

# README for the Reproducibility Package for Beyond Green Jobs: Advancing Metrics And Modelling Approaches For A Changing Labor Market

## Contents

1. Overview
2. Data Availability
3. Instructions for Replicators
4. List of Exhibits
5. Requirements

## Overview

This package reproduces Figures 1,2,4,6,7, and the text for B1. The other figures were taken from previous publications or otherwise publicly available.

In the top folder, you will find this README, the final manuscript, the `envir_bg.yml` file specifying the python packages required, plus `main.py`, the main script. It also includes the following folders:

Folder structure: - Analysis: Jupyter notebooks to generate results - Data: Original data sources from ESCO (in Data -> ESCO) and the Georgia LFS (in Data -> Georgia) - Data\_intermediate: Analysis output created for intermediate use during the main execution will be stored here (ESCO-specific files in Data\_intermediate -> ESCO) - Out: Final output folder (this will contain the final figures after execution)

## Data Availability

Some data cannot be made publicly available in this package. These are: 1) For Brazil – RAIS data (Relação Anual de Informações Sociais: Annual Report of Social Information) for 2020. This can be requested from IBGE (<https://ces.ibge.gov.br/base-de-dados/metadados/mte/relacao-anual-de-informacoes-sociais-rais>). The data used in this manuscript had been pre-cleaned and processed for the purpose of Figure 7. – Access year-month: 2023-03 – PNADC data for 2019 as cleaned by the World Bank GLD team (in particular `BRA_2019_PNADC_v01_M_v02_A_GLD_STATA14`). This can be requested from <https://microdatalib.worldbank.org/index.php/catalog/13106/get-microdata> and included in the Data/Brazil folder if available. – Access year-month: 2025-09 2) Bhutan labor force survey (LFS) data for 2022. This can be requested from <https://www.nsb.gov.bt/services/statistical-data-request/>. This can be included in the Data/Bhutan folder if available. – Access year-month: 2023-11

## Data Sources publicly available

All required data is available in the Data folder

- Bhutan data (restricted: see Data Availability) in Data/bhutan
- Bureau of Labor Statistics (BLS) data in Data/bls – The Occupational Employment and Wