Mapping Employment Projections and O*NET data: a methodological overview

Because of differences in data collection purposes and practices, combining data from different federal statistical programs that use the Standard Occupational Classification system can be complicated. This article addresses this problem by presenting a method for mapping occupational data from the U.S. Bureau of Labor Statistics Employment Projections program and the U.S. Department of Labor Occupational Information Network.

The EP program develops 10-year projections of future labor market outcomes, including employment and number of job openings, for around 800 occupations.[1] The program does not collect its own data from a survey or a census, as is common in other BLS programs. Instead, it uses data from other sources as inputs to a multistep process that produces the projections.[2] The key data source for current occupational employment estimates, as well as staffing patterns (which describe the occupational composition of each industry), is the BLS Occupational Employment and Wage Statistics (OEWS) program.[3] Because OEWS data are integral to the formation of occupational projections, EP and OEWS (hereafter referred to collectively as "EP/OEWS") share identical occupational taxonomies. Therefore, any method for mapping the EP and O*NET taxonomies applies equally well to mapping the OEWS and O*NET taxonomies.

O*NET, which is sponsored by the DOL Employment and Training Administration, collects, analyzes, and disseminates information on various occupational characteristics. Examples of variables found in O*NET data include the primary tasks performed in an occupation or the knowledge, skills, and abilities required in that occupation.[4]

It may not be obvious why a discussion about mapping data from these sources is necessary, because both EP/OEWS and O*NET provide crosswalks from their occupational taxonomies to the SOC.[5] These crosswalks can be merged to connect occupational information from the EP/OEWS and O*NET programs. However, the existing crosswalks have no guidance for users on how to combine or impute data in the absence of a one-to-one match between occupations. Because of differences in the purpose and methods of data collection across programs—

Amy Hopson
hopson.amy@bls.gov

Amy Hopson is a research economist in the Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics.

differences that result in a different set of occupations included in each program—mapping EP/OEWS and O*NET data requires resolving various types of nonmatches.[6]

There is not necessarily one "right" way to do this mapping, but the method outlined in this article is generalizable across applications and results in every EP/OEWS occupation being mapped to O*NET data. Alternative approaches are also possible and can be tailored to the specific topic of research or the level of occupational specificity desired in the analysis. For example, in a 2019 working paper, Matthew Dey and Mark A. Loewenstein impute missing O*NET data for an OEWS occupation by using the O*NET characteristics of the most similar occupations in terms of wages (the authors' primary variable of interest).[7] Another example in which O*NET occupational characteristics are initially matched to OEWS occupations appears in a 2011 publication by Daron Acemoglu and David Autor, who summarize results for occupations aggregated into either 4 or 10 large groups.[8] Dey and Loewenstein's method works well in the context of a particular research question, whereas Acemoglu and Autor's approach minimizes the effects of classification discrepancies by requiring less occupational detail.

While these approaches are valid, the purpose of the method proposed here is to create a general mapping that is not tied to any particular application. The method involves understanding the conceptual framework behind different taxonomies, exploiting the hierarchical structure of the SOC, and, whenever possible, using employment data to weight more detailed O*NET data.

The article proceeds as follows. First, it establishes some historical context by reviewing the development of the relevant occupational taxonomies. Second, it defines the types of nonmatches that arise because of inherent methodological differences across programs. Researchers who want to combine the rich data available from the EP/OEWS and O*NET programs will need to decide how to manage these discrepancies. Third, the article offers a method of mapping the current EP/OEWS occupational taxonomy to the O*NET-SOC 2019 taxonomy.[9] Although the focus is on current classifications, the order and principles behind the mapping can be applied whenever a new taxonomy is implemented. Finally, the article concludes with a simple example that illustrates one possible use of the completed mapping.

History of occupational taxonomies

This section provides a brief history of the development of the SOC, OEWS, and O*NET occupational taxonomies. The timeline of that development is summarized in table 1.

Table 1. Timeline of taxonomies, 1998–2022

Year	O*NET database	O*NET taxonomy	OEWS data	OEWS taxonomy	SOC taxonomy
1998	O*NET 98	O*NET 98	OEWS 1997	OEWS 97-98	1998 SOC
1999	O*NET 98	O*NET 98	OEWS 1998	OEWS 97-98	1998 SOC
2000	O*NET 3.0	O*NET-SOC 2000	OEWS 1999	2000 SOC[1]	2000 SOC
2001	O*NET 3.1	O*NET-SOC 2000	OEWS 2000	2000 SOC[1]	2000 SOC
2002	O*NET 4.0	O*NET-SOC 2000	OEWS 2001	2000 SOC[1]	2000 SOC
2003	O*NET 5.0, 5.1	O*NET-SOC 2000	OEWS 2002	2000 SOC[1]	2000 SOC
2004	O*NET 6.0, 7.0	O*NET-SOC 2000	OEWS May 2003, OEWS November 2003	2000 SOC[1]	2000 SOC

See footnotes at end of table.

Table 1. Timeline of taxonomies, 1998–2022

Year	O*NET database	O*NET taxonomy	OEWS data	OEWS taxonomy	SOC taxonomy
2005	O*NET 8.0, 9.0	O*NET-SOC 2000	OEWS May 2004, OEWS November 2004	2000 SOC	2000 SOC
2006	O*NET 10.0, 11.0	O*NET-SOC 2006	OEWS May 2005	2000 SOC	2000 SOC
2007	O*NET 12.0	O*NET-SOC 2006	OEWS May 2006	2000 SOC	2000 SOC
2008	O*NET 13.0	O*NET-SOC 2006	OEWS May 2007	2000 SOC	2000 SOC
2009	O*NET 14.0	O*NET-SOC 2009	OEWS May 2008	2000 SOC	2010 SOC
2010	O*NET 15.0	O*NET-SOC 2009	OEWS May 2009	2000 SOC	2010 SOC
2011	O*NET 15.1, 16.0	O*NET-SOC 2010	OEWS May 2010	Hybrid 2000/2010 SOC	2010 SOC
2012	O*NET 17.0	O*NET-SOC 2010	OEWS May 2011, OEWS November 2011	Hybrid 2000/2010 SOC	2010 SOC
2013	O*NET 18.0	O*NET-SOC 2010	OEWS May 2012	2010 SOC[2]	2010 SOC
2014	O*NET 18.1, O*NET 19.0	O*NET-SOC 2010	OEWS May 2013	2010 SOC[2]	2010 SOC
2015	O*NET 20.0, O*NET 20.1	O*NET-SOC 2010	OEWS May 2014	2010 SOC[2]	2010 SOC
2016	O*NET 20.2, 20.3, 21.0, 21.1	O*NET-SOC 2010	OEWS May 2015	2010 SOC[2]	2010 SOC
2017	O*NET 21.2, 21.3, 22.0, 22.1	O*NET-SOC 2010	OEWS May 2016	2010 SOC[2]	2010 SOC
2018	O*NET 22.2, 22.3, 23.0, 23.1	O*NET-SOC 2010	OEWS May 2017	2010 SOC[2][3]	2018 SOC
2019	O*NET 23.2, 23.3, 24.0, 24.1	O*NET-SOC 2010	OEWS May 2018	2010 SOC[2][3]	2018 SOC
2020	O*NET 24.2, 24.3, 25.0	O*NET-SOC 2010	OEWS May 2019	Hybrid 2010/2018 SOC	2018 SOC
2020	O*NET 25.1	O*NET-SOC 2019	OEWS May 2019	Hybrid 2010/2018 SOC	2018 SOC
2021	O*NET 25.2, 25.3, 26.0[4], 26.1[4]	O*NET-SOC 2019	OEWS May 2020	Hybrid 2010/2018 SOC	2018 SOC
2022	O*NET 26.2[4], 26.3[4], 27.0[4], 27.1[4]	O*NET-SOC 2019	OEWS May 2021	2018 SOC	2018 SOC

[1] Residual SOC occupations not included in OEWS data; small additional set of occupations not included in OEWS data.

[2] OEWS published data for substitute teachers separately.

[3] OEWS aggregated 21 SOC occupations into 10 OEWS occupations.

[4] These database names and dates are expected if O*NET continues its recent publication schedule. O*NET does not publish its exact publication schedule for database updates.

Note: O*NET = Occupational Information Network; OEWS = Occupational Employment and Wage Statistics; SOC = Standard Occupational Classification.

Source: U.S. Bureau of Labor Statistics and O*NET.

SOC

The years 1998–2000 marked a turning point in the classification of occupations in the United States. Until then, different government agencies used different occupational taxonomies, each developed for agency-specific purposes. Although a first attempt at a more unified system was made with the introduction of the 1977 SOC, it never gained much traction.[10] The 2000 SOC resulted from years of work and discussions involving government agencies, experts, and the public, reflecting efforts to create a standardized occupational classification system.[11]

The SOC has been updated twice since 2000, and these updates resulted in the 2010 SOC and the 2018 SOC. The purpose of periodically updating the SOC is to capture a more current snapshot of the occupational landscape. As the economy fluctuates and technology changes, new occupations arise, and some existing ones become obsolete. The SOC captures these shifts by introducing new detailed occupations into its taxonomy, splitting a single occupation into two or more new occupations, combining two or more occupations into a single occupation, or removing a detailed occupation by moving it into a residual occupation.[12]

OEWS

The OEWS program began in 1971 as a survey of manufacturing establishments. Two years later, the program was expanded to include nonmanufacturing firms. For the next 25 years, OEWS used its own occupational classification system, updating it as needed with information on occupational composition gathered by its surveys.[13] Because of its size, scope, and relatively frequent updates, the OEWS occupational taxonomy served as a starting point for both the newly developing SOC and O*NET classification systems.

After a new SOC was finalized in 1998, OEWS adopted a SOC-based taxonomy in its 1999 and subsequent datasets.[14] A crosswalk between the 1998 OEWS occupational codes and the 1999 SOC occupational codes provides a mapping between the two systems.[15] The crosswalk represents a complex, many-to-many mapping, but the 1999 OEWS data contain an indicator for whether an occupation is comparable across the 1998 and 1999 coding systems.[16]

In May 2017, OEWS made some small adjustments to its occupational classification system by combining 21 detailed SOC occupations into 10 occupational aggregates.[17] Each aggregate comprises similar SOC occupations that cannot be reliably distinguished by the survey questions and responses. The use of these aggregates is expected to continue.

O*NET

O*NET was created to replace and enhance the occupational information contained in the *Dictionary of Occupational Titles* (*DOT*).[18] One of the biggest tasks in the initial development of O*NET was to consolidate around 12,000 *DOT* occupations into an occupational taxonomy about a tenth of that size.[19] The starting point for the new O*NET taxonomy was the taxonomy used by OEWS at the time. The O*NET developers sometimes split a single OEWS occupation into two or more "occupational units" (OUs) that were more similar to one another than to the broader occupation.[20]

After developing this new taxonomy in 1998, O*NET had to quickly adapt to the newly introduced 1998 SOC. The result was the O*NET-SOC 2000 taxonomy, which was developed to match its OUs to the SOC taxonomy. Many (482) OUs had a one-to-one match to a SOC occupation. In those cases, O*NET adopted the matching SOC codes and titles. Sometimes, the O*NET occupations were at a more detailed level than the SOC occupations. In this case, the OUs determined in O*NET's initial taxonomy were kept and fit into the SOC taxonomy by creating eight-digit occupational codes that matched the more detailed OUs with the corresponding SOC occupations. Finally, SOC occupations that were more detailed than a corresponding OU or that did not link to any OU were adopted into the O*NET-SOC 2000 taxonomy.

O*NET next updated its taxonomy in 2006, aiming to "identify any overlap, redundancy, or gaps in the way [O*NET-SOC occupations] represent the SOC occupations to which they are linked."[21] This update resulted in a better overall mapping of the O*NET and SOC taxonomies, reducing the number of eight-digit O*NET-SOC occupations and adding data collection for 70 SOC-level occupations. In 2009, O*NET again updated its taxonomy in order to incorporate 153 new and emerging occupations. To identify these occupations, O*NET focused on a set of 17 high-growth, high-demand industry clusters.[22] The next two updates to the O*NET taxonomy, O*NET-SOC 2010 and O*NET-SOC 2019, were made in response to SOC revisions.

Sources of nonmatches

OEWS collects data through a semiannual survey of nonfarm establishments. The survey's sample design is at the establishment level, and data are collected on wage and salary workers within a sampled establishment.[23] To obtain reliable estimates at the desired occupational level of detail, OEWS combines data from the six most recent survey panels. With a final combined sample of over 1,000,000 establishments, OEWS can estimate occupational employment and wages for every SOC occupation (a few exceptions are discussed below).

O*NET collects data differently. Instead of gathering information on all occupations within a sampled establishment, it targets specific occupations from a predetermined taxonomy.

Level of detail

One reason for nonmatches is that, for some occupations, O*NET collects data at a more detailed level than does the EP program. While the most detailed SOC occupation is classified at the six-digit level of detail, some O*NET-SOC occupations are classified at the eight-digit level. Because the SOC system covers all occupations, every eight-digit O*NET-SOC occupation is subsumed under a six-digit SOC occupation, and no eight-digit O*NET-SOC occupation captures, by itself, an entire SOC occupation.

A challenging feature of merging O*NET data with other occupational data is that the O*NET dataset contains descriptive occupational characteristics but does not contain estimates of employment. This means that the data for O*NET-SOC occupations cannot be easily aggregated to a higher level in the taxonomy. Without employment numbers to weight the estimates, one cannot capture the relative magnitudes of more detailed O*NET-SOC occupations when they are aggregated to the SOC level of detail.

A second reason for nonmatches between EP/OEWS and O*NET-SOC occupational data is that EP/OEWS has a small group of occupations that are classified at a less detailed level than those in the SOC. O*NET only collects data at the most detailed SOC level or beyond, so any occupations aggregated to a higher level do not have matching O*NET-SOC occupations. The less detailed EP/OEWS occupations are based on two types of combinations in the OEWS data: (1) permanent aggregations of two or three detailed occupations that OEWS could not reliably differentiate in its survey, and (2) temporary hybrid occupations that OEWS uses while it transitions to a new classification system.

Because of the panel nature of the survey design, OEWS must have a transitional period for adopting a new SOC taxonomy once one is introduced. The most recent version of the SOC—the 2018 SOC—was finalized in November 2017. Beginning with the November 2018 survey panel, OEWS has collected and coded occupational data by using this new taxonomy. Data for older panels were collected and coded by using the 2010 version of the SOC. As a result, OEWS estimates for May 2019 and May 2020 rely on data collected and coded under two

different taxonomies. For these 2 years (data for which were released in spring 2020 and spring 2021), OEWS created and published its estimates under a hybrid taxonomy that combines the 2010 and 2018 SOC structures.[24]

Residual occupations

The SOC system contains many residual occupations—usually identified by the phrase "all other" in their titles—that include any employment that belongs within a minor or broad group level in the SOC hierarchy but is not included in a distinct detailed occupation. The EP program produces estimates for all occupations, including the residual occupations. However, because O*NET targets specific occupations and does not collect information on the universe of all jobs, it has no data on residual occupations at the SOC level. As mentioned previously, O*NET collects data on some occupations at the eight-digit level of detail. These occupations are sometimes matched to a residual SOC occupation because they do not belong under a more specific SOC occupation. However, they can never account for an entire six-digit SOC occupation. A residual SOC occupation always includes many unnamed occupations that O*NET data have not captured.

Table 2 summarizes the types of mappings that exist between EP/OEWS and O*NET data. The table's top data row shows that 698 occupations have a one-to-one match in O*NET, and that this set of matches contains 87 percent of total EP employment. The temporary OEWS hybrid occupations are the biggest source of nonmatches, containing 8.5 percent of total employment. Some of these nonmatches will be resolved once the OEWS transition to the 2018 SOC is complete with the release of May 2021 data in spring 2022. A little less than half of the employment in this group is in aggregate occupations that include a residual SOC occupation. The residual occupations within these aggregates will still require a mapping procedure after the 2018 SOC is fully implemented. The rest of the nonmatches are split equally between the permanent OEWS aggregate occupations and the remaining residual occupations not accounted for elsewhere.[25]

Table 2. Types of mappings between EP/OEWS and O*NET data

Category	EP/OEWS occupations	O*NET-SOC occupations	Percent of EP employment[1]
Match	698	698	87.4
Extra O*NET	52	84	—
Permanent OEWS aggregate	8	17	2.0
OEWS aggregate with residual	3	6	1.2
Temporary OEWS hybrid	34	90	8.5
OEWS hybrid with residual	19	59	3.9
Remaining residual	50	34	2.0
Total	790	923	100.0

[1] Estimates in this column may not add to 100 because of rounding.

Note: EP = Employment Projections; OEWS = Occupational Employment and Wage Statistics; O*NET = Occupational Information Network; SOC = Standard Occupational Classification.

Source: U.S. Bureau of Labor Statistics, 2019 EP data; and O*NET.

Mapping procedure

The mapping procedure is carried out in several steps that must be completed sequentially. Each step deals with a certain type of nonmatch, and later steps build on results from earlier steps. The steps are as follows:

- Step 1: Eliminate extra O*NET occupations
- Step 2: Transform EP/OEWS taxonomy into SOC taxonomy
    - Step 2.A: Permanent OEWS aggregates
    - Step 2.B: Temporary OEWS hybrids
- Step 3: Fill in residual SOC occupations
    - Step 3.A: Residual occupations at the most detailed level
    - Step 3.B: Residual occupations at higher levels

Step 1: Eliminate extra O*NET occupations

For a set of occupations, O*NET collects data both at the six-digit SOC level and at the more detailed eight-digit level. For example, O*NET collects data on occupations O*NET-SOC 11-1011.00, chief executives, and O*NET-SOC 11-1011.03, chief sustainability officers. The former occupation is at the six-digit level of detail and maps exactly to SOC 11-1011, chief executives. Chief sustainability officers, on the other hand, are a more detailed subset of the occupation and have no corresponding SOC code. Because a one-to-one match at the correct level of detail already exists, the extra eight-digit O*NET occupation should be thrown out.

The general mapping procedure for this step involves matching the six-digit O*NET-SOC occupations to the corresponding EP/OEWS occupations and then discarding the more detailed eight-digit O*NET-SOC occupations.

Step 2: Transform EP/OEWS taxonomy into SOC taxonomy

The 2019 EP/OEWS taxonomy contains two types of occupational aggregates that are at a less detailed level than the corresponding occupations in the 2018 SOC. The first type of aggregate stems from a permanent change in OEWS procedures. The second type of aggregate contains temporary occupational combinations that will be used only while OEWS transitions from the 2010 SOC to the 2018 SOC. These aggregates are cases in which the 2018 SOC occupations are more detailed than the occupations in the 2010 SOC. This situation can occur when a single detailed occupation is split into two or more detailed occupations, or when a residual occupation is split into one or more new occupations and a new residual occupation.[26]

Step 2.A: Permanent OEWS aggregates

For permanent aggregates, the procedure uses the most recently available employment data for the component occupations—that is, May 2016 OEWS estimates—to determine their relative sizes within each aggregate.[27] Because the SOC structure is comprehensive, the combination of detailed occupations should span the entirety of the new, higher level occupation. The proportions based on relative occupational employment are used as weights on individual components.[28] Because OEWS was still using the 2010 SOC in 2016, this initial calculation maps the 2016 EP/OEWS taxonomy to the 2010 SOC taxonomy. Columns three through six of table 3 represent the process described up to this point.

