



THE CENTER FOR  
**OPEN DATA ENTERPRISE**

# **Open Data for the Labor Market**

**An Open Data Roundtable with the  
U.S. Department of the Labor**

# TABLE OF CONTENTS

Foreword .....	1
Introduction.....	2
Background: Federal Labor Data.....	2
The Department of Labor Open Data Roundtable.....	6
Findings.....	7
Successes .....	7
Challenges .....	7
Cross-Cutting Issues.....	7
Occupational Information .....	8
Skills Information .....	8
Career Pathways .....	9
Proposed Solutions.....	9
Cross-Cutting .....	9
Occupational Information .....	10
Skills Information .....	11
Career Pathways .....	12
Next Steps.....	13
Appendices .....	14
A - Acronyms.....	14
B - Agenda.....	15
C - Participants.....	16
D - Sponsors.....	17
E - About the Open Data Roundtables.....	23
F - About the Center for Open Data Enterprise.....	24

*Special thanks to Pam Frugoli for her initiative and ongoing support.*

For questions about this report or the Open Data Roundtables, please contact the Center for Open Data Enterprise.

Joel Gurin, President & Founder - [joel@odenterprise.org](mailto:joel@odenterprise.org)

Laura Manley, Partnerships & Programs - [laura@odenterprise.org](mailto:laura@odenterprise.org)

Audrey Ariss, Research & Design - [audrey@odenterprise.org](mailto:audrey@odenterprise.org)

Katherine Garcia, Communications & Outreach - [katherine@odenterprise.org](mailto:katherine@odenterprise.org)

# FOREWORD

The U.S. Department of Labor (DOL) is leading the effort to ensure that America's workforce is equipped with the skills, education, and experience needed to succeed in the modern global economy. Critical to that effort is the collection and dissemination of timely, accurate, and relevant workforce data. These data can inform us about the state of our economy, what industries and occupations are in demand, and skills and credentials that workers need to stay competitive in the labor market and develop competencies that are sought by business. The Department of Labor has been working over the past two decades to be a leading provider of that information. Jobseekers, dislocated workers, recently graduated students, veterans, new immigrants, and businesses can all benefit from this effort. The Department of Labor is continually seeking ways to improve data collection to increase efficiency and quality, and enhance the way in which data are presented and made available to make it most beneficial to users.

The DOL Open Data Roundtable on November 16, 2015 was a forum of some of the leading individuals in data collection and analysis from the fields of business, academia, and government. The DOL roundtable had a dual focus; in addition to seeking ways to make government data more open and accessible to users, it also focused on identifying ways to utilize data publicly available on the Internet to inform and augment government data about the labor market. In particular, DOL is exploring ways to identify emerging skills and occupations and to monitor the pace and nature of change in labor market demand for a skilled workforce. The roundtable provided an opportunity to discuss the efforts being undertaken by various organizations, identify opportunities to work together, and discuss potential for meeting the challenges that still exist.

As with any undertaking of such scope, we understand that this effort cannot be done alone. The world of 'big data' is one which holds great potential but also presents challenges. Such challenges must be met with collaboration. This roundtable was designed to initiate idea sharing and partnerships between government and stakeholders from the private sector and academic research. In our world of rapidly-changing technology, there are many entities that have similar interests along with different strengths and perspectives — and perhaps even data — to contribute to the ongoing goal of developing a better workforce information ecosystem.

Learning from these discussions is one of the ways that the Department can ensure that it is providing the useful information to equip workers to face a new age in technology and globalization and to meet the needs of business. We look forward to working together to learn from and test the options and ideas discussed at the roundtable and presented in this report.

*Byron Zuidema*

Deputy Assistant Secretary

U.S. Department of Labor

Employment and Training Administration

*Michael Horrigan*

Associate Commissioner

U.S. Department of Labor

Bureau of Labor Statistics



# INTRODUCTION

On November 16, 2015, the U.S. Department of Labor and the Center for Open Data Enterprise co-hosted an Open Data Roundtable in Washington, DC. The Roundtable focused on federal occupational data, particularly information about jobs and the skills needed for them, and the ways that this data can be made more accessible and usable. It brought together experts from the Department of Labor, the Executive Office of the President, academia, nonprofit organizations, and the private sector.

This event was a part of the Open Data Roundtable Series conducted by the Center for Open Data Enterprise.<sup>1</sup> Open data – free data, accessible online, that anyone can use and republish without restrictions – is being recognized as a major public resource. The Roundtables are action-oriented dialogues that bring together government agencies and the organizations that use their data. These dialogues are designed to help extract the value of each agency’s data reserves by taking a demand-driven approach and focusing on data users.

This report combines participant observations, feedback, and suggestions related to the topics discussed at the event. It does not represent the views of all participants, the Department of Labor, or all users of Labor data. The report is being released as a public document to encourage further input, dialogue, and improvements in open data. It is designed to be used by officials in the Department of Labor and other government agencies, users and providers of occupational and skills data, and members of the media and the general public.

## BACKGROUND: FEDERAL LABOR DATA

### O\*NET DATA

The Occupational Information Network, or O\*NET, is a comprehensive federal source of data on occupational characteristics and requirements, and was a focus of the Labor Open Data Roundtable. O\*NET was first published in December 1998 by the Department of Labor Employment and Training Administration (ETA) and is updated on an annual basis. The network uses standardized questionnaires to collect data from statistically random samples of businesses and workers. (For more information on O\*NET visit the O\*NET Resource Center website.<sup>2</sup>)

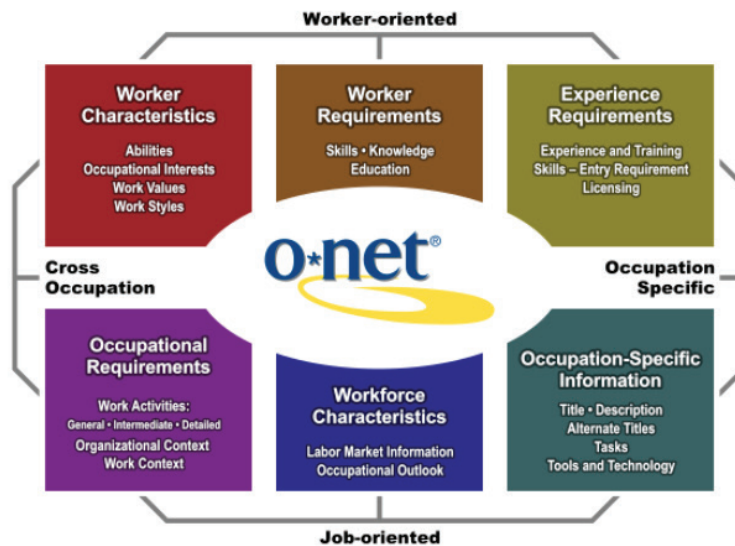
For each occupation, O\*NET provides the following information:<sup>3</sup>

