Mapping UK SOC2020 to O*NET

Rosie Day, Andy Dickerson, Gennaro Rossi and Rob Wilson

May 2023

1. Introduction

This note documents the process used to match UK SOC unit groups (or '4-digit' occupations) to O*NET occupations for the purposes of linking the occupation-specific skills measures in O*NET to occupations in the UK. It also provides a summary of the resulting one-to-many match between the two occupational classifications.

2. Matching SOC2020 and O*NET2019

There are 412 unit groups in UK SOC2020, the latest version of the UK's standard occupational classification. There are 1,016 occupations in O*NET-2019, which is the latest O*NET taxonomy (first implemented in O*NET in November 2020) and is itself linked to the US SOC2018 which is the current US SOC.

Matching between UK SOC2020 and O*NET-2019 utilised a Computer Assisted Structured Coding Tool (CASCOT), together with an expert (human) CASCOT coder. Initially, a manual matching exercise was carried out based on the names of the occupational groups in SOC and in O*NET. This was followed by examining the job titles in each occupational group using CASCOT. CASCOT uses the detailed job title index file for SOC2020 in order to produce a range of possible O*NET occupation matches for each UK SOC2020 4-digit occupation based on the similarities in the job title indexes (there are approximately 30,000 UK job titles and 55,000 US job titles). The final selection of the O*NET occupation code(s) chosen was then determined by the expert CASCOT coder.

There were 64 UK SOC2020 codes for which this SOC-to-O*NET process failed to identify a suitable O*NET code. For these, the reverse process was employed – i.e. mapping from O*NET-to-SOC in order to identify which O*NET code(s) best matched to these SOC codes. These were then used to fill in the gaps on the SOC-to-O*NET mapping.

This two-stage process therefore produced a one-to-many match for each of the 412 SOC2020 unit groups to one or more of the 1,016 O*NET-2019 occupations. However, not all of the 1,016 occupations in O*NET-2019 have data on skills – only 923 are 'datalevel' occupations. The (1016 – 923 =) 93 occupations for which data is not available (non-datalevel occupations) are typically small in employment terms, and so representative data on skills utilised in these occupations would be difficult to compile. Additionally, no data are collected for any military occupations in O*NET. Finally, there are also some occupations that are scheduled to be data-level occupations in O*NET-2019 but, as yet, no data have been collated and processed.

There are 22 UK SOC2020 unit groups which were matched to only non-datalevel occupations. For these, the CASCOT expert identified the next closest matches within O*NET which are datalevel occupations using a combination of CASCOT and examining the O*NET-related occupations. This was supplemented by a manual comparison of the descriptions and tasks of the SOC unit groups and O*NET occupations, well as checking the O*NET educational requirements against the skill level in SOC to help identify the most

relevant match/es. If no close datalevel matches could be found by comparing the job titles (for example, for SOC2020 n.e.c. (not elsewhere classified) unit groups), the O*NET matches already identified for the other SOC2020 codes in the 'parent' minor group (3-digit occupation) were considered.

This process enabled a match for all of the 412 SOC2020 unit groups to datalevel occupations in O*NET-2019. At present (May 2023), there are 10 UK SOC2020 occupations which match to O*NET codes which have yet to be populated with data for O*NET2019 (eg because they are new occupations to O*NET2019 as result of a disaggregation of a previous more aggregated occupation), but are scheduled to be datalevel in O*NET-2019.

The mapping process therefore yields a one-to-many mapping between the 412 unit group occupations in UK SOC2020 and O*NET-2019 occupations. This enables us to read across the different skills that are utilised in each occupation in order to generate 'occupational skills profiles' for UK occupations.

3. Summary of match between SOC2020 and O*NET-2019

The match has an average (median) of 2.81 (2) O*NET-2019 occupations for each UK SOC2020 unit group, with a minimum of 1 and a maximum of 38 O*NET-2019 occupations for each of the 412 unit groups in UK SOC2020. There are 168 1:1 mappings (41% of all UK SOC2020 occupations), while another 102 (25% of all UK SOC2020 occupations) unit groups match to only two O*NET-2019 occupations. The distribution of the number of O*NET codes per UK SOC2020 unit group is presented in the figure below.