Table 3. Step 2.A, OEWS aggregate occupations and weights

2019 EP/ OEWS code	EP/OEWS title	2018 EP/ OEWS code	2010 SOC code	2010 SOC title	2010 weight	2018 SOC code	2018 SOC title	2018 weight	Final step-2.A weight
13-1020	Buyers and purchasing agents	13-1020	13-1021	Buyers and purchasing agents, farm products	0.03	13-1021	Buyers and purchasing agents, farm products	1.00	0.03
			13-1022	Wholesale and retail buyers, except farm products	0.26	13-1022	Wholesale and retail buyers, except farm products	1.00	0.26
			13-1023	Purchasing agents, except wholesale, retail, and farm products	0.71	13-1023	Purchasing agents, except wholesale, retail, and farm products	1.00	0.71
21-1018	Substance abuse, behavioral disorder, and mental health counselors	21-1018	21-1011	Substance abuse and behavioral disorder counselors	0.39	21-1011	Substance abuse and behavioral disorder counselors	1.00	0.39
			21-1014	Mental health counselors	0.61	21-1014	Mental health counselors	1.00	0.61
29-2010	Clinical laboratory technologists and technicians	29-2010	29-2011	Medical and clinical laboratory technologists	0.51	29-2011	Medical and clinical laboratory technologists	1.00	0.51
			29-2012	Medical and clinical laboratory technicians	0.49	29-2012	Medical and clinical laboratory technicians	1.00	0.49
39-1013	First-line supervisors of gambling services workers	39-1010	39-1011	Gaming supervisors	0.74	39-1013	First-line supervisors of gambling services workers	—	1.00
			39-1012	Slot supervisors	0.26				
39-7010	Tour and travel guides	39-7010	39-7011	Tour guides and escorts	0.93	39-7011	Tour guides and escorts	1.00	0.93
			39-7012	Travel guides	0.07	39-7012	Travel guides	1.00	0.07
47-4090	Miscellaneous construction and related workers	47-4090	47-4091	Segmental pavers	0.05	47-4091	Segmental pavers	1.00	0.05
			47-4099	Construction and related workers, all other	0.95	47-4099	Construction and related workers, all other	1.00	0.95
51-2028	Electrical, electronic, and electromechanical assemblers, except coil winders, tapers, and finishers	51-2028	51-2022	Electrical and electronic equipment assemblers	0.83	51-2022	Electrical and electronic equipment assemblers	1.00	0.83
			51-2023	Electromechanical equipment assemblers	0.17	51-2023	Electromechanical equipment assemblers	1.00	0.17
51-2090	Miscellaneous assemblers and fabricators	51-2098	51-2092	Team assemblers	0.83	51-2092	Team assemblers	1.00	0.83
			51-2099	Assemblers and fabricators, all other	0.17	51-2099	Assemblers and fabricators, all other	1.00	0.17
53-1047	First-line supervisors of transportation and material moving workers, except aircraft cargo handling supervisors	53-1048	53-1021	First-line supervisors of helpers, laborers, and material movers, hand	0.48	53-1042	First-line supervisors of helpers, laborers, and material movers, hand	1.00	0.48
			53-1031	First-line supervisors of transportation and material-moving machine and vehicle operators	0.52	53-1043	First-line supervisors of material-moving machine and vehicle operators	0.33	0.17

See footnotes at end of table.

Table 3. Step 2.A, OEWS aggregate occupations and weights

2019 EP/ OEWS code	EP/OEWS title	2018 EP/ OEWS code	2010 SOC code	2010 SOC title	2010 weight	2018 SOC code	2018 SOC title	2018 weight	Final step-2.A weight
						53-1044	First-line supervisors of passenger attendants	0.33	0.17
						53-1049	First-line supervisors of transportation workers, all other	0.33	0.17

Note: OEWS = Occupational Employment and Wage Statistics; EP = Employment Projections; SOC = Standard Occupational Classification.

Source: U.S. Bureau of Labor Statistics, May 2016 OEWS data.

The next task is to transform the older classifications into the 2019 EP/OEWS and 2018 SOC taxonomies. If there is only a code or title change for one or more relevant occupations, the mapping remains the same; it changes only if the composition of the crosswalked occupations changes.

Most of the occupations and their mappings carry over exactly in the updated taxonomies, with two exceptions. In one case, which involves first-line supervisors of gambling services and workers, the 2018 SOC combines previously separated occupations into an aggregate OEWS occupation. This means that relative weights are no longer necessary, and there is now a one-to-one mapping. In another case, one of the detailed occupations in the 2010 SOC is further split into three new occupations in the 2018 SOC. As a result, a single EP/OEWS occupation that previously mapped to two SOC occupations now maps to four SOC occupations. In this situation, one should start with the proportions calculated for the two 2010 SOC occupations and then use the strategy discussed in step 2.B to further break down the single SOC occupation that is split into three new occupations.

The weights associated with the transformation to the most recent taxonomies are shown in the second-to-last column of table 3. To complete the mapping for this step, one should multiply these weights by the 2016 employment weights.

Step 2.B: Temporary OEWS hybrids

As noted earlier, the second type of aggregate is made up of new occupations whose relative sizes are not captured in existing data. For this reason, each detailed occupation making up the temporary hybrid aggregate receives equal weight. Once OEWS collects data for the new occupations in all six panels required for its estimates, this second type of aggregate will no longer be necessary.

Overall, the general mapping procedure for step 2 involves using the most recently available OEWS employment data as weights; if no such data are available, the procedure uses equal weights.

Step 3: Fill in residual SOC occupations

Because O*NET targets specific occupations for data collection, it does not have data on any residual SOC occupations. Although O*NET collects data on some specific occupations that fit within a residual SOC code in the occupational taxonomy, these more detailed occupations have no corresponding OEWS employment data.

Therefore, no information exists on what percentage of a residual SOC occupation is covered by the detailed O*NET occupation(s) matched to it.

Beyond this lack of information, using the O*NET data poses a conceptual problem. All O*NET occupations originally matched to residual SOC occupations are "new and emerging" occupations as determined by O*NET. Because the occupations were selected on the basis of particular characteristics, they are not necessarily representative of the residual occupation as a whole.

Given these coverage and conceptual issues, the first task in this step is to drop any O*NET-SOC occupations belonging to a residual SOC occupation. These residual occupations are now essentially missing data, and the missing matches are imputed by using a weighted average of related occupations. This method exploits the hierarchical SOC structure by assuming that occupations placed in a residual category are similar to the other occupations grouped under the next-highest level in the hierarchy. This imputation process starts by splitting the residual occupations into two groups, A and B, depending on their position in the SOC hierarchy.

Group A includes residual occupations whose SOC code does not end in 99. These occupations are at the most detailed level and capture occupations not otherwise classified in a particular broad group within the SOC taxonomy. An example of a residual occupation in this category is SOC 21-1019, counselors, all other. This occupation captures any occupations that belong in SOC broad group 21-1010, counselors, but that do not belong in any of the five other detailed SOC occupations within that group.