- Worker requirements: the skills and knowledge required to perform the work
- Worker characteristics: the abilities, interests and values needed to perform the work
- Experience requirements: the training and level of education and experience needed for the work
- Job requirements: the work activities and context, including the physical, social, and organizational factors involved in the work
- Work performed: the tasks and work activities carried out on the job, along with the tools and technology required to perform those activities
- Labor market information: data including current and projected employment, the industries where workers are employed, and the pay scale for the work. These types of data are not collected directly by O\*NET but come from other linked sources, such as the Department of Labor’s Bureau of Labor Statistics (BLS) and state labor market information agencies.

1 The Open Data Roundtables were originally launched as an initiative of The GovLab at New York University.

2 O\*NET Resource Center. <http://www.onetcenter.org>.

3 The O\*NET Content Model. <http://www.onetcenter.org/content.html>.



## STANDARD OCCUPATIONAL CLASSIFICATION (SOC)

O\*NET uses an occupational taxonomy based on the Standard Occupational Classification (SOC) system, which is issued by the Office of Management and Budget (OMB) and used by federal statistical agencies to classify workers into occupational categories for the purpose of data collection, analysis, and dissemination. The SOC makes it possible to compare data from different federal sources, to aggregate data on employment from a wide range of job titles, and to track data over time to identify changes in the labor market and the growth of different occupations. The O\*NET occupational taxonomy, called the O\*NET-SOC, builds on the SOC by adding detailed information about the kinds of skills required for each type of job, including the tasks and work activities performed, the tools and technology utilized on the job, and related preparation requirements.

All federal agencies that collect data on occupations are required to use the SOC system to provide comparable information. The current version of the SOC was published in 2010, and the next edition is scheduled for publication in 2018. This schedule aligns with the 2017 update of the industry classification system (North American Industrial Classification System, or NAICS), making it possible to add new occupations and industry classifications simultaneously for data series that use both.

## BUREAU OF LABOR STATISTICS OCCUPATIONAL DATA

The Bureau of Labor Statistics Occupational Employment Statistics (OES) program includes information on occupational employment and wages as well as distribution of occupational employment by industry. It publishes employment information and wage estimates by area for over 800 occupations. Occupational data is provided for 580 local areas and 430 industry aggregations. This data is acquired through a national survey of 1.2 million business establishments, collected over 3 years with 400,000 establishments surveyed each year. These samples are stratified by industry, size, and area. Because the data is currently collected from a rolling sample, the OES provides a limited ability to compare wages over time.

Beginning in November 2015, all employers in the OES survey have been asked to provide job titles for all their workers. Small employers are mailed short, open-ended forms and are asked to provide job titles, duties, and wage ranges. Large employers are asked to provide an electronic file with job titles and wage rates. Employers will sometimes send additional information such as department, hours worked, gender, and race for each staff member.

Each job title is assigned an SOC code by OES staff. BLS is currently developing a system of auto-coding based on each

job title to make the process more efficient and consistent. This will also allow for greater access to employer job titles to identify potential new and emerging occupations, and more specific wage estimates. It also shifts the task of coding each occupation from the employer to the state and BLS employees.

Other data published by BLS includes 10-year employment projections for SOC detailed occupations. These projections account for job openings to replace workers leaving the occupation, as well as growth in those occupations. The projections program also provides typical entry level requirements for education and training for each occupation. Another resource is the Occupational Outlook Handbook, which describes over 300 occupations including information on occupation requirements. The Current Population Survey provides some demographic data, and the Survey of Occupational Injuries and Illness provides data on worker fatalities, illness, and injury rates.

## DATA APPLICATIONS

The labor market involves three kinds of stakeholders in what can be considered an “employment triangle”: Employers, employees/job-seekers, and intermediaries (job counselors and educators). For the job market to function efficiently, all three need a common understanding of the nature and requirements of different occupations. Employers can use information about occupations and skills to make decisions on compensation, training, and where their business should be located. Job-seekers can use labor data to explore their career options and make decisions on education and training. Job counselors and other labor market intermediaries use the data for career guidance and helping with job searches.

Labor data fuels research on both a macro and micro level. On a national level, this data provides insight into national competitiveness, growth projections for different industries, potential labor shortages, and skills gaps in the U.S. workforce. On a detailed level, labor data can give an in-depth picture of the skills that industries, occupations, and tasks require.

## RECENT INITIATIVES

O\*NET is a widely used, core resource, and efforts are under way to make it as timely, accurate, and relevant as possible. In 2010, the National Research Council of the National Academies conducted a review of the O\*NET system and approaches to improving it.<sup>4</sup> Elsewhere, a number of recent articles and blog posts have tracked updates to O\*NET.<sup>5</sup> The Department of Labor is working to identify ways to make O\*NET an even more effective tool for skills training and for matching workers to jobs. OMB and an interdepartmental Policy Committee are working to update the SOC classification system to assist both data collectors and users.

The Department of Labor is leading several projects, discussed at the Roundtable, that are designed to improve occupational data:

- Work is under way to convert the Occupational Employment Statistics program to a national time series, making it possible to compare job categories on a year-by-year basis. This analysis should help identify occupations that are growing rapidly and where workers are in demand.
- The BLS is now exploring “autocoding” of job titles and wage levels to increase the speed, efficiency, and timeliness of data collection for the OES program. This approach relies on computational linguistics and automates the classification of textual information. It is now being tested with O\*NET codes and titles on employer-reported job titles.
- The Department of Labor and other researchers are working to find ways to map career pathways more ac-

4 A Database for a Changing Economy. <http://www.nap.edu/catalog/12814/a-database-for-a-changing-economy-review-of-the-occupational>

5 Tyrone Grandison, “O\*NET today and beyond.” 18F, March 11, 2015, <https://18f.gsa.gov/2015/03/11/onet-today-and-beyond>.

curately. A number of career maps have been developed, but the career paths they outline – such as “help desk to software developer to CIO” – are often not realistic. Researchers at the City University of New York, Georgetown University, and elsewhere are studying online work histories to identify actual career paths empirically and learn how to help people advance their careers.

