7-14-2020

Workforce Data and Government in **Data Science in the Public Interest: Improving Government Performance in the Workforce**

Joshua D. Hawley

The Ohio State University, Ohio Education Resource Center, and Center for Human Resource Research

Data Science in the Public Interest

Improving Government Performance in the Workforce

Joshua D. Hawley

2020

W.E. Upjohn Institute for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

Names: Hawley, Joshua D., author.
Title: Data science in the public interest : improving government performance in the workforce / Joshua Hawley.
Description: Kalamazoo, Michigan : W.E. Upjohn Institute for Employment Research, 2020. | Includes bibliographical references and index. | Summary: "This book is about how new and underutilized types of big data sources can inform public policy decisions related to workforce development. Hawley describes how government is currently using data to inform decisions about the workforce at the state and local levels. He then moves beyond standardized performance metrics designed to serve federal agency requirements and discusses how government can improve data gathering and analysis to provide better, up-to-date information for government decision making"— Provided by publisher.
Identifiers: LCCN 2020007603 (print) | LCCN 2020007604 (ebook) | ISBN 9780880996747 (paperback) | ISBN 9780880996754 (ebook)
Subjects: LCSH: Manpower policy—United States. | Labor policy—United States. | Labor supply—United States—Statistical methods. | Labor market—United States—Statistical methods. | Big data.
Classification: LCC HD5724.H346 2020 (print) | LCC HD5724 (ebook) | DDC 331.12/042028557—dc23
LC record available at https://lccn.loc.gov/2020007603
LC ebook record available at https://lccn.loc.gov/2020007604

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Avenue
Kalamazoo, Michigan 49007-4686

Cover design by Carol A.S. Derks.
Index prepared by Diane Worden.
Printed in the United States of America.
Printed on recycled paper.

1
Workforce Data and Government

A healthy workforce is a good indicator of a successful economy. We measure this health with key indicators such as the unemployment rate, income, and productivity, which help to define expectations. Is the economy strong or weak? What job search parameters do new graduates face? The business news diligently tracks and reports economic indicators in an effort to assess the current state of the economy and to predict its future direction.

One of the most critical economic issues is squarely in the workforce domain: How will changes in technology impact jobs? Workers fear that they will lose their jobs to automation; employers worry that they will struggle to find workers with the appropriate skills to utilize new technology. And governments are concerned about the economic impact to their communities as they work to stay competitive and attract businesses to build their tax base.

The research literature identifies the kinds of jobs that can be replaced by technology (Autor, Levy, and Murnane 2002, 2003). For example, occupations with routine job tasks are expected to be automated, leading to displacement of workers. Estimates on the numbers of jobs that will be replaced by technology vary considerably. One estimate from McKinsey Global Institute (Manyika et al. 2017) predicts that up to 73 million jobs will be lost in the United States by 2030. The report forecasts significant job losses across occupations and highlights a staggering 375 million jobs where people "may need to switch occupational categories and learn new skills" (p. 11). The same report, however, also projects an increase of 20–50 million new jobs globally in information technology in the same period. The potential upsides of technological advances include the creation of jobs that will require a higher set of skills and offer commensurate pay.

Minimum wage is another pressing workforce issue. Many cities and states are wrestling with the idea of mandating increases to the minimum wage to address the gap between earnings and the cost of living. Wage growth has been weak, while housing, medical, and other

costs have rapidly increased. States and cities are increasingly focused on public policies that will increase earnings for its residents.

The economic recovery from the Great Recession created significant labor problems. Many unemployed individuals were unable to find work, while businesses could not find qualified workers. The recession was particularly difficult for workers just entering the job market, especially recent high school dropouts. Men without college degrees who were laid off during the recession had an especially hard time finding their way back to full employment. The reaction of this population played a major role in the outcome in the 2016 presidential election, as these workers voted for candidates whom they believed understood their plight and would restore jobs.

The changing nature of work combined with the inevitable struggle to adapt raises the stakes for government. What should the role of government be in the workforce? Can we improve the overall skills of new graduates as they make their way in this new uncertain economy? Federal, state, and local government have distinct roles in workforce training specifically and workforce development more generally. Local and state government have statutory authority over K–12 education and most of the higher education authorities. While federal policies govern vocational training for adults, states maintain influence and govern the local institutions. Although the overall funding levels for workforce development have decreased over time, recent administrations have shown renewed interest in specific workforce development programs, including the Registered Apprenticeship Program and Career and Technical Education for high school students (Hawley 2017).