Group B includes residual occupations whose SOC code ends in 99. Each occupation in this set captures occupations not otherwise classified in a particular minor group within the SOC taxonomy. An example of a residual occupation in this category is SOC 21-1099, community and social service specialists, all other. This occupation captures any occupations that belong in SOC minor group 21-1000, counselors, social workers, and other community and social service specialists, but that do not belong in any of the other detailed SOC occupations within that group (including SOC 21-1019, counselors, all other).

Step 3.A: Residual occupations at the most detailed level

Using occupations in group A, this step involves creating a weighted average of the other SOC occupations within the broad group. The calculation uses the current EP baseline employment estimates as weights.[29] The weights are then applied to the O*NET-SOC codes that correspond to the SOC codes in the broad group.

Tables 4 and 5 illustrate this procedure for SOC 21-1019, counselors, all other. Table 4 presents a snapshot of the location of this residual occupation within the SOC structure, and table 5 works through the calculation of weights for the corresponding O*NET-SOC codes.

Table 4. 2018 SOC structure for broad group 21-1010, counselors

2018 SOC detailed occupation	2018 SOC title	2019 EP/OEWS code	2019 EP/OEWS title
21-1011	Substance abuse and behavioral disorder counselors	21-1018	Substance abuse, behavioral disorder, and mental health counselors
21-1012	Educational, guidance, and career counselors and advisors	21-1012	Educational, guidance, and career counselors and advisors

See footnotes at end of table.

Table 4. 2018 SOC structure for broad group 21-1010, counselors

2018 SOC detailed occupation	2018 SOC title	2019 EP/OEWS code	2019 EP/OEWS title
21-1013	Marriage and family therapists	21-1013	Marriage and family therapists
21-1014	Mental health counselors	21-1018	Substance abuse, behavioral disorder, and mental health counselors
21-1015	Rehabilitation counselors	21-1015	Rehabilitation counselors
21-1019	Counselors, all other	21-1019	Counselors, all other

Note: SOC = Standard Occupational Classification; EP = Employment Projections; OEWS = Occupational Employment and Wage Statistics.

Source: U.S. Bureau of Labor Statistics.

Table 5. Calculating mapping codes and weights for SOC 21-1019, counselors, all other

EP/OEWS code	SOC code	2019 projected employment (thousands)	OEWS weights (from step 2)	2019 SOC employment (thousands)[1]	SOC weights[1]	O*NET-SOC code
21-1018	21-1011	319.4	0.39	126.0	0.15	21-1011.00
21-1012	21-1012	333.5	1.00	333.5	0.40	21-1012.00
21-1013	21-1013	66.2	1.00	66.2	0.08	21-1013.00
21-1018	21-1014	319.4	0.61	193.4	0.23	21-1014.00
21-1015	21-1015	120.2	1.00	120.2	0.14	21-1015.00
Total	—	—	—	841.3	1.00	—

[1] Estimates may not sum to total because of rounding.

Note: SOC = Standard Occupational Classification; EP = Employment Projections; OEWS = Occupational Employment and Wage Statistics; O*NET = Occupational Information Network.

Source: U.S. Bureau of Labor Statistics, 2019 EP and May 2016 OEWS data; and O*NET.

As seen in table 4, broad group 21-1010 contains the aggregate EP/OEWS occupation 21-1018, substance abuse, behavioral disorder, and mental health counselors. The fourth column of table 5 shows how employment in EP/OEWS occupation 21-1018 is disaggregated by using the weights calculated in step 2. SOC 21-1011, substance abuse and behavioral disorder counselors, has a weight of 0.39, and SOC 21-1014, mental health counselors, has a weight of 0.61. Applying these weights to the current employment estimates yields employment estimates for each SOC occupation.

Once all occupations within the broad group have employment estimates at the SOC level, the difference between the broad group's total employment and the employment in the residual occupation serves as the base for determining the relative weight of each occupation. After the SOC proportions are calculated, the O*NET-SOC code or codes are mapped to each SOC code, taking on the corresponding SOC weight.

Step 3.B: Residual occupations at higher levels

In this step, the process described for step 3.A is repeated at the minor-group level for occupations in group B. Occupations in group A must be mapped prior to this step. These occupations refer to a more detailed level within the SOC hierarchy and thus may be included within the relevant minor group for a given group-B occupation.

Overall, the general mapping procedure for step 3 involves deleting any existing O*NET codes matched to a residual SOC occupation and then filling in the residual SOC mapping with the employment-weighted average of the mappings of other occupations within the corresponding broad or minor group.
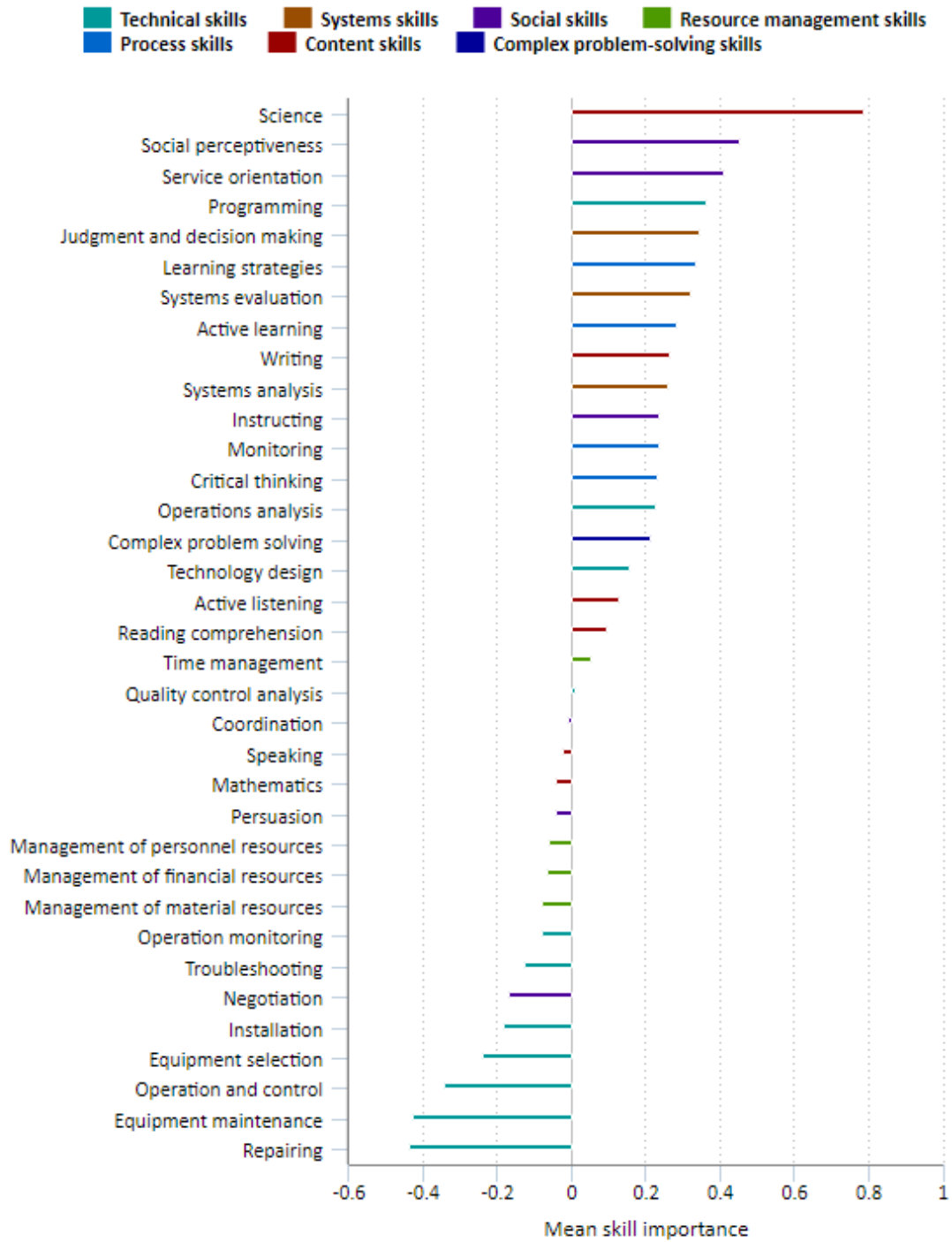
Use-case example