- The O\*NET project has developed a website called My Next Move<sup>6</sup> to make O\*NET information more accessible. There is a customized version of this website for veterans and a version in Spanish.

At the same time, the White House is leading collaborative efforts, the Workplace Data Initiative and the Skills Genome project, involving industry and government to develop a new skills taxonomy. This work draws on private-sector data from resumes, job postings, course information, and other sources. By combining government and private-sector data, this program aims to develop more real-time information, define the skills and roles needed for different occupations, and keep better track of the rapidly evolving labor market. The project’s immediate goal is to release a national dynamic skills library or co-op for use in 2016.

Additionally, in the 2017 Budget Request President Obama proposed to create “a \$500 million Workforce Data Science and Innovation Fund to create dynamic data sets on jobs, skills, and to help training providers and workers keep pace with rapidly changing job needs.”<sup>7</sup>

---

<sup>6</sup> My Next Move. <http://www.mynextmove.org>.

<sup>7</sup> FACT SHEET: President Obama Proposes New ‘First Job’ Funding to Connect Young Americans with Jobs and Skills Training to Start Their Careers. <https://www.whitehouse.gov/the-press-office/2016/02/04/fact-sheet-president-obama-proposes-new-first-job-funding-connect-young>

# THE DEPARTMENT OF LABOR OPEN DATA ROUNDTABLE

The Open Data Roundtable was planned as part of the Department of Labor's initiative to make its data more accessible and useful. The Roundtable brought together governmental and non-governmental stakeholders with a wide range of perspectives and expertise, including academic researchers, companies that work with Department of Labor data and companies that provide platforms for employers and job-seekers. The Roundtable was designed to discuss strategies to improve data on new and emerging occupations and skills, and to better link O\*NET with other occupational data sources.

The Roundtable included in-depth presentations from the Department of Labor's Chief Economist and officials from the Employment Training Administration and Bureau of Labor Statistics; the National Center for O\*NET Development; the National Economic Council; and the City University of New York. (A full list of presenters is shown on the Agenda in Appendix B.) The presenters described initiatives that are under way to improve and apply data related to the labor market.

With context from these presentations, Roundtable attendees participated in discussion groups of eight to ten people each for breakout sessions on three key elements of labor data: skills, occupations, and career pathways. The following sections summarize their insights on successes to date, critical challenges, and potential solutions.



# FINDINGS

## Successes

Roundtable participants described O\*NET as an essential resource with widespread use. Many companies and other organizations in the United States and internationally use Department of Labor websites for online career tools, ranging from information on skills and occupations to resume-writing templates on CareerOneStop.<sup>8</sup> Clear policies on O\*NET data and Labor's public reports on technical changes to the database have made O\*NET continually valuable and useful.<sup>9</sup>

Data on skills and occupations is holistic, free, and available online. The O\*NET information is globally relevant and is being used internationally as well as in the U.S. Occupational data from BLS and O\*NET is collected rigorously; the system is validated, reliable, and trustworthy. Roundtable participants especially cited the value of Labor's data on unemployment and the fact that its data covers "soft skills".

Finally, participants said that Labor's information on career pathways provides a good starting point for career counselors and college support staff. In particular, the My Next Move website, one of the suite of O\*NET portals, was held up as an example of a useful and user-friendly approach.

## Challenges

The Roundtable focused on current obstacles to using Labor data most effectively and possible ways to address them. This section summarizes some of the most important issues that participants highlighted.

### CROSS-CUTTING ISSUES

- **Lagging currency of information.** In a dynamic labor market, descriptions and definitions of occupations and skills change rapidly. Keeping up with the rapid emergence of new job categories and requirements is a major challenge for BLS and ETA. Participants recommended that industry help contribute to occupations and skills data to address this problem.
- **Increased personalization of jobs and the "long tail."** Jobs and the skills required to do them are becoming more personalized, making it difficult to aggregate data in a meaningful way. There is a need to standardize these descriptions or centralize the information, or develop another solution to solve this challenge.
- **Data presentation.** The O\*NET<sup>10</sup> website could be improved by identifying users and use cases, tailoring presentation to key users, and organizing information with a user-friendly interface that supports real-world search behavior. While O\*NET provides web services and application programming interfaces (APIs), some participants suggested that O\*NET become even more of an "API shop" to encourage developers to build on its data.
- **Multiplicity of related data sources.** Participants noted the need to organize the ecosystem of occupational data by connecting DOL data to other resources on skills, available jobs, or educational requirements. In addition to O\*NET, occupationally related information is now available from:
  - ▶ National Compensation Survey (NCS)
  - ▶ Occupational Employment Statistics (OES)

8 CareerOneStop <http://www.careeronestop.org>.

9 O\*NET Resource Center What's New? <http://www.onetcenter.org/whatsnew.html>.

10 O\*NET Web Services. <https://services.onetcenter.org>.

- ▶ Occupational Outlook Handbook
  - ▶ Competency Model Clearinghouse
  - ▶ Sector-specific reports
  - ▶ BLS employment projections
  - ▶ Current Population Survey (CPS) and American Community Survey (ACS)
  - ▶ The National Labor Exchange (NLx)
- **Conflict between the need for standardization, user needs and experience.** Overall, participants noted that SOC is not a good user-facing taxonomy and serves government better than job seekers or employers. In particular, both employers and employees describe their work in ways that are different from SOC categories or O\*NET taxonomy.
  - **Lack of regular and organized feedback.** The system would benefit from input on the types of data that would be useful to job seekers, employers, and other data users. Ongoing feedback could also help DOL refresh job titles and descriptions more quickly. Beyond feedback channels, crowdsourcing could help make data more current and complete.
  - **Incomplete and inconsistent data definitions and taxonomies.** Improvements are needed for consistency and to help clarify the relationship between skills, job tasks, tools and technologies. It is also important for taxonomies and data to be open, machine-readable, and applicable across many different systems and stakeholders. Both the O\*NET Center and efforts such as the Cass Project<sup>11</sup> and similar projects are trying to provide more general open source tools and infrastructure for this purpose.