Governments have turned to a new and important instrument to address workforce issues: data. In particular, they are increasingly deploying administrative data to design and build data systems relating to student information, human services, and workforce or employment. When it enacted the Workforce Innovation and Opportunity Act (WIOA) in 2014, the federal government outlined for the first time a common set of national performance standards, which are in fact mandated for a range of workforce efforts, such as Career and Technical Education, Adult Education, and the Temporary Assistance for Needy Families. Broadly speaking, these performance standards look at employment outcomes of training funded by the government. WIOA requires states to use common files, such as the wage records kept by

the Unemployment Insurance (UI) program. States are just beginning to develop the data systems to implement these new accountability systems. Ohio has incorporated WIOA's performance standards into its statewide strategic plan for economic development and provides the metrics online in a dashboard format. Counties are using these dashboards to monitor outcomes for different WIOA programs, such as Adult and Dislocated Worker. Counties also have the ability to use the performance tools to monitor related workforce programs, such as Registered Apprenticeship, using the same common measures. Chapter 4 will review Ohio's program in more detail and discuss the modifications the state has made.

In addition to holding public agencies accountable, data improve the performance of the workforce development system through "learning" from prior experience (Eberts 2015). Although some workforce development programs are being developed according to best practices, there are challenges with incorporating lessons from local experiments into current practice. For example, Ohio rolled out the Comprehensive Case Management and Employment program in 2015, which required significant changes to the organization and delivery of local services to eligible youth. The state collected data on how counties changed their institutional structure to follow state laws and implement the program, but it did not necessarily build systems to capture lessons learned from these new policy shifts (Hawley and Munn 2017). Thus, it is very difficult to adjust program design, change the curriculum of the programs, or know if big shifts in program delivery can be made to improve performance. As a result, job counselors, career advisors, and administrators operate without a standardized set of practices to improve performance.

The problems governments face in addressing gaps in the workforce are magnified by the lack of a clear and standardized definition of the concept of workforce development. Workforce development is not taught in college. There is no codified body of practice that forms the basis for the field. Some scholars have defined similar concepts, such as workforce education, which have more in common with vocational or career and technical education (Gray and Herr 1998). Others have defined workforce development to include fields such as adult education or human resource development (Bernick 2005; Harrison and Weiss 1998), but very few scholars have attempted to outline a comprehensive

definition for workforce development as a coherent field of educational practice (Jacobs 1990, 2000; Jacobs and Hawley 2005).

Why does this matter? Since workforce development is not a formal field or discipline, it can be difficult to define the government programs that form the core offerings of a workforce development program. In my own writing I have focused on the commonalities between programs that serve both adults and youth. Jacobs and Hawley (2009) outline the scope of workforce development programs, including in our definition programs such as career and technical education, human resource development, and training for entry-level workers and incumbent workers. The following section describes attributes of workforce development that programs share.

Skills Matter

Workforce development in essence is about skills, specifically those that are in demand. Without the right skills, the unemployed will not find jobs, and the employed will have a tough time working their way up the ladder. Workforce development is also about job search skills, as well as helping individuals reduce barriers to employment (e.g., a criminal justice record, lack of sufficient transportation, or poor housing).

It can be challenging to measure the overall skill level of the workforce. Governments use proxy measures such as the education level to determine if the workforce is more skilled or less (i.e., the more educated the workforce is, the more skilled it is), but that can be misleading. While the percentage of the current workforce aged 25–64 that has an associate's degree or greater increased from 34.6 percent in 2005 to about 39 percent in 2015, overall skill readiness has not kept pace.

Historically, the education levels used to measure workforce skills have moved through different stages. In *The Race between Education and Technology*, Goldin and Katz (2008) describe the movement beginning in the early 20th century toward universal high school education. Before 1910, only 9 percent of Americans graduated from high school. By 1940, the graduation rate was over 50 percent. Moreover, 73 percent of youth aged 15–18 in 1940 were enrolled in high school, compared to just 19 percent a generation earlier. The United States experienced a dramatic increase in the number of high schools in the period between the two world wars. However, access to education after high school

spread unevenly across the nation, lagging in urban northern cities and in the South. But by the 1950s, states moved to build higher education institutions to meet increased demand for college based in large measure on the GI bill.

The real growth in higher education did not happen until the 1960s. Until that time, the bulk of the working population made do with high school or elementary schooling. During the 1960s, college enrollment grew by 120 percent compared to the 1950s. In 1969–1970 alone, the number of bachelor's degrees issued by the United States (792,000) more than doubled the number issued a decade earlier (392,000) (National Center for Education Statistics 2016). Currently, colleges issue about 1.9 million bachelor's degrees annually.

Significant gaps in educational attainment between men and women have always existed. According to data from the American Community Survey, 26 percent of women aged 25–64 and 28 percent of men in the same age bracket had at least a bachelor's degree in 2015. However, when you narrow the age group, the differences are more noticeable. Women aged 25–34 have a 5 percent advantage, with 32.5 percent of women completing bachelor's degrees compared to 27 percent of men (Goldin and Katz 2008).

Levy and Murnane (2004) demonstrate another way of measuring skills. They use the *Dictionary of Occupational Titles* to tabulate the 12,000 occupational descriptions and label skills changes in the U.S. economy. The changes in job skills show that work requiring complex communication or expert thinking grew between 1969 and 1999, while various types of manual work declined substantially. These data show that as education levels increased, so did jobs that required higher skill levels.