This section presents an example illustrating a potential application of the mapping procedure. This example is not meant to present a detailed analysis of the data but rather to show how the mapping procedure might be used to analyze interesting questions. Suppose we want to know how a changing mix of occupations may affect demand for skills. One way to explore this question is to merge EP data on projected occupational employment change with O*NET data on occupational skill requirements. Specifically, the present example shows the average importance of each skill measured by O*NET for the 50 occupations projected to have the fastest growth according to 2019 EP data.

O*NET data capture 35 skill elements spread across 7 aggregated skill categories. Respondents to the O*NET survey answer the following question: "How important is the skill to the performance of your current job?" Each skill is rated on a scale of 1 (not important) to 5 (extremely important). Because the average importance of the skills varies widely (i.e., some skills are important across most occupations or not important across most occupations), each O*NET measure is normalized to have a mean of 0 and a standard deviation of 1. This normalization makes it easier to place the value of a particular occupation's skill within the overall distribution of that skill across all occupations.

The skill data in this example come from the O*NET 25.0 database. This version of the O*NET database relies on an older taxonomy (O*NET-SOC 2010), but it is used here because skills data for many of the new occupations in the most recent version are not yet available.[30] The steps of the mapping procedure are almost identical to those outlined in the previous section, but one additional step is needed. This extra step addresses a type of occupational nonmatch that is not present in the most recent taxonomies, and this nonmatch is discussed in the appendix.

Chart 1 shows the average skill importance, weighted by employment, for the 50 fastest growing occupations. These occupations require science skills that are three-quarters of a standard deviation above the mean, on average. Social perceptiveness and service orientation skills are also more important in the 50 fastest growing occupations than they are in the average occupation. On the other hand, most elements in the technical skills category have below-average importance for this group.

Chart 1. Average skill importance for top 50 occupations with fastest growth

Legend:
- Technical skills
- Process skills
- Systems skills
- Content skills
- Social skills
- Complex problem-solving skills
- Resource management skills

Click legend items to change data display. Hover over chart to view data.
Source: Author's calculations based on U.S. Bureau of Labor Statistics data and O*NET data.

Conclusion

The method outlined in this article uses current occupational taxonomies to map O*NET data on occupational characteristics to every EP/OEWS occupation. The method involves understanding the conceptual framework behind different taxonomies, exploiting the hierarchical structure of the SOC, and, whenever possible, using employment data to weight more detailed O*NET data. Besides providing a step-by-step mapping procedure, this article identifies various types of nonmatch between the EP/OEWS and O*NET data sources and the reasons why their taxonomies differ in some areas. Researchers who want to combine data from these sources will benefit from accounting for these nonmatches, even if they use different strategies to better address a particular research question.

Appendix

One type of nonmatch, falling under the broad category of "level of detail," has only been resolved with the most recent update of the O*NET taxonomy. This type of nonmatch refers to the case of a six-digit SOC occupation being matched to O*NET-SOC occupations only at the eight-digit level of detail. This case is similar to the type of nonmatch described in step 1 of the mapping procedure, but the latter scenario has O*NET-SOC data at both the six-digit SOC level and the eight-digit O*NET-SOC level.

This appendix presents an additional methodological step for resolving this type of occupational nonmatch. Within the broader methodological order outlined in the main body of the article, this step falls between step 2 (transform EP/OEWS taxonomy into SOC taxonomy) and step 3 (fill in residual SOC occupations), and can be called step 2.1 (transform O*NET-SOC taxonomy into SOC taxonomy).

The main idea underlying the approach for this additional step is that eight-digit O*NET-SOC occupations may be represented in the pre-SOC OEWS data from 1998. For O*NET-SOC occupations with that feature, the employment estimates from the 1998 OEWS may be used to determine the relative weights within the broader SOC occupation. The approach uses the O*NET-SOC 2010 and 2010 SOC taxonomies, but its procedures can be applied to earlier versions of the O*NET-SOC taxonomy and to the 2000 SOC taxonomy. Because the approach uses older data and changing classification systems, it is not completely mechanical and may require some judgment.

The underlying idea that current eight-digit O*NET-SOC occupations may be present in pre-SOC OEWS data is based on the following facts, which were discussed in the section reviewing the history of occupational taxonomies:

- The original O*NET occupational units (OUs) were based on the pre-SOC OEWS taxonomy.
- When O*NET transitioned to a SOC-based classification system in the O*NET-SOC 2000 taxonomy, OUs at a more detailed level than occupations in the 2000 SOC were kept as eight-digit O*NET-SOC occupations. Some of these O*NET-SOC occupations remained in later O*NET-SOC taxonomies.
- It is possible to determine which 2010 SOC occupations are conceptually consistent across the relevant timeframe, from the 1998 OEWS occupational taxonomy to the 2010 SOC.
- OEWS allowed both one-to-one and many-to-one matches between its 1998 and 1999 taxonomies to be identified as consistent.

Taken together, these facts imply that the mapping strategy depends on the answers to two questions:

- Is the 2010 SOC occupation conceptually consistent all the way back to the pre-SOC OEWS taxonomy?
- Are the O*NET-SOC 2010 occupations equivalent to the pre-SOC OEWS occupations?

The SOC occupations are grouped according to the responses to these questions, and each occupational group has a different mapping procedure.

