

Data description and codebook

The data consists of a random sample of 12,000 projects per year posted on the online platform during the period 2016-2020. The data includes the following project-level and worker-level variables:

Variable Name	Description
<u>project-level</u>	
project_ID	An anonymised unique identifier for each project.
date_published	The publication date of the project on the platform.
category_main	The broad job category of the project.
category_minor	The specific job category of the project.
hourly_rate	The hourly rate in USD paid for this project.
project_skills	A list of skill tags required to do the project (defined by the employer).
employer_country	The country location of the employer.
<u>worker-level</u>	
worker_ID	An anonymised unique identifier for each worker.
worker_country	The country location of the worker.
worker_skills	A list of skill tags of the worker (as specified on their individual profile page).
worker_gender	The gender of the worker (derived from the first name of the worker and the country location information).