Technology plays an important role in the skilled workforce. Since the 1960s, computers have replaced many jobs in the United States. Levy and Murnane (2004) describe four ways that the introduction of mainframe computers affected jobs: 1) by changing the number of or demand for different jobs, 2) by altering the kinds of jobs available for workers, 3) by increasing or decreasing wages, and 4) by changing the types of skills required. Computers can substitute for labor in one area, but jobs change in other ways. For example, as banks automate check processing, the number of people needed to verify funds and read legible handwriting decreases (Autor, Levy, and Murnane 2002). In con-

trast, the number of high-paying jobs to program machines to read the checks or to carry out exceptions processing increases. It is very difficult to connect technological advancement to a single shift in employment. There are changes in the level of employment, but the number of jobs impacted by secular shifts in job tasks complicates forecasting job demands in the future.

Immigration is also an important factor in the labor force. Public policies have changed over time to accept or reject immigrants from different countries and with different education or skills backgrounds. Currently, immigrants make up about 13.5 percent of the population in the United States, whereas in 1970 they composed only 4.7 percent. In 2015, there were approximately 43 million immigrants. The trend in recent decades is that immigrants are overall less educated than in prior decades. The change in education level among immigrants affects the mix of workers available in the economy. That is, less-educated immigrants are competing with less-educated native-born Americans for the same jobs. However, there are discernible differences between these two groups. Low-skilled immigrants tend to gravitate toward different types of work than that of low-skilled Americans. The kinds of jobs that immigrants hold differ depending on local economic and labor market conditions. Ohio has many small and medium-sized cities, where a large fraction of workers are Hispanic. While Hispanics are only 3.6 percent of the state as a whole, they represent over 6 percent of the local population in seven counties in the 2010 census (Ohio Department of Development, n.d.).

Conversely, immigrants also make up a disproportionate fraction of our educated workforce. Of the 64.6 million individuals with college degrees in 2014, 10.5 million were immigrants. The percentage of immigrants with college degrees has been increasing at higher rates than that of native-born college educated workers (Zong and Batalova 2016). College-educated immigrants tend to be concentrated in high-demand sectors, such as health care or technology.

Public Policy and the Workforce

WIOA focuses on training and workforce opportunities for disadvantaged workers. When enacted in 2014, it laid the groundwork for common performance metrics, unified planning, and the development

of legal frameworks and data standards. The intent of these mandates is to give states a firm push to better align resources across historically isolated and disparate training efforts. The challenge to compliance lies in the consistency and availability of data and in translating that data to coherent practices. Governments generate an immense variety of data on participants, the programs they complete, and the outcomes, but they fail to link data on a regular basis with different agencies to improve performance. Moreover, the speed with which data are generated creates expectations that data can be analyzed and used rapidly. There is a mismatch between the needs of government for actionable data and local communities' willingness to change practice based on data.

The increasing use of data—something seen across all areas of business and government—offers opportunities for workforce development professionals to combine data resources to improve the quality of information available for decision making. Data-based decisions can improve forecasting for planning and budgeting future programs. For example, Ohio's Strategic Plan at the Governor's Office outlines the common measures used to compare the performance of the workforce programs, including wage and employment outcomes on all key federal workforce programs, including WIOA programs. Additionally, the state developed a dashboard website that compares program-level performance and can be used over time to improve the management of workforce development programs.

Workforce development data systems are continual works in progress. Early efforts from the U.S. Department of Labor (USDOL) and the Census Bureau in the 1960s and 1970s aimed to develop longitudinal data systems, including the use of UI wage records and the Continuous Wage and Benefit History program (Blanchard and Corson 1982; Borus and Tash 1970). Building on those efforts, several states worked together in the 1990s and early 2000s under a labor department mechanism called the Administrative Data Research and Evaluation Project (ADARE). The ADARE states collaborated on research and evaluation projects. The initiative improved understanding about the different data systems that states use in workforce development, resulting in additional government investment at the federal level in 2010 with the Workforce Data Quality Initiative (WDQI). WDQI built infrastructure at the state level, focusing on networks of professionals who have experience working with workforce, higher education, and K–12 data

linkages within state government and university systems. WDQI funds have supported 35 states since 2010 with over $57 million, with some states receiving multiple grants. The funds also have been used to build statewide longitudinal data systems, often combining UI wage records with other longitudinal data files in systematic ways. As a point of comparison, the U.S. Department of Education built the Statewide Longitudinal Data Systems program, which has provided about $640 million in funds to virtually every U.S. state until 2014 (Government Accountability Office 2014).

State-level investments in longitudinal data systems vary significantly. Many states had some experience building longitudinal data systems for the Continuous Wage and Benefits History or through developing UI wage records more generally. Florida's state government was quite active in building the Florida Education Training Placement Information Program (FETPIP) in the 1980s and 1990s, and, despite a lack of a comprehensive source of information on state systems, many other states used state resources to develop similar statistical systems. In Chapter 4 we describe the work done in two states, Washington and Ohio, as a way to understand the more general process to develop statewide longitudinal data systems.