Group 1

In this group, the 2010 SOC occupations are comparable across the entire period, and the O*NET-SOC 2010 occupations match the corresponding 1998 OEWS occupations. This comparability across time is verified by including only occupations that OEWS indicated were consistent during the transition from the pre-SOC 1998 OEWS taxonomy to the SOC-based system. Consistency is also verified by checking whether the occupations were altered in the update from the 2000 SOC to the 2010 SOC. For this group, the procedure uses the proportions of 1998 employment as weights on the corresponding O*NET-SOC occupations. (See table A-1.)

Table A-1. Step 2.1, occupations and weights for group 1

2010 SOC title[1]	2010 SOC code	O*NET-SOC 2010 code	O*NET-SOC 2010 title	Weight
Appraisers and assessors of real estate (E)	13-2021	13-2021.01	Assessors	0.40
		13-2021.02	Appraisers, real estate	0.60
Marine engineers and naval architects (N)	17-2121	17-2121.01	Marine engineers	0.76
		17-2121.02	Marine architects	0.24
Police and sheriff's patrol officers (E)	33-3051	33-3051.01	Police patrol officers	0.83
		33-3051.03	Sheriffs and deputy sheriffs	0.17
Billing and posting clerks (T)	43-3021	43-3021.01	Statement clerks	0.05
		43-3021.02	Billing, cost, and rate clerks	0.95
Court, municipal, and license clerks (E)	43-4031	43-4031.01	Court clerks	0.49
		43-4031.02	Municipal clerks	0.27
		43-4031.03	License clerks	0.25
Stock clerks and order fillers (N)	43-5081	43-5081.01	Stock clerks, sales floor	0.55
		43-5081.02	Marking clerks	0.01
		43-5081.03	Stock clerks—stockroom, warehouse, or storage yard	0.34
		43-5081.04	Order fillers, wholesale and retail sales	0.10
Welders, cutters, solderers, and brazers (N)	51-4121	51-4121.06	Welders, cutters, and welder fitters	0.92
		51-4121.07	Solderers and brazers	0.08
Captains, mates, and pilots of water vessels (E)	53-5021	53-5021.01	Ship and boat captains	0.54
		53-5021.02	Mates—ship, boat, and barge	0.35
		53-5021.03	Pilots, ship	0.11

[1] Letters in parentheses indicate changes between 2000 and 2010 SOC taxonomies: N = no change, T = title change, E = definition editing change.

Note: SOC = Standard Occupational Classification; O*NET = Occupational Information Network.

Source: U.S. Bureau of Labor Statistics, 1998 Occupational Employment and Wage Statistics data; and O*NET.

Group 2

In this group, each 2010 SOC occupation has a part that can be traced back to the 1998 OEWS taxonomy, but the entire 2010 SOC occupation is not consistent over this timeframe. Once the consistent part of the 2010 SOC occupation has been identified in the 1998 OEWS taxonomy, a second qualification requires that at least one of

the consistent OEWS occupations from 1998 corresponds to one or more detailed O*NET-SOC occupations. The mapping procedure for this group involves the most judgment, because it may not be possible to verify exactly which occupational components are comparable over time. Once the continuous components have been identified, the next step of the mapping procedure is to calculate their employment in 1998 as a share of the total employment for the corresponding SOC occupation. Any O*NET-SOC occupations that match the component 1998 OEWS occupations are weighted according to the resulting shares. The remaining share of employment is then split equally between any unmatched O*NET-SOC occupations connected to the SOC occupation. (See table A-2.)

Table A-2. Step 2.1, occupations and weights for group 2

2010 SOC title	2010 SOC code	O*NET-SOC 2010 code	O*NET-SOC 2010 title	Weight
Musicians and singers	27-2042	27-2042.01	Singers	0.13
		27-2042.02	Musicians, instrumental	0.87
Detectives and criminal investigators	33-3021	33-3021.01	Police detectives	0.71
		33-3021.02	Police identification and records officers	0.00
		33-3021.03	Criminal investigators and special agents	0.29
		33-3021.05	Immigration and customs inspectors	0.00
		33-3021.06	Intelligence analysts	0.00
Jewelers and precious stone and metal workers	51-9071	51-9071.01	Jewelers	0.49
		51-9071.06	Gem and diamond workers	0.03
		51-9071.07	Precious metal workers	0.49

Note: SOC = Standard Occupational Classification; O*NET = Occupational Information Network.

Source: U.S. Bureau of Labor Statistics, 1998 Occupational Employment and Wage Statistics data; and O*NET.

For example, SOC occupation 51-9071, jewelers and precious stone and metal workers, is associated with three detailed O*NET-SOC occupations:

- 51-9071.01, jewelers
- 51-9071.06, gem and diamond workers
- 51-9071.07, precious metal workers

The mapping between the 1998 and 1999 OEWS occupations shows that SOC 51-9071 was matched with three 1998 OEWS occupations:

- 89123, jewelers and silversmiths
- 89126, precision hand workers, jewelry and related products
- 89926, gem and diamond workers

1998 OEWS occupation 89926, gem and diamond workers, is the only component of the SOC occupation that matches an O*NET-SOC code. In 1998, employment for gem and diamond workers was 1,100, and employment for all three 1998 OEWS occupations composing SOC 51-9071 was 36,710. Therefore, the O*NET-SOC occupation 51-9071.06, gem and diamond workers, receives a weight of 0.03 (1,100 divided by 36,710). The remaining employment is split equally between the other two O*NET-SOC occupations, each of which receives a weight of 0.49.

Group 3

In this group, no occupations are consistent across the 1998 and 1999 OEWS taxonomies and match a detailed O*NET-SOC occupation. Therefore, given *J* O*NET-SOC occupations mapped to a single SOC occupation, each O*NET-SOC occupation receives a weight of 1/*J*.

NOTES

1 For more information about the Employment Projections (EP) program, see https://www.bls.gov/emp/.

2 For information on all data sources and the methodology for producing projections estimates, see "Employment projections," *Handbook of Methods* (U.S. Bureau of Labor Statistics), https://www.bls.gov/opub/hom/emp/pdf/emp.pdf.

3 Note that, in spring 2021, this program changed its name from Occupational Employment Statistics (OES) to Occupational Employment and Wage Statistics (OEWS). This article uses the updated program name throughout, but references and URLs mostly use the old name. For more information about the OEWS program, see https://www.bls.gov/oes/.

4 For more information about O*NET, see "The O*NET® content model," *O*NET Resource Center* (Raleigh, NC: National Center for O*NET Development), https://www.onetcenter.org/content.html.

5 The crosswalk from OEWS 2019 to the 2018 SOC is available at https://www.bls.gov/oes/soc_2018.htm. The crosswalk from O*NET-SOC 2019 to the 2018 SOC is available at https://www.onetcenter.org/taxonomy/2019/soc.html.

