

# Online Labour Index codebook \*

Otto Kässi, Vili Lehdonvirta

November 3, 2016

## 1 Background

This file describes the contents of the data files used for the visualisation here <http://ilabour.oii.ox.ac.uk/online-labour-index/>

The visualisation is based on two files: `OLIdata_YYYY_mm_dd.txt` and `bcountrydata_YYYY_mm_dd.txt` (e.g. `OLIdata_2016-11-02.txt`), and can be found here [https://figshare.com/articles/Online\\_Labour\\_Index\\_Measuring\\_the\\_Online\\_Gig\\_Economy\\_for\\_Policy\\_and\\_Research/3761562](https://figshare.com/articles/Online_Labour_Index_Measuring_the_Online_Gig_Economy_for_Policy_and_Research/3761562)

### OLIdata

The `OLIdata` file contains information on the daily new and filled vacancies across all platforms broken down by occupation and normalised to May 2016 = 100. It acts as a basis for the 'OLI' and 'By occupation' tabs of the visualisation. The file contains four columns:

1. *Date* - Date of collection. The data collection is scheduled to start every day at 00:01, and when the data from all platforms is collected, `OLIdata` is appended with new data.
2. *occupation* - Occupation classification. Total is a sum of all other occupations. For details, see paper
3. *status* - either "new" or "filled". If a vacancy was not present in data collected at  $t-1$ , but is present on  $t$ . It is a new project at time  $t$ . Conversely, if a vacancy was present at  $t-1$ , but is not present at time  $t$ , it is filled on time  $t$
4. *count* - count of vacancies with a given status/date/occupation combination. These are normalised so that the mean observed daily vacancies in May 2016 equals 100 index points.

---

\*This document describes the data files related to the paper Online Labour Index: Measuring the Online Gig Economy for Policy and Research.

## **Bcountrydata**

The bcountrydata is based on a weighted sample of of project from the platforms observed. The file contains four columns:

1. *timestamp* - same as 'Date'
2. *count* - count of vacancies from *country* within sample
3. *country* - self-explanatory
4. *country\_group* - since the employer country distribution is highly skewed in such a way that the top five countries add up to over 90 percent of the vacancies, we group the smaller countries into geographical groups for visualization purposes.
5. *occupation* - same as in OLIdata

*Notice that since the country by occupation distribution counts are calculated from a sample, the values in count field cannot be interpreted as absolute numbers of vacancies across countries. On the other hand, it is possible to use these numbers to calculate relative shares of different country and occupation combinations.*

The code used to create the visualisation is available from Github [https://github.com/ox-it/OxfordIDN\\_Case-Studies/tree/gh-pages/2016/OnlineLabourIndex](https://github.com/ox-it/OxfordIDN_Case-Studies/tree/gh-pages/2016/OnlineLabourIndex).