Governments traditionally have used data to make decisions for workforce planning and policy, as evidenced in military-run workforce programs in World Wars I and II and with the advent of modern workforce development policies (Dorn 2007; Jacobs 2000). Currently, large workforce programs, such as WIOA or the Federal Perkins Loan Program, require concrete data on the number of jobs by occupation in the economy and the level of education that is required for those jobs. These core decisions in workforce planning might be stated as follows:

1) How many jobs are required in the future?
2) What occupations will these jobs be from?
3) What educational levels do the jobs require?
4) How many new workers are needed at what time?

Traditionally, government and academics have used manpower planning to answer these questions. Manpower planning is defined as "the development of human resources for efficient use" (Richter 1984, p. 678). Developed during reconstruction after World War II, manpower

planning used labor force statistics to assist national governments with centralized decision making. It is still used in some form by institutions such as the Bureau of Labor Statistics (BLS), which estimates labor demand and attempts to tease out the required education level in the jobs for the future. This model has fallen out of favor in developed nations, although it continues to be used in some developing countries (Psacharopoulos 1991; Richter 1984). Russia, China, and South Korea have an easier time generating estimates using manpower planning tools. Moreover, countries with a large fraction of employment in manufacturing or government can more easily use the methodology.

It may be possible to use population estimates, supply information, and macroeconomic data and arrive at an estimate of how many workers will be needed. However, governments cannot use manpower planning to identify precise numbers of skilled workers required by employers throughout the economy. We discuss this problem in more detail in Chapter 4 and show that the solution requires more complex data-based tools.

Researchers have criticized manpower planning (Spalletti 2008). The primary criticisms center on the difficulties of forecasting demand for skilled workers in a dynamic economy. As discussed earlier, the number of jobs lost to technology and the number created in new fields of work are hard to predict. However, governments are still faced with the questions of how and where to invest in specific training programs and educational planning. Current decisions might focus more on the kinds of jobs by industry or career cluster and provide a range of data to state and local government. As these questions evolve, our data systems require changes to ensure that government has the necessary information to make the decisions.

Expectations Based on Business

Current thinking on workforce planning at the state and local levels has been influenced by the larger trends from technology and economic development. Data collection is ubiquitous and deployed extensively to improve business. Businesses are constantly assessing the value of specific investments in everything from credit card offers to pay rates for sales workers, and determining how these investments impact core business outcomes (Levenson 2015). Government has dramatically

increased the amount of data that it collects and is consistently improving data utilization.

Let us examine in detail the ways in which government ensures businesses have a supply of skilled college graduates. Based on national data, it is presumed that the country has too few skilled workers in key areas such as engineering or medicine. Unfortunately, the data do not guarantee that the jobs will actually materialize four years later after students graduate from college. Indeed, many scholars do not agree with the conclusion that there is pent-up demand for college-level workers (Rosenbaum 2003; Salzman 2013). Occupations that demand higher levels of education may in fact be growing more quickly than others, but still represent only a small fraction of total job openings annually (Sommers and Franklin 2012). Therefore, government is investing in policies that essentially are no better than informed guesswork regarding future business hiring. Indeed, an article by colleagues from MIT (Xue and Larson 2015) reflects this conundrum.

Government systems to measure the effectiveness of investments in workforce training have greatly improved in recent years. Traditional tools for measuring results require evaluations that have additional costs and can take years to complete. And after making changes to federal programs, government often does not evaluate outcomes to determine success. For instance, even though vocational training increases earnings for workers, only a small minority of WIOA participants complete any vocational credentials (Bloom and Michalopoulos 2001; Mangum 2000).

Unfortunately, the process of creating public policies is more political than scientific. Government policies on training or workforce often do not reflect the mountain of hard-won evaluation evidence. The evaluation evidence uniformly notes that workforce programs remain isolated and fragmented. State agencies have a hard time integrating services among disparate programs. Secondly, effective workforce programs, the subject of multiple random-assignment evaluation studies, have not been replicated successfully. Moreover, the successful models identified in evaluation studies have not been consistently translated into changes at the local level (Government Accountability Office 2000). The United States still maintains multiple funding streams and programs, leading to duplication of effort and confusion for the job-training clients. For example, a substantial body of evidence shows

the value of integrating academic and vocational content into training (Bottoms, Presson, and Johnson 1992; Hawley and Chiang 2016; Rosenstock 1991; Wachen et al. 2012). However, many programs that serve adults still operate by offering remedial coursework separately from vocational or technical content.

Big data offers an opportunity for government, as it has for businesses. Airline companies use data collected on plane arrivals and departures to make efficient decisions about pricing (McAfee and Brynjolfsson 2012). Companies collect data on consumers' buying and web search practices to tailor marketing strategies to specific individual preferences. Similarly, governments need data on individuals to identify and target needed services. The big data revolution offers an opportunity to develop the tools needed to guide long-term investments in the public sector. Moreover, the past few years have seen a resurgence in the use of predictive modeling tools in conjunction with big data resources to help government understand the long-run impacts of potential policy changes (Ghaffarzadegan, Hawley, and Desai 2013; Hur, Ghaffarzadegan, and Hawley 2015; Kuhn and Johnson 2013). While big data needs to be seen through privacy regulations and government legal rules designed to ensure confidentiality of administrative records, there is a wealth of practical experience at the state and local levels that can help government make the best use of data for performance improvement.

