Where the Jobs Are: Using Real-Time LMI for Green Jobs

Accurate and timely labor market information (LMI) helps to ensure a well-functioning labor market. Real-time LMI is a new source of information and has the potential to improve our understanding of current labor market conditions and emerging fields, including the green economy. As part of the American Recovery and Reinvestment Act, the U.S. Department of Labor received $50 million to help state workforce agencies improve LMI on green jobs and develop practices for possible broader use and impact. Sixteen of the grantees used real-time LMI to explore the green economy, and one grantee, the Northeast Consortium, dedicated its entire grant to understanding the strengths and limitations of real-time LMI. This brief shares some of the lessons from their efforts.

What Is Real-Time LMI?

Real-time LMI is created by scraping job postings and resumes from the internet daily and aggregating this information to make inferences about labor market conditions. The data can illustrate trends in supply and demand, emerging occupations, skill requirements, and demand for education and certifications. Web crawlers or spiders gather data from public and private labor exchanges and websites that host job postings, such as online job boards, employer sites, newspapers, and government sites. The software parses job listings into data fields for analysis. The information can include required skills, education levels and certifications, wages, and geographic locations.

Although real-time and traditional LMI can address the same questions, data sources and data collection methods differ. Traditional LMI is collected from employer surveys and interviews, as well as administrative records.

How Well Does It Work?

Data captured by spidering or web-scraping have strengths as well as limitations. Some postings provide details about the skills and qualifications needed and can identify emerging skills; however, the scraping program may have difficulty distinguishing skills (such as advanced computer skills) from qualifications (such as a technical certificate).

Most job postings identify the employer, which enables LMI analysts to study the types of jobs a firm posts online and the frequency of the postings. Analysts may also be able to use the firm name to link real-time information with administrative or survey firm data. In addition, they may examine the frequency of postings for particular job titles—for example, to see whether openings for a "wind plant technician" are increasing or decreasing.

In terms of education or training requirements, job postings are not always helpful. The job title (such as lawyer or doctor) may imply the need for a particular degree. Or the posting may include preferred or required education levels (for example, "one year of college or a technical school certificate, or 12 months of related experience or training, or equivalent combination of education and experience"). Parsing such postings into data fields may be difficult.

At its best, real-time LMI offers information on the universe of online job postings. However, jobs not advertised online (such as certain low-wage or union positions) are underrepresented in any labor market analysis based on real-time data. In addition, online postings are designed to meet the needs of employers, who may include vague details to attract a large number of applicants. Many firms use online postings to collect resumes for projected rather than actual vacancies.

Geographic information may also be misleading. For example, postings may state the location of headquarters rather

Real-Time Data

Web Scraping

Data

LOCATION
OCCUPATION
INDUSTRY
EMPLOYER NAME
SKILLS REQUIRED
WAGES

Trends in Supply and Demand
Emerging Occupations

Increased Understanding of Labor Markets

Real-time LMI is created by scraping job postings and resumes from the internet daily and aggregating this information to make inferences about current labor market conditions.

than of the vacancy. If so, an analysis may yield inaccurate data on local labor markets and locations of job growth.

Many job postings also omit detailed salary information. Instead, they may offer "competitive piece rate and benefits." Analysts may need to use traditional LMI on average salaries to supplement the information in online postings.

Shedding Light on Green Jobs

Real-time LMI can highlight employer demand and required skills for green jobs and potentially be used to project job growth.

In our study, grantees who wanted to use real-time LMI to better understand green jobs faced significant challenges. The first was defining "green." Most grantees used occupation and industry as a starting point for this definition, often based on the list developed by O*NET. Others included additional occupations identified through surveys. However, not all job listings clearly identify occupation and industry. Furthermore, this approach flags all postings under a given code (such as chemical engineer), and may over-identify green jobs because not all positions in a given occupation are green. Conversely, certain green postings might not be flagged because the occupation or industry is not on the green list.

The Northeast Consortium attempted to identify jobs based on green words or phrases in the postings. A software tool parsed online job listings, searching for green phrases. The software and list were developed iteratively, by examining postings identified as green (or not identified as such) to determine the success of the parsing tool and then refining it. By the end of the grant, the list included nearly 900 key phrases. The consortium supplemented this approach with a list of green firms identified using traditional LMI sources. Developing accurate lists required knowledge of the labor market and a commitment to updating the lists as new information became available, a

time- and labor-intensive approach.

Analyzing Green Jobs. After identifying jobs as green, grantees used real-time LMI to assess demand at a specific point and determined skills, certifications, and training requirements.

One state in our study used real-time data to assess demand but acknowledged these data were inexact, and postings did not always reflect actual vacancies. The state used real-time data to assess the strength of current employment demand by calculating the ratio of real-time postings to total state employment. Grantees also examined job posting data to determine occupations and industries placing the most ads for green jobs.

… real time LMI always needs to be linked to traditional sources of LMI.
—Northeast Consortium

Real-time data help track demand for skills and certifications necessary for green jobs. For instance, postings from known green firms that require certifications, such as Leadership in Energy and Environmental Design (LEED), provided insight into what was required for positions in green industries. The Northeast Consortium determined that the top three green "skills" in job ads were environmental health, water treatment, and LEED certification.

This consortium also explored using these data for short-term employment projections. However, analysts need repeated, consistent measures of employment demand to construct a demand model. If a group decides to spider additional websites, or changes the frequency of spidering, the measures are not consistent and job posting data may spike purely as an artifact of the change in data collection. The consortium worked with real-time vendors to develop a limited set of websites that are spidered consistently to ensure data are not generating the appearance of changes in employment demand.

What Have We Learned?

Real-time data have value as a point-in-time measure to capture information on new and emerging occupations and skills. They provide an opportunity to analyze current labor market conditions without the lag that occurs in traditional LMI. As always, data quality is key. Furthermore, real-time LMI provides better information when combined with traditional LMI. Our study suggests the following.

Be cautious about data quality. Real-time data can be noisy, producing false positives from duplicate observations, inaccurate flagging of green jobs, and postings from employers that do not actually have openings but want a constant applicant pool. Human monitoring ensures high quality data, and real-time LMI may need the same level of quality control used in other surveys, such as those from the Bureau of Labor Statistics. Researchers need to understand which sites are being spidered and how frequently. Changes in web-scraping procedures can produce an inaccurate picture of changes in economic conditions. It is particularly challenging to construct a time series when the data collection process is evolving.

Combine real-time LMI with traditional LMI. Real-time data provide a snapshot of the market and may be the best data source for certain credentials, such as industry-based certifications not captured in current surveys. These data also provide information on new and emerging occupations that may take time to appear in traditional LMI. For the best use of real-time data, users should combine it with traditional LMI to provide context.

Our country's interest in preparing the workforce for new and better jobs requires a strong understanding of current and future labor market conditions. Real-time LMI is an important tool that can help us move toward this goal.

Read more about Mathematica's study at www.mathematica-mpr.com/labor/green_jobs_economic_growth.asp.