## OCCUPATIONAL INFORMATION

- **Usability of the SOC system.** SOC classifications do not keep pace with the labor market, and companies may need help assigning SOC codes to wage records in states where it is mandatory. Overall, SOC data would be more useful if it was released in a more granular level. The SOC includes 840 detailed level occupational categories and the O\*NET system has added a number of additional detailed occupations totaling 974. However, some data users feel that neither captures additional variations within categories that can make a difference in wages.
- **Need for more “real-world” job classifications.** Many existing classifications and job definitions are out of date. The Department of Labor is seeking additional strategies to identify and collect data on new and emerging occupations from private sector and other sources. Participants also recommended considering more colloquial naming conventions for occupations and applying user feedback to refresh job titles and descriptions more quickly. However, the need to update job classifications should be balanced against the value of consistency, which makes occupations comparable across a time series.
- **Regional variations.** While occupational employment and wage data are localized there is also interest in more targeted data to reflect differences relevant to specific geographic areas.

## SKILLS INFORMATION

- **Interrelated skills data and taxonomy.** Participants flagged the need for a better taxonomy and hierarchies of skills, including distinctions between professional and soft skills and between skills and specific tasks. Other needs included:

<sup>11</sup> CASS. <http://www.cassproject.org>.

- ▶ A larger dictionary for skills.
- ▶ Better information on transferable skills and related skills. (For example, C++ and Java are closely related, and skill in one usually indicates ability to learn and perform in the other.)
- ▶ Linking to education: It is difficult to match skills to educational training. The system should connect to the national curriculum code, which is very detailed but disconnected from the labor market.
- ▶ Alignment with industry taxonomies and connections to industry credentials and licenses. For example, LinkedIn has a database of over 45,000 skills that can help inform Labor's skills taxonomy.
- ▶ Better ways to map skills to jobs, with better measures of the relevant technical skills needed for each job.
- ▶ Adoption of standards such as schema.org or HR Open Technology standards for job postings to assist in parsing emerging skills information from jobs postings.

## CAREER PATHWAYS

- ▶ **Need to connect information on occupational skills, real-world input, and curriculum development and training programs.** Job training curricula could be improved with a more data-driven approach. This could include adding both O\*NET data and complementary data from real-world job experience.

## PROPOSED SOLUTIONS

Roundtable breakout sessions proposed the following solutions in response to the challenges outlined above.

### CROSS-CUTTING

1. **Create a 'Wikipedia' for jobs and occupational information and skills.** A wiki could make it easy for employers, companies that manage occupational data, and others to contribute ideas and real-time information on skills and jobs data.

Proposal: Build a wiki that can serve as a platform for public engagement on jobs and skills data. This wiki would complement O\*NET and other official labor data sources with crowdsourced data. The Department of Labor could:

- Consider other possible government models. One example: OpenEI<sup>12</sup>, developed by the U.S. Department of Energy with semantic wiki software. Consider making the wiki like GitHub, which is designed to make it possible to track changes that users make between versions.
- Take the O\*NET typology as a base and allow users to contribute new occupational information that would extend the typology.
- Clearly state that much of the information on this wiki is not official government information. Provide metadata on source and provenance, so people who want to use it know the origin of the data and can decide how to use it.

Desired impact: Enable employers, job-seekers, and job counselors to improve skills training and job-matching with timely, easily understandable information on different job categories. Since employers and employees often describe jobs in ways that differ from SOC or O\*NET categories, this wiki would enable them to access information in a way that matches their frame of reference. At the same time, the wiki can provide

<sup>12</sup> Open Energy Information. <http://en.openei.org>

information that can be considered for integration into O\*NET as well.

2. **Undertake outreach and engagement around O\*NET.** More potential users could benefit from awareness of and information on ways to use the O\*NET information and websites.

Proposal: Develop a cohesive outreach and engagement program to communicate O\*NET's value and find partners to help enhance it. Outreach can include the following components:

- More visualizations, including using third-party plug-ins to help present the data visually
- MOOCs (massive open online courses) to train users on how to use the information
- Links to other relevant databases
- Highlighted use cases (e.g. community colleges using My Next Move) to encourage new users

Engagement strategies can include:

- Engaging the Small Business Administration to encourage small businesses to use O\*NET, perhaps by familiarizing them with the O\*NET Toolkit for Business.<sup>13</sup>
- Coordinating with other agencies on outreach, such as BLS and the Census Bureau.
- Developing ongoing feedback loops with users of Labor data. This could include formalizing groups that use Labor data to share information regularly.
- Working more closely with the National Labor Exchange (NLx). The NLx is a partnership designed to match job-seekers to job openings found on corporate career websites and state job banks. The NLx was mentioned as a resource in several contexts at the Roundtable, suggesting the value of working more closely with the Exchange overall.

Desired impact: Increased and more effective use of job and occupational information, including O\*NET data, as a resource for job-seekers and employers alike. As a result, improved skills training, job matching, and overall employment levels and job satisfaction.

## OCCUPATIONAL INFORMATION

3. **Draw on non-governmental sources to track the changing job market.** An overriding concern is the need to predict how the job market will change so that workers can be effectively trained in the future. Both academia and the private sector have information that reflects the evolving job market and that would benefit the federal labor data system.

Proposal: Leverage educational institutions and employers to capture evolving and emerging occupations.

- Work through U.S. and state Departments of Education to collaborate with university career services. Ask them to provide a feed of online postings from recruiters and send surveys to alumni communities. (Note: This may require universities to change their terms of agreement with employers.)
- Develop a templated online form companies can use to enter occupation and skills descriptions for job postings.
- Approach employers in select rapidly changing industries to develop more real-world naming conventions

<sup>13</sup>O\*NET Resource Center O\*NET Toolkit for Business. <http://www.onetcenter.org/toolkit.html>.

for occupations. (Example: BLS has worked with the film industry to develop occupational descriptions.)

Desired impact: National workforce development will be focused on the right capabilities for the future; educators and job counselors will teach the most marketable skills; job-seekers will learn the skills that will most benefit them.