WHAT IS BIG DATA?

Big data emerged in the 1990s as a term that was applied to the physical size of data, and generally referred to data that would not fit on a single computer or mainframe. In more recent years the term has been applied to the volume of data meant for analysis, irrespective of the size of the computer (Lewis 2018). In this regard, data from a single county unemployment insurance system might not qualify as big data because there are very few people who receive benefits. In contrast, data from Medicaid will qualify as big data. A single file from the Ohio Medicaid department in 2017 included 4.7 million rows just on prescription drugs.

A second way big data is distinguished from survey or administrative data in general is the complexity of the data elements. Unemployment Insurance wage records cover all people working in a state, and therefore contain a lot of records. But wage records usually have only four to six columns worth of data. In contrast, the K–12 education data in Ohio include well over 100,000 variables on approximately 1,000,0000 students annually. This complexity makes the education data much more difficult to analyze using traditional methods, and also reinforces the need for different methods of data storage and linkage using a database instead of relying on a traditional analytical file (Foster et al. 2017).

The early work of D.J. Patil (Lewis 2018, p. 147), who was the first chief data scientist for the U.S. federal government under President Obama, provides an interesting example of big data. In the early 1990s, Patil required massive amounts of data to test mathematical theories while working on his PhD. At that time, one of the largest stores of data was on weather systems and held by the National Oceanic and Atmospheric Administration. Weather data generates as much information as the entire Library of Congress on a daily basis. It is also complex, in that weather data includes information on different physical and human systems (Lewis 2018). Patil downloaded a great deal of this data and went on to conduct data science work in defense, education, and the private sector before his role as the chief data scientist.

WHAT KIND OF DATA DOES GOVERNMENT HAVE ON THE WORKFORCE?

Big data is challenging for government, as the amount and type of data resources maintained on the workforce increase in size. USDOL and associated agencies have always needed to maintain data on wages and salary to enforce regulations and comply with requirements from Congress on performance reporting. Federal labor and education data can be thought of in three categories: 1) survey data, 2) data reported from employers or states to comply with certain policies, and 3) real-time labor market data. Each is dealt with in turn in the following sections.

Survey Data

Three primary survey files produce data on employment on an annual basis: the Current Population Survey (CPS), a monthly household survey of 60,000 units; the Current Employment Survey (CES), a payroll survey of 147,000 firms representing 634,000 employees monthly; and the Occupational Employment Statistics (OES) program, done twice a year to a random sample of 200,000 establishments. Table 1.1 provides an overview of the data sources from survey data.

These three surveys provide different data to help us understand the workforce. The CPS is best for labor market information and provides demographic data. The CES is better for measures of employment, hours, earnings, and industry of employment. The OES collects data on occupations and wage rates.

The data in the CES are divided into 10 sectors (e.g., manufacturing, finance, education), and within each sector there is information on union membership and representation, average hours worked, average hourly wages, and workplace-related deaths and injuries.[1] The OES categorizes data into 800 occupations. In each occupation, it provides wages at percentiles, the percentage of the industry covered by that occupation, and geographic profiles to show concentrations of each occupation.[2] The CPS focuses on documenting employment and unemployment.

Table 1.1 Survey Data Collected on the U.S. Workforce

Survey name	Organization collecting data	Sample	Regularity
Current Population Survey	Bureau of Labor Statistics	60,000 households	Monthly
Current Employment Survey	Bureau of Labor Statistics	147,000 firms	Monthly
Occupational Employment Survey	Bureau of Labor Statistics	200,000 firms	Semiannual
Job Openings and Labor Turnover Survey	Bureau of Labor Statistics	16,000 firms	Monthly
O*NET (Occupational Information Network)	Employment and Training Administration	1,110 occupations	Ongoing

The three surveys have strengths and weaknesses. One shared weakness is that they are samples of companies, households, or individuals, which means that specifically for small units, such as small cities, the number of cases representing that frame is very small. The CES and OES use samples to generate estimates and do not provide exact employment numbers. The sampling results in margins of errors, especially for smaller metropolitan statistical areas. The surveys are corrected after the fact, and data are used on an annual basis differently from the monthly data.

Other survey data are important for workforce planning at the federal level. The Job Openings and Labor Turnover Survey provides general information on workers who quit or were laid off in the last month.[3] The decennial census offers the best estimate for total number of citizens in the workforce and categorizes workers by race and age. Many workforce analysts use the American Community Survey (ACS), an annual census sample that provides extensive information on demographic, workforce, and employment characteristics of residents at varying levels of granularity. The Occupational Information Network (O*NET), funded by the Employment Training Administration, operates a survey system that provides information on occupations and is used extensively in career development or supply and demand work by state and local government.