6 In this article, the term "nonmatch" refers to any correspondence between occupations in two classification systems that does not represent a one-to-one mapping.

7 Matthew Dey and Mark A. Loewenstein, "On job requirements, skill, and wages," Working Paper 513 (U.S. Bureau of Labor Statistics, March 2019), https://www.bls.gov/osmr/research-papers/2019/pdf/ec190030.pdf.

8 Daron Acemoglu and David Autor, "Skills, tasks and technologies: implications for employment and earnings," in Orley Ashenfelter and David Card, eds., *Handbook of labor economics*, vol. 4, part B (Amsterdam: Elsevier, 2011), pp. 1043–1171.

9 Both of these taxonomies are based on the 2018 Standard Occupational Classification (SOC). For details on the EP/OEWS taxonomy, see https://www.bls.gov/oes/soc_2018.htm. For details on the O*NET-SOC 2019 taxonomy, see https://www.onetcenter.org/taxonomy.html.

10 Chester Levine, Laurie Salmon, and Daniel H. Weinberg, "Revising the Standard Occupational Classification system," *Monthly Labor Review*, May 1999, p. 36, https://www.bls.gov/mlr/1999/05/art4full.pdf.

11 The 2000 SOC was a minor revision of the 1998 SOC, affecting only a few occupations. See "What's new with the SOC? Changes to the SOC structure" (U.S. Bureau of Labor Statistics, last modified September 15, 2000), http://data.widcenter.org/download/soc2000/socnew.txt.

12 Because the SOC covers all occupations in the economy, the language of adding new occupations or deleting old ones can be confusing. Technically, a "new" detailed occupation is one that has been separated from a residual occupation. "Deleted" occupations are moved into the residual occupation rather than being deleted, so they are no longer separately specified.

13 *Occupational employment and wages, 1999*, Bulletin 2545 (U.S. Bureau of Labor Statistics, April 2001), https://www.bls.gov/oes/bulletin_1999.pdf.

14 From 1999 to 2003, OEWS collected and published data on around 700 of 801 detailed SOC occupations. The program also collected data—at the level of the broad group within the SOC taxonomy—for 8 additional occupations comprising 22 detailed occupations. Starting with its 2004 dataset, OEWS published data on every detailed occupation, mirroring the 2000 SOC.

15 The crosswalk is available at http://data.widcenter.org/download/soc1998/socoes98.xls.

16 According to a 1999 OEWS bulletin, "wage estimates for detailed occupations which changed under the SOC are based only on data collected in the 1999 survey, while wage estimates for detailed occupations which are unaffected by the SOC are based on data collected in the 1997, 1998, and 1999 surveys" (*Occupational employment and wages, 1999*, p. iii). The 1999 OEWS data contain an indicator for whether 1 or 3 years of wage data were used, so occupations with 3 years of wage data are comparable across the 1998 and 1999 OEWS taxonomies.

17 For a list of the detailed and aggregate occupations, see "Upcoming occupational and industry aggregations in the May 2017 Occupational Employment Statistics estimates," *Occupational Employment and Wage Statistics* (U.S. Bureau of Labor Statistics), https://www.bls.gov/oes/changes_2017.htm.

18 *The new DOT: a database of occupational titles for the twenty-first century* (U.S. Department of Labor, 1993), https://www.onetcenter.org/dl_files/omb2002/AppendixC.pdf.

19 "Appendix D—The development of the occupational information (O*NET™) analyst database" (Raleigh, NC: National Center for O*NET Development, 1998, revised June 12, 2002), https://www.onetcenter.org/dl_files/appendix_d.pdf.

20 Ibid.

21 "Updating the O*NET-SOC taxonomy: summary and implementation" (Raleigh, NC: National Center for O*NET Development, March 2006), p. 5, https://www.onetcenter.org/dl_files/UpdatingTaxonomy_Summary.pdf.

22 These industry clusters are advanced manufacturing, aerospace, automotive, biotechnology, construction, education, energy, financial services, geospatial technology, green economy, healthcare, homeland security, hospitality, information technology, nanotechnology, retail trade, and transportation. See "New and emerging occupations of the 21st century: updating the O*NET-SOC taxonomy" (Raleigh, NC: National Center for O*NET Development, March 2009), https://www.onetcenter.org/dl_files/UpdatingTaxonomy2009_Summary.pdf.

23 OEWS defines the term "establishment" as "the physical location of a certain economic activity, for example, a factory, mine, store, or office. Generally a single establishment produces a single good or provides a single service. An enterprise (a private firm, government, or nonprofit organization) could consist of a single establishment or multiple establishments. A multi-establishment enterprise could have all its establishments in one industry (i.e., a chain), or could have various establishments in different industries (i.e., a conglomerate)" (https://www.bls.gov/oes/oes_ques.htm#def).

24 The hybrid structure can be downloaded at https://www.bls.gov/oes/soc_2018.htm.

25 Also, the nonmatched occupations vary substantially in size. The 3 residual occupations that are part of OEWS aggregate occupations make up 1.2 percent of all employment, whereas the 50 remaining residual occupations account for only 2 percent of total employment.

26 For lists of these new detailed occupations, see tables 2 and 3 in *Standard Occupational Classification manual* (Executive Office of the President, Office of Management and Budget, 2018), pp. 6–8, https://www.bls.gov/soc/2018/soc_2018_manual.pdf.

27 Using the EP program's 2016 baseline employment estimates also is an option, but the data are less readily available. For this set of occupations, the difference in the weights calculated by using the EP and OEWS estimates never exceeds 2 percentage points.

28 The only exception is OEWS occupation 15-2090, miscellaneous mathematical science occupations, which is the OEWS aggregate of SOC 15-2091, mathematical technicians, and SOC 15-2099, mathematical science occupations, all other. According to O*NET, "mathematical technicians, as defined in the SOC, could not be found in sufficient numbers to support data collection"; see

"Updating the O*NET-SOC taxonomy," p. 15. Mathematical technicians are removed as a detailed occupation in the 2018 SOC, so the broad OEWS occupation is treated as a single residual occupation when creating the mapping.

29 This procedure requires the prior division of the OEWS occupations from step 2, because some of the aggregates include a residual and a nonresidual occupation.

30 O*NET usually publishes a significant update to approximately 100 occupations every summer, and since 2015, it has consistently released new occupational data every August (see https://www.onetcenter.org/db_releases.html).

RELATED CONTENT

Related Articles

Implementing the 2010 Standard Occupational Classification in the Occupational Employment Statistics, *Monthly Labor Review*, May 2013.

Revising the Standard Occupational Classification system for 2010, *Monthly Labor Review*, August 2010.

Revising the Standard Occupational Classification system, *Monthly Labor Review*, May 1999.

Related Subjects

Projections | Statistical programs and methods | Occupations