## SKILLS INFORMATION

4. **Publish a new skills taxonomy.** There is currently no single skills taxonomy to match the Department of Labor's occupation taxonomy. O\*NET has taxonomies for Work Activities, for Knowledges, Skills, and Abilities, and for Tools and Technology. Many items in these taxonomies are classified to a global taxonomy—the United Nations Standard Products and Services Classification (UNSPSC)<sup>14</sup>. Adoption of a single, unified taxonomy would help to: describe how skills relate to specific jobs and industries, identify where there is an emerging demand for new skills, and show where skills need to be developed in the workforce.

Proposal: Develop a multi-stakeholder approach for refining and adopting a taxonomy of skills.

- Draw on international research – including the EU competency model; the potential OES autocoding of job titles and descriptions; the OECD classification of occupations; the European Skills, Competences, Qualifications and Occupations (ESCO); and the network for international policy making and cooperation in education and training, or NORRAG.<sup>15</sup>
- Work with the Department of Education and National Center for Education Statistics.
- Draw on private sector input. Companies have an incentive to improve potential employees' skills by incorporating information from human resources and payroll systems as well as industry associations.
- Explore getting skills information on an industry-by-industry basis. Do this by addressing the particular needs of each sector, perhaps starting with a structured industry such as healthcare. One example of this approach is the hierarchical model CompTIA has developed for the IT industry.
- Explore using a crowdsourced approach. Develop a Wikipedia for jobs (previously mentioned) that is crowdsourced for corrections. Use an iterative top down/bottom up approach, with interaction between government experts and the public.
- Connect to existing O\*NET data. Bundle O\*NET descriptions by skills as well as by occupations.

Desired impact: A more accurate, timely, and usable picture of necessary job skills will lead to improvements on many levels. Individuals will receive more helpful job training, employers will be able to recruit more effectively, and the national workforce will meet national labor needs.

5. **Use data mining and artificial intelligence to capture emerging skills.** A number of technical approaches now make it possible to extract information on job skills computationally. On the industry side, some companies have begun analyzing resumes, postings, and potentially position descriptions with machine learning to develop skills taxonomies. This approach could be more broadly used by the Department of Labor and others to develop a skills taxonomy (as described above) for national use.

Proposal: Build and use data mining and natural language processing capabilities within the Department of Labor, taking funding, resources, and capabilities needed into consideration. Include such approaches as:

<sup>14</sup> O\*NET Resource Center UNSPSC Reference. [http://www.onetcenter.org/dictionary/20.1/mysql/unspsc\\_reference.html](http://www.onetcenter.org/dictionary/20.1/mysql/unspsc_reference.html).

<sup>15</sup> NORRAG is an internationally recognised, multi-stakeholder network which has been seeking to inform, challenge and influence international education and training policies and cooperation for almost 30 years. <http://www.norrag.org>.

- Autocoding of job titles and resumes.
- Text analysis tools for content analysis.
- Automated analysis of changes over time.

Desired impact: Accelerate the timeline to adopt a national skills taxonomy with greater speed and efficiency than a strictly survey approach.

## CAREER PATHWAYS

### 6. Use longitudinal data sources to develop better data on career pathways.

Longitudinal data can help show how individuals have moved between jobs and determine what actual career pathways match the realities of the job market. The Department of Labor could analyze a number of these data sources to develop an understanding of career pathways to inform career counseling and education.

Proposal: Develop a new understanding of career pathways using data from the following sources:

- National Labor Exchange (NLx).
- Applicant tracking services such as Oracle SAP, SuccessFactors, payroll companies like ADP.
- Social Security and Census data.
- Occupational mobility data from the Bureau of Labor Statistics.
- National Longitudinal Survey of Youth (NLSY).

Additionally, work with partners who can contribute knowledge and data in other ways, for example:

- Private-sector companies such as Burning Glass, Indeed, and LinkedIn. Use these as sources of real-time data on skills to add to the Longitudinal Employer-Household Dynamics (LEHD) databases.
- Unemployment offices at the state level. If privacy concerns can be managed, it would be possible to get data from unemployment insurance (UI) and add occupational identifiers to the UI wage records. This would be a highly valuable data source since UI records have already been validated. (Note, however, that most states currently do not collect occupation data as part of their unemployment insurance system.)
- Universities that can capture alumni information on career paths after graduation.

Desired impact: The Department of Labor will have a clearer, more accurate picture of career pathways in different fields, which counselors and skills trainers can use to help their clients find jobs and advance their careers.

## NEXT STEPS

The insights and suggestions from this Roundtable can help inform two initiatives now under way: the Department of Labor's work to improve O\*NET, and the White House Workforce Data Initiative to develop better data and a taxonomy on skills for employment. The ideas and resources discussed at this Roundtable can be helpful to these current efforts.

The Employment and Training Administration (ETA) in the U.S. Department of Labor (DOL) is coordinating its work with the White House initiative, which Ryan Burke from the National Economic Council described at the Roundtable. ETA is participating in this effort through sharing of open data and utilization of APIs from new data repositories.

ETA is also working to develop strategies and greater capacity to leverage information from the millions of job postings and resumes on the Internet. DOL is considering a variety of ways to collaborate with other stakeholders involved in data aggregation and data mining efforts relevant to emerging skills and occupations.

The O\*NET project is currently leveraging tools that already mine data from job postings and resumes as well as providing curated data by occupation. For example, the O\*NET system has recently utilized real-time data tools to add over 3,000 new technology terms to the Tools and Technology file, has associated those technologies with relevant O\*NET occupations, and has classified them according to the global United Nations Standard Products and Services Classification. In addition, O\*NET has added indicators for 'hot' technologies or technology skill terms - the ones found most frequently in job postings in the past year - and made that information searchable and useful for both job seekers and curriculum developers. This information can help educators focus on important current skills when designing curricula for job training, and can help job seekers make better education and training choices.

Many of the emerging technology skills have names that may be familiar to those already working in the high tech sector but can be almost unintelligible to those who are started to learn these skills and hope to join the tech sector. For example, some data-mining efforts have found that the top 10 skills for the emerging occupation of "Data Scientist" are: SQL, Hadoop, Python, Java, R, Hive, MapReduce, NoSQL, Pig, and SAS. O\*NET research organizes these terms within the UNSPSC to define them and provide context. Moreover, customers who come to O\*NET OnLine can learn that R is "object or component oriented development software," and that SQL is "software related to database user interface and querying, and database reporting and management."

