Cornell University. This site has a collection of statistical tools, but users must sign an agreement, are assumed to be knowledgeable statistical practitioners, and observe the terms and conditions of Cornell University. Any extraction of data is subject to 24-hour notice and approval by the CRADC administrator.

Individuals or organizations that wish to obtain access to the complete QWI files must obtain explicit approval from the corresponding state labor market information agency. Upon receipt of the approval, usually an email, LED will send a DVD or provide HTTP access to the requester. Access to CRADC must also be approved by Cornell University in writing.

AVAILABLE METHODS

The following methods and applications are currently available at the LED website (http://lehd.dsd.census.gov). LED has started to place these applications into the websites of the state partners (a practice called "skinning"):

- **QWI Online.** QWI Online is a signature LED product. Through its use of selection menus and pivot tables, the user can generate dynamic tables that are consistent with the Census Bureau disclosure avoidance rules, with the option of downloading the selected results into Excel spreadsheet files.

- **Top Industries.** This is a new online application designed primarily for Workforce Investment Boards so that the user can find the top industries ranked on a variety of measures, by local geography, age, and gender.

- **Local Workforce.** This is a new online application designed primarily for employers so that the user can describe the workforce composition of selected NAICs by local geography and industry. The reporting is by age and gender.

- **Template Reports.** A series of template reports has been produced profiling older workers for ten LED state partners, and two are under preparation. A new template report on older workers is being reviewed at this time. These reports are produced jointly with the individual state partners in hardcopies and made available electronically on the LED website.

METHODS UNDER STUDY OR DEVELOPMENT

During the past year, LED has been studying and developing additional dissemination methods that include:

- **Data Ferret (http://dataferrett.census.gov/TheDataWeb/index.html).** Data Ferret is a data extraction software and a data mining tool developed by the Census Bureau. It accesses data through the Internet and can be used to develop custom web pages. LED has received a demonstration of Data Ferret and is favorably impressed by some of its features. Additional discussions are being held to explore the possibilities of implementation for LED.

- **MDDB (http://www.sas.com/technologies/dw/storage/mddb/).** MultiDimensional DataBase (MDDB) is an application being developed by LED based on the SAS software package. It is intended to provide quick access to pre-summarized data that are generated from vast amounts of detailed data. The prototype has been demonstrated in LED training sessions, and is currently available in the Partners Only section of the LED website.

- **ALMIS (http://www.doleta.gov/business/rsrclib/mktinfo/).** America's Labor Market Information System (ALMIS) is a centralized database for the maintenance of labor market and occupational information. It is a system supported by the Employment and Training Administration (ETA) at the Department of Labor. LED is committed to supply and integrate its data with ALMIS on or before June 2005 under agreement with ETA.

- **Beyond 20/20 (http://www.beyond2020.com).** This is a commercially available software package for data dissemination, visualization and analysis. It is a tool that is

currently used by the American Community Survey to disseminate data. LED received a demonstration of Beyond 20/20 and was impressed with its basic features. The preferred Linux platform will become available around September 2005, and it is envisioned to be the next version of QWI Online. Negotiations on building a prototype are ongoing.

- **Advanced Query (http://advancedquery.census.gov).** This is a custom Census Bureau application that produces tabulations from the Census 2000 microdata files on the Internet. It allows the user the ability to construct tabulations based on rules for electronic disclosure limitation developed by the Census Bureau. Features of Advanced Query are similar to those of the American FactFinder in the Census Bureau website (http://www.census.gov).

LED DATA DISSEMINATION STRATEGY

Data dissemination will be a continuing challenge to LED due to a number of factors:

- **Growth of Volume.** The QWI files will grow rapidly in size and number as historical data accumulate, new indicators are introduced, and new needs are identified for development.

- **Growth of Users.** As LED and its visibility expand, the number of data users, as well as the diversity of their program and technical knowledge, access devices, and appetite for information, will also increase.

- **Growth of Functional Requirements.** As the potential value of the LED data is increasingly recognized, additional functional requirements will be identified, usually at a rate much faster than the rate of development.

- **Evolving Technologies.** Technology continues to evolve as yesterday's solution may rapidly become obsolete while tomorrow's solution requires both knowledge and resources to develop.

- **Constraints of Budget.** At the same time, funding is always limited that it requires delicate balance between needs, growth, technology, and resources.

LED will continue to strive to meet these challenges with innovative solutions, as it has in the past. Four strategic approaches are being formulated and outlined here:

- **Use of Internet.** Internet will be increasingly the mode of choice to disseminate dynamic, public-use data to an ever-expanding and diverse user community. The original LED website was launched in February 2004, and it was redesigned by December 2004. As new applications such as On The Map, Data Ferret, and MDDB, are developed, the features and functionalities of existing applications such as QWI Online, Top Industries, and Local will also be enhanced or consolidated.

- **Identifying User Needs.** LED will improve its means to identify the user needs for data dissemination by feedback from the LED Steering Committee, state partners, sponsors such as ETA, stakeholders, annual workshop, and its website. In addition, periodic online opinion surveys will be conducted; focus groups organized; and comments from training sessions and conferences collected.

- **Planning and Organization.** To the extent practicable, the identified user needs will be matched with the state of technology, resources, and program objectives to enhance the overall data dissemination operation and its performance. This will in turn require more careful planning and organization to ensure that the LED data dissemination methods will continue to be relevant, efficient, timely, and innovative that will also meet confidentiality protection, accessibility, and similar Census Bureau and federal regulations.

- **Training.** Classical classroom training and new electronic training are important components of a successful data dissemination strategy. Knowledge should be passed on to the "trainers" who can train others to train others. LED is committed to build on its training program with this renewed "train-the-trainer" emphasis. An emerging vision and outline for training will be presented for comments during the 2005 annual workshop.

COMMENTS AND QUESTIONS

Please direct all comments, suggestions, and questions about this paper or LED data dissemination to Jeremy Wu, program manager, at Jeremy.S.Wu@census.gov.