Additional government agencies, private companies, and associations provide employment data through surveys. For example, the Department of Health and Human Services estimates the supply and demand of 35 different types of health care workers, and the Association of American Medical Colleges provides a more detailed analysis of physicians in the workforce. In the private sector, Economic Modeling Specialists International uses data from the Departments of Labor, Education, and Commerce to create customized labor market data systems for its customers.

Administrative Data

One of the most important forms of data used by analysts and government are administrative records. These include UI wage records, statutory program data kept or reported to federal agencies to comply with legislation or regulation, and local data kept by states (Table 1.2).

Table 1.2 List of Administrative Data Systems Commonly Included in Longitudinal Data

Labor and Economics	Education	Other
Unemployment Insurance Wage Records (UI)	Education Management Information System (EMIS)	Temporary Assistance for Needy Families
Quarterly Census on Employment and Wage (QCEW)	Higher Education Information System	National Student Clearinghouse
Workforce system administrative data (UI benefits, WIOA Title I, Wagner-Peyser Act/ES	Teacher licensure files Value added data on student/teacher performance	Bureau of Motor Vehicles

These data have always existed but have only recently been used systematically across states.

The scope of administrative records depends on the state. The core UI wage record file exists in all states because it is required under the terms of the federal-state partnership for UI. The legal foundation of the current wage record system is based in the Federal Unemployment Tax Act of 1937, which set up a federal tax to cover unemployed workers. As part of the tax, states were asked to build (over time) a way of reporting earnings on a quarterly basis. The statute at the federal level currently establishes the framework for reporting wage records as part of the administration of UI (Workforce Information Council 2014). Moreover, at the state level, the statute defines specific elements to be covered by wage records.

These wage record data are universal and come directly from firms, meaning they are the single most important source of data on employment outcomes. However, records are very thin, containing information on only the quarter of employment, the amount earned in that period of time, and the number of weeks worked in a specific quarter. It is important to note that state coverage of wage records varies by state. Some states, such as Alaska, have occupation in the record, and others such as Oregon or Minnesota have hours worked.

A wide range of other state-administered data sources are used to understand workforce programs. Virtually all state systems exist as

record-keeping sources for federal programs. The largest are the data systems that supported the Workforce Investment Act (1998–2014) (WIA) and now support the Workforce Innovation and Opportunity Act. Specific statutory programs, such as the Adult, Dislocated Worker, or Youth formula programs and adult literacy programs, each have their own state data systems. The systems themselves are not standardized in terms of the data elements they include or the legal and governance systems that allow them to be used for evaluation and planning purposes. Some important exceptions in data collection practices across states will be discussed in Chapter 4.

Administrative record systems exist to respond to federal—and to a lesser extent—state statute. As a result, they are usually created for a specific programmatic purpose, and not for a general policy or research activity. States such as Ohio, for example, have separately maintained data systems for all of the statutory WIOA programs (i.e., WIA Title 1, Adult Literacy, Wagner-Peyser, and Vocational Rehabilitation). These systems capture common aspects of programs, such as enrollment, participation, program services, and exit. However, states capture these data in different ways, with no standardized definition of key terms, such as *participant*. The lack of common definitions makes the production of dashboards and scorecards challenging, as we will describe in Chapter 4.

Moreover, the technical development of the systems is highly variable. In Ohio, for example, the adult literacy data system is quite extensive and maintained by a state university. It has a well-documented technical structure. Other programs do not have the same investment in data collection and maintenance. For example, the data system for the Perkins-funded adult programs has not received the same level of support and is maintained mostly by analysis of data collected for other state purposes. This gap in technical sophistication of the data systems is driven in part by federal requirements for performance reporting.

Beginning with Florida in the 1980s and then Missouri and Maryland, states have sought to use administrative records to enable better understanding of student progress and success, often in the vocational or higher education sectors. Florida, for example, built the Florida Education and Training Placement Information Program system (FETPIP) to answer critical questions about student outcomes in higher education and employment. Currently, FETPIP uses wage record data to follow up

on graduates from high school and higher education, as well as workforce development programs.

Early studies using state records by Stevens (1989) describe emerging programs in Missouri, Texas, and Florida—states that were specifically focused on expanding the use of administrative data systems for tracking vocational education completers. Over the years, Stevens and other scholars developed wage records as a tool for understanding the outcomes of many different state-specific and federal programs. The ADARE consortium, funded by the USDOL between 1998 and 2012, illustrates the varied uses to which states and researchers have put wage records. ADARE sponsored studies on a range of topics, including vocational education, welfare reform, food stamp dynamics, and child support (Stevens 2012).

At the federal level, some new tools take state-collected administrative records and use them in significantly different ways. The Longitudinal Employer Household Dynamics program (LEHD) is a U.S. Census project that collects state-level UI wage records and makes these data available through the census research application process. The LEHD also provides access to aggregate records that offer very useful information to researchers and planners in state and local government (Abowd, Haltiwanger, and Lane 2004). In recent years, economists and social scientists have increasingly used the LEHD to study labor markets and human capital, including workers in scientific fields and the low-wage labor market (Andersson, Holzer, and Lane 2005; Lane et al. 2015; Weinberg et al. 2014).