DOL is also considering the feasibility of developing a wiki of occupational information that business and experienced workers can add to voluntarily, and evaluating the pros and cons of this approach. A wiki could permit great flexibility and access, but would require awareness and action on the part of participants. It also has the potential to be less comprehensive or representative than other approaches. DOL would appreciate thoughts or feedback on a potential "occupational wiki" approach to help assess its potential value and utility.

In addition, the leaders of the O\*NET project want to make sure that individuals interested in open data know how they can now submit new items for inclusion in the O\*NET Tools and Technology database. When a user conducts a search in O\*NET OnLine by the Advanced Search for Tools and Technology<sup>16</sup> it will be matched to a UNSPSC classification and related O\*NET occupations if they exist, or the system will indicate that that particular Tool or Technology is to yet to be in the file. In either case, users will see a screen with a link indicating that they can provide feedback to help keep that file up-to-date.

---

<sup>16</sup> O\*NET Online Tools & Technology Search. <http://www.onetonline.org/search/t2>.



We welcome feedback on the Tools & Technology database. We encourage you to suggest tools and technologies where our coverage may not reflect the full range of an occupation's duties. All suggestions will be considered for a future update of the T2 database.

[Provide T2 feedback](#)



[Help](#)

[Find Occupations](#)

[Advanced Search](#)

[Crosswalks](#)

[O\\*NET Sites](#)

Was this page helpful?  

[Job Seeker Help](#) • [Contact Us](#)

Share:      

[Link to Us](#) • [Cite this Page](#)

Follow us:    

[About this Site](#) • [Privacy](#) • [Disclaimer](#)

O\*NET Web Services and APIS can be found at this link: <https://services.onetcenter.org>. DOL encourages all interested parties to use the system, provide feedback, and supply new information. DOL encourages all interested parties to use O\*NET Web Services, provide feedback, and share new information.





## APPENDIX A: ACRONYMS

ACS	American Community Survey
BLS	Bureau of Labor Statistics
CPS	Current Population Survey
DOL	U.S. Department of Labor
ESCO	European Skills, Competences, Qualifications and Occupations
ETA	Employment and Training Administration
LEHD	Longitudinal Employer-Household Dynamics
NAICS	North American Industrial Classification System
NCS	National Compensation Survey
NLx	National Labor Exchange
NLSY	National Longitudinal Survey of Youth
OES	Occupational Employment Statistics
OMB	Office of Management and Budget
O*NET	Occupational Information Network
SOC	Standard Occupational Classification
UI	Unemployment Insurance
UNSPSC	United Nations Standard Products and Services Classification

## APPENDIX B: AGENDA

- 9:00 a.m. Registration, coffee and refreshments
- 9:30 a.m. Welcome and opening remarks  
Byron Zuidema, Deputy Assistant Secretary, Employment and Training Administration, U.S. Department of Labor  
Mike Horrigan, Associate Commissioner for Employment and Unemployment Statistics, Bureau of Labor Statistics, U.S. Department of Labor
- 9:45 a.m. Structure of the day  
Joel Gurin, President, Center for Open Data Enterprise
- 9:50 a.m. Agency data briefings
- ❖ O\*NET Overview: Content, Taxonomies, Data Collection and Web Services  
Phil Lewis, O\*NET Technical Officer, National Center for O\*NET Development, North Carolina Department of Commerce
  - ❖ Bureau of Labor Statistics: Occupational Data & Standard Occupational Classification (SOC) Update Process  
Laurie Salmon, Division Chief, Occupational Employment Statistics, Bureau of Labor Statistics, U.S. Department of Labor
- 10:30 a.m. Break
- 10:40 a.m. Session 1: O\*NET data use cases - successes, challenges and opportunities
- 11:40 a.m. Reports out & group discussion
- 12:00 p.m. Lunch
- 12:45 p.m. Overview of the Skills Co-op Initiative
- 1:00 p.m. Session 2: Occupational and skills data - Improving relevance and timeliness
- 2:00 p.m. Reports out & group discussion
- 2:20 p.m. Break
- 2:35 p.m. Developing career pathways
- 2:45 p.m. Session 3: Leveraging O\*NET data for career development and job mobility
- 3:15 p.m. Reports out & group discussion
- 3:35 p.m. Next Steps  
Heidi Shierholz, Chief Economist, U.S. Department of Labor
- 3:45 p.m. Adjourn
- 4:00 p.m. Happy Hour at the Dubliner

# APPENDIX C: LIST OF PARTICIPANTS

## Government Agencies and Offices

### U.S. Department of Labor

The Department of Labor (DOL) is in charge of programs and laws that cover all facets of employment and work. It's mission is to foster, promote, and develop the welfare of the wage earners, job seekers, and retirees of the United States; improve working conditions; advance opportunities for profitable employment; and assure work-related benefits and rights.

#### Bureau of Labor Statistics

The Bureau of Labor Statistics of the U.S. Department of Labor is the principal Federal agency responsible for measuring labor market activity, working conditions, and price changes in the economy. Its mission is to collect, analyze, and disseminate essential economic information to support public and private decision-making.

**James Franklin** - Economist

**Mike Horrigan** - Associate Commissioner

**Kristen Monaco** - Research Chief

**Rebecca Rust** - Assistant Commissioner

**Laurie Salmon** - Occupational Employment Statistics Program Manager

**Michael Wolf** - Division Chief

#### Chief Evaluation Office

The Chief Evaluation Office (CEO) coordinates, manages, and implements DOL's evaluation program. CEO works closely with all offices and agencies throughout DOL to develop and implement evaluations that address priorities set by the Secretary and the agencies. CEO also collaborates externally with colleagues in other Federal Departments and in the professional evaluation and research community.

**Jon Simonetta** - Deputy Chief Evaluation Officer

#### Employment and Training Administration

The Employment and Training Administration (ETA) administers federal government job training and worker dislocation programs, federal grants to states for public employment service programs, and unemployment insurance benefits.

**Tiffany Smith** - Senior Policy Advisor

#### Office of Workforce Investment

The Office of Workforce Investment (OWI) is responsible for implementing an integrated national workforce investment system that supports economic growth and provides workers with the information, advice, job search assistance, supportive services, and training in demand industries and occupations needed to get and keep good jobs.

**Lauren Fairley** - Workforce Analyst

**Pam Frugoli** - Manpower Analyst

**Don Haughton** - Supervisory Economist

**Alex Nallin** - Workforce Analyst

#### National Center for O\*NET Development

The National Center for O\*NET Development develops the Occupational Information Network (O\*NET), a comprehensive system for collecting, organizing, describing, and disseminating data on occupational characteristics and worker attributes.

**Phil Lewis** - Technical Officer

**Jeremiah Morris** - Programmer and Web Developer

**David Rivkin** - Technical Officer

**Harry Russos** - Project Manager

---

## U.S. Department of Commerce

The mission of the U.S. Department of Commerce is to create the conditions for economic growth and opportunity. The Department works with businesses, universities, communities, and the Nation's workers to promote job creation, economic growth, sustainable development, and improved standards of living for Americans.

**David Langdon** - Economist and Senior Policy Advisor

---

## U.S. General Services Administration, 18F

18F is a civic consultancy for the government, inside the government, consisting of a team of designers, developers, and product specialists inside the General Services Administration.

**Lori Foster** - Fellow  
**EJ Kalafarski** - Presidential Innovation Fellow

---

## National Economic Council

The National Economic Council (NEC) coordinates policy-making for domestic and international economic issues, coordinates economic policy advice for the President, ensures that policy decisions and programs are consistent with the President's economic goals, and monitors implementation of the President's economic policy agenda. It resides within the Office of Policy Development and is part of the Executive Office of the President.

**Ryan Burke** - Senior Policy Advisor

---



## Companies, Nonprofit, Academic and Research Organizations

### Abt Associates

Abt Associates Inc. researches, evaluates, and implements programs in the fields of health, social and environmental policy, and international development.

**Karin Martinson** - Principal Associate

**John Trutko** - President

**Matt Zeidenberg** - Senior Scientist

### ADP

ADP is a global provider of cloud-based Human Capital Management solutions that cover HR, payroll, talent, time, tax and benefits administration.

**Tom Crowley** - Director, UI Agency Liaison

### Amazon Web Services

Amazon provides cloud computing services through Amazon Web Services to a range of clients.

**Ariel Gold** - Program Manager World Wide Public Sector

### Booz Allen Hamilton

Booz Allen Hamilton is a provider of management consulting, technology, and engineering services to the US government in defense, intelligence, and civil markets, and to major corporations and not-for-profit organizations.

**Cynthia Schuman Ottinger** - Deputy Director for Philanthropy Programs

### Booz Allen Hamilton

Booz Allen Hamilton is a provider of management consulting, technology, and engineering services to the US government in defense, intelligence, and civil markets, and to major corporations and not-for-profit organizations

**Julie Bernard**  
**Greg Caramanica**  
**Bryce Pippert** - Principal

### Burning Glass

Burning Glass is a developer of solutions for resume parsing, job matching, and real-time labor market analytics.

**Dan Restuccia** - Chief Analytics Officer

### Center for American Progress

The Center for American Progress is a progressive public policy research and advocacy organization.

**Ethan Gurwitz** - Research

### Center for Data Science and Public Policy, University of Chicago

The Center for Data Science and Public Policy brings together data science and public policy experts. Its mission is to conduct research and create computational and data-driven solutions to large-scale social problems in areas such as healthcare, education, sustainability, and community development.

**Matt Gee** - Data Scientist

### Center for Open Data Enterprise

The Center for Open Data Enterprise is an independent nonprofit organization that develops smarter open data strategies for governments, businesses, and other nonprofits by focusing on data users.

**Audrey Ariss** - Research & Design

**Katherine Garcia** - Communications

**Joel Gurin** - President

**Laura Manley** - Director of Partnerships & Programs

**Lorinc Thurnay** - Research Intern

---

### Center for Regional Economic Competitiveness

The Center for Regional Economic Competitiveness (CREC) is an independent, not-for-profit organization founded to provide policy-makers from around the world with information and technical assistance needed to formulate and execute innovative, regional, job-creating economic strategies.

**Haden Springer** -  
Manager Technical  
Assistance & Training

---

### The Conference Board

The Conference Board is a global, independent business membership and research association working in the public interest, providing economic data and analyses, research and best practices concerning management, leadership, and corporate citizenship, as well as public and private forums in which executives learn with and from their peers.

**Rick Clayton** - Associate  
Advisor

---

### Eduworks

Eduworks provides software-based solutions for converting digital content and courseware into new formats and for analyzing documents such as resumes, reports and design documents for content and quality.

**Robby Robinson** - CEO

---

### George Washington Institute of Public Policy

The George Washington Institute of Public Policy is a center for faculty and graduate students engaging in research on important public policy issues. Current areas of research include state and local fiscal policy, urban and regional economic development, active living research, the role of federal statistics in monitoring U.S. economic activity, housing policy, the economics of retirement, workforce development in the 21st century, and social welfare policy.

**Robert Sheets** - Research  
Professor

---

### The Georgetown University Center on Education and the Workforce

The Georgetown University Center on Education and the Workforce is an independent, nonprofit research and policy institute affiliated with the Georgetown McCourt School of Public Policy that studies the link between education, career qualifications, and workforce demands. The Center conducts research in three core areas with the goal of better aligning education and training with workforce and labor market demand: jobs, skills, and people.

**Nicole Smith** - Senior  
Economist

---

### George Washington University Department of Economics

The Department of Economics is part of the social and behavioral sciences in the Columbian College of Arts and Sciences at George Washington University.

**Tara Sinclair** - Associate  
Professor

---

### Trachtenberg School of Public Policy and Public Administration, George Washington University

The Trachtenberg School of Public Policy and Public Administration is a center for public affairs education, research, and public service at the George Washington University.

**Burt Barnow** - Amsterdam  
Professor of Public Service  
and Economics

---

### Indeed

Indeed.com is an employment-related search engine for job listings available in over 50 countries and 28 languages.

**Tara Sinclair** - Chief  
Economist

---

---

### **Integra Management Associates**

Integra Management Associates is an independent association management firm, specializing in the management of common interest realty associations.

**Darrel Sandall** - President & CEO

### **LinkedIn**

LinkedIn is a networking tool that helps users discover professional connections, recommended job candidates, industry experts and business partners.

**Pablo Chavez** - Vice President Global Public Policy and Government Affairs

**Craig Martell** - Senior Engineering Manager Learning and Skills Team

---

### **Markle Foundation**

Markle works to realize the potential of information technology to address previously intractable public problems for the economic security, health, and national security of all Americans.

**Robert Khedouri** - Managing Director and Chief Operating Officer

---

### **Microsoft**

Microsoft develops, manufactures, licenses, supports and sells computer software, services, devices and solutions that help people and businesses realize their full potential.

**NS Rana** - Solutions Architect

---

### **National Association of State Workforce Agencies**

The National Association of State Workforce Agencies (NASWA) is an organization of state administrators of unemployment insurance laws, employment services, training programs, employment statistics and labor market information and other programs and services provided through the publicly-funded state workforce system.

**Pam Gerassimides** - Assistant Executive Director

---

### **New York City Labor Market Information Service, City University of New York**

The New York City Labor Market Information Service develops customized, data-driven products and services for education and workforce practitioners and policy makers to help guide their efforts in achieving positive, meaningful labor market outcomes for the students and jobseekers they serve.

**Lesley Hirsch** - Director

---



# APPENDIX D: SPONSORS

The Center for Open Data Enterprise thanks its sponsors for their support of the Open Data Roundtables.



## Open Data Partner

**Microsoft** is a worldwide supplier of devices and services that help people and businesses realize their full potential. Customers use Microsoft's products to find creative solutions to business problems, develop breakthrough ideas, and stay connected to what's most important to them. For more information, see [www.microsoft.com](http://www.microsoft.com).



## Series Sponsors

**Amazon Web Services** offers a broad set of global compute, storage, database, analytics, application, and deployment services that help both large enterprises and start-ups move faster, lower IT costs, and scale applications. Amazon Web Services Worldwide Public Sector is helping government and education customers employ cloud services to reduce costs, drive efficiencies, and increase innovation across the globe. Public Sector organizations of all sizes use AWS to build applications, host websites, harness big data, store information, conduct research, improve online access for citizens, and more. For more information, see [www.aws.amazon.com/gov](http://www.aws.amazon.com/gov).



**PricewaterhouseCoopers** is a network of firms in 158 countries that delivers quality in assurance, tax and advisory services. PwC helps federal agencies meet the challenge of integrating financial and performance systems to help make decisions and improve accountability. Clients look to the PricewaterhouseCoopers Public Sector Practice to bring direct hands-on knowledge of federal standards for systems, internal controls, and financial reporting. The Practice assists clients through creating interactive data, developing an agile reporting and analytic framework, and identifying and implementing improvements to the data and information supply chain. To find out more, visit [www.pwc.com/publicsector](http://www.pwc.com/publicsector).



## Event Sponsor

**Booz Allen Hamilton** provides management and technology consulting and engineering services to leading Fortune 500 corporations, governments, and not-for-profits across the globe. Booz Allen partners with public and private sector clients to solve their most difficult challenges through a combination of consulting, analytics, mission operations, technology, systems delivery, cybersecurity, engineering, and innovation expertise. For more information, visit [www.boozallen.com](http://www.boozallen.com).



# APPENDIX E: ABOUT THE OPEN DATA ROUNDTABLES

## DESCRIPTION

The Open Data Roundtables ([opendataenterprise.org/convene](http://opendataenterprise.org/convene)) developed by the Center for Open Data Enterprise are action-oriented dialogues that bring together government agencies and the organizations that use their data. By hosting the Roundtables, the Center offers a low-tech solution to a high-tech problem. The Roundtables are designed to:

- Identify high-value datasets so agencies can address them as a priority;
- Develop solutions to make data more accurate, complete, and easy to work with; and
- Connect data providers and users for ongoing collaboration

The Center develops the agenda for each Roundtable in collaboration with the agency. Preparation includes research and questionnaires to stakeholders inside and outside government. Each Roundtable combines presentations from agency officials and staff with breakout sessions that bring government and data customers together in groups of eight to ten.

After each Roundtable, the Center for Open Data Enterprise issues a public report summarizing the participants' discussion of data challenges and opportunities, their proposals for solutions, and agencies' commitments to action. The Open Data Roundtables were originally launched as an initiative of the GovLab at NYU and are now run by the Center for Open Data Enterprise.

## PREVIOUS ROUNDTABLES

U.S. Department of Commerce and White House Office of Science and Technology Policy

U.S. Department of Agriculture and the White House Climate Data Initiative

U.S. Patent and Trademark Office

U.S. Department of Transportation

U.S. Department of Energy

U.S. Department of Veterans Affairs

U.S. Department of the Treasury

## PARTICIPANTS

The Roundtable aims to bring together the agency's key data stakeholders, including representatives from companies, nonprofits, academic institutions, other government agencies using their data. Representatives include decision-makers with technical and/or business understanding of how the organizations use government data.

# APPENDIX F: ABOUT THE CENTER FOR OPEN DATA ENTERPRISE

The **Center for Open Data Enterprise** is an independent nonprofit organization that develops smarter open data strategies for governments, businesses, and nonprofits by focusing on data users. Our mission is to maximize the value of open data as a public resource. We work to unleash this value through a better understanding of open data users and greater engagement of stakeholders.

## What We Do

Our user-centered approach aims to improving the open data ecosystem in three ways. We **map** the uses of open data from around the world; **convene** data users and providers to identify challenges and opportunities; and **implement** solutions driven by user input.

### We MAP.



The first global view of the uses of open data.

The Open Data Impact Map is a searchable, sortable database of the uses of open data, providing a deeper understanding of the demand for this resource.

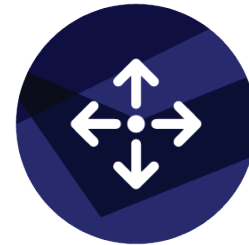
### We CONVENE.



Action-oriented dialogues for government agencies & their data users.

Our Open Data Roundtables in the U.S. and abroad help identify high-value datasets, find solutions to data problems, and establish new collaborations.

### We IMPLEMENT.



Improving the management & quality of open data.

We work with public and private sector partners to develop solutions to key data challenges informed by user feedback.

## Contact Us

For general inquiries, contact Katherine Garcia at [katherine@odenterprise.org](mailto:katherine@odenterprise.org).

For partnership opportunities, contact Laura Manley at [laura@odenterprise.org](mailto:laura@odenterprise.org).

Learn more at [opendataenterprise.org](http://opendataenterprise.org).