Real-Time Labor Market Data

Many organizations have built real-time labor market data systems. No list will be comprehensive, but the tools that appear to have the widest reach are discussed below.

Since 2008, the Conference Board has been using human capital analytics to develop an Employment Trends Index (ETI), which is a forecasting tool for future employment outcomes and is published monthly, soon after the BLS report. The first indicator of the ETI is the Consumer Confidence Survey®, which measures job seekers' confidence in finding a job. These data are combined with Real Manufacturing and Trade Sales computed by the Bureau of Economic Analysis, industrial

production statistics provided by the Federal Reserve Board, and the percentage of firms unable to fill jobs, as calculated by the National Federation of Independent Business Research Foundation.[4]

The ETI is useful for making predictions about the labor market. For example, according to the Conference Board, an increase in the percentage of firms unable to fill positions is strongly correlated with a future peak in employment. In the beginning stages of economic weakness, there is a decrease in the ratio of temporary to full-time workers, a decline in confidence in finding a job, and an increase in UI claims (Levanon 2008).This information may be helpful during the development of workforce policy and for businesses to plan for future hiring and compensation.

A second tool, from Boston-based Burning Glass, is called Labor Insight, which uses web postings of jobs to provide information on hiring and training trends. Burning Glass and the Conference Board sell analytical services to government and the private sector to help them understand how to improve analysis of the labor market and hiring trends.

Moreover, Monster.com produces reports on the labor market for cities, which is combined with labor market information, and the ADP Research Institute offers a monthly set of employment reports that are used as a national estimate on employment and unemployment.

WHAT DOES GOVERNMENT DO WITH THE DATA IT HAS?

Government makes decisions with workforce and economic data at regular intervals. These kinds of decisions are understood but not documented in any consistent way. For example, reporters and business analysts remain fixated on the internet when USDOL releases the *Employment Situation Report* rate each month. A simple search of stories on National Public Radio's Planet Money podcast shows at least two shows per month on the unemployment rate, usually the day the data are released. But these reports on the employment situation are actually based on two complex surveys from the CPS and the CES.

Based on these surveys, the *Employment Situation Report* provides detailed data on employment, unemployment, and the condition of the labor force more generally.

I am not aware of any formal study of how these data are used in decision making, although employment statistics have an impact on public opinion, hiring, wages, and even stock prices. In effect, government gathers these data to measure the temperature of the labor market as an indicator of relative health. However, because consumers and analysts alike are liable to misinterpret or see phantom patterns in the data from the *Employment Situation Report*, it is difficult to make an accurate diagnosis of the underlying problems.

A second example might help explain why we need better theory on how data are used in decision making. Government consistently focuses on using economic data to forecast training and education supply. The BLS carries out formal projections on the demand for workers on a regular basis under its Employment Projections Program.[5] The current BLS methodology is strongly related to the historical methods used by the manpower planning practitioners beginning in the 1960s. Obviously the BLS has improved its methods for forecasting labor demand by occupation and educational level, but the fundamental goal is very similar to those outlined in the original work by international organizations (Richter 1986; Spalletti 2008). However, other organizations, such as Georgetown University, also produce estimates of the number and occupational distribution of these workers. The Georgetown methodology is quite different in some respects, particularly in how it relaxes the assumptions on the distribution of educational credentials in specific occupations (Carnevale, Smith, and Strohl 2010). The two sets of estimates offer quite different numerical estimates of the labor needed in the current economy.

Neither federal nor private estimates of labor supply and demand answer all the questions that local government, educational institutions, or state entities have. Stating that an occupation is supposed to grow does not answer questions about how much it should expand, or if the expansion should be static or needs to vary over time.

WHAT DATA ARE THE GOVERNMENT LACKING?

Despite the massive amount of data collected on the workforce, there are some key areas where government does not collect data, as well as some gaps in the data systems themselves. The Workforce Data Quality Campaign (WDQC) is a nonprofit association that promotes the use of administrative data in state policy. Its annual survey captures data on a state-by-state basis on a set of standardized workforce questions, such as what linkages exist between postsecondary education data and the workforce. The data consistently indicate problems with workforce data coverage. The most fundamental gaps tend to be around industry credential data and the availability of data in consumer-friendly forms. For example, 28 states reported in 2015 that they had access to an interagency council on workforce data, but only 9 states produce dashboards for policymakers, and only 3 provide data on industry credentials (WDQC 2015). The problem with the WDQC work, however, is that it just documents the availability of state data. Additional work needs to be done to understand the quality of the data. The following data are available in a very limited fashion from government and require action to develop the kinds of data systems that government needs.

Occupational Data

In general, states do not collect data on the occupation or job title for workers (Workforce Information Council 2014). One exception is Alaska, which provides these data only because of the state's unique income support provided to all residents. Without occupation measured directly, states are unable to document the kinds of work that are in demand in ways that can be conveyed clearly to educational institutions. The only alternative is to use crosswalks between degree credentials or industry of employment instead of the job title to generate an occupational code for each job. However, the crosswalks are often outdated and designate occupation based on what workers were trained for in college or the industry where they are currently employed. For people who graduate in welding or another technical area, the lack of a job title is not as critical as for business or humanities graduates, who often end up working in a wide range of industry and occupational careers.

Rate of Pay

The current UI wage record data—the core data used by researchers to understand employment outcomes at the state level—do not include data on hourly pay, they only include data on the total amount of pay received during that time period. There is no information on pay rate because it is not a required element under federal statute, and also it is not required to pay unemployment insurance. Some states have other items such as number of weeks worked or the occupation of the industry (Workforce Information Council 2014). This lack of data creates significant problems for understanding the wage outcomes of workforce training programs. For example, if someone graduates from an adult education program and earns a very low quarterly amount, as happens almost universally, we do not know if this is due to a low pay rate or if the individual is working part time. Without these data, we cannot calculate hourly earnings or examine the number of full- or part-time workers.

Industry Training Data

There is virtually no information regarding on-the-job training in industry. The last surveys the federal government conducted of employers on this issue were the 1995 Survey of Employer Provided Training and the 1997 National Employer Survey (Lerman, McKernan, and Riegg 2004).[6] Moreover, there is no mechanism by which participation in on-the-job training is captured in administrative records. Any training workers receive after they leave a government institution is not captured by workforce data systems, preventing analysis of larger questions about what happens after workers begin careers. Government spends a significant amount of money on companies through tax subsidies to bring jobs to communities or through state-sponsored training programs like California's Employment and Training Panel. However, very little of this information makes its way into evaluation statistics (Gorman et al. 2004; Moore et al. 2003).

CONCLUSION AND DESCRIPTION OF THE BOOK

This chapter has provided an introduction to workforce data and an overview of some of the key concepts necessary to the book. Specifically, I articulate an idea that while we have a great deal of data on the workforce in the United States, governments are underutilizing these data in decision making.

This chapter also asks how data systems can be made more responsive to business and government and applied to workforce systems. We are looking for a quicker and more effective turnaround between evaluation evidence and organizational improvement.

As a baseline, the chapter describes the kinds of data governments collect. We differentiate between three kinds of workforce data: survey data, administrative records, and real-time employment information. Each data source provides a piece of what governments require to establish data systems. While the primary focus of the book will be on the use of administrative data, the data tools we describe in Chapters 4 and 5 make use extensively of both survey and real-time data. Using the concepts from this chapter, we turn in Chapter 2 to the workforce system and its evolution. In so doing we describe the public policy landscape that we develop data systems within.

Chapter 2 describes government's current role in the workforce. I briefly discuss important characteristics of government in the workforce, both for high-skilled workers and low-wage individuals. I describe the range of government interventions, from the high-skill training programs run by the National Institutes of Health to programs like WIA, designed to improve the work participation and earnings of low-wage workers.

Chapter 3 provides an overview of how governments use data to make decisions and what the scholarly literature says about the ways government can strengthen decision making. Additionally, the chapter describes federal and state performance management systems in the four workforce areas. I offer examples from prior policy regimes under WIA and the Job Training Partnership Act that illustrate some of the different objectives of the current performance systems.

Chapter 4 gives specific examples of technical systems used to provide state performance management systems in the workforce area. We

review two key systems, Ohio and Washington, both of which have been implemented to allow the states to monitor performance of workforce programs. The chapter concludes with some advice for readers interested in the legal and governance issues that arise when a state data system is established.

Finally, Chapter 5 explores technical developments that states have made in performance management for the workforce. Specifically, I describe examples in three areas: scorecards, dashboards, and data visualizations.

The overall goal of this book is to describe new ways government is using tools to inform decisions about the workforce at the state and local levels. The book moves beyond standardized performance metrics designed to serve federal agency requirements and discusses how government uses tools that can be used to provide up-to-date information for government.

Notes

1. See Current Employment Statistics, http://www.bls.gov/jlt/jltover.htm#scope and Job Openings and Labor Turnover Survey, http://www.bls.gov/web/empsit/cesprog.htm#Coverage (last updated February 1, 2019).
2. See Occupational Employment Statistics Overview, http://www.bls.gov/oes/oes_emp.htm (last updated February 1, 2019).
3. "Job Openings and Labor Turnover Survey Overview," Bureau of Labor Statistics, https://www.bls.gov/jlt/jltover.htm#scope (last modified July 14, 2014).
4. "The Conference Board Employment Trends Index," Conference Board, https://www.conference-board.org/data/eti.cfm (accessed January 14, 2020).
5. The Employment Projections program uses data from the Occupational Employment Statistics, Current Employment Statistics, and the Current Population Survey. The methodology is clearly outlined in. https://www.bls.gov/emp/documentation/projections-methods.htm (accessed October 10, 2019).
6. Survey of Employer Provided Training was provided in 1993 and 1995. My research shows that this was the last time the federal government actually did a formal survey of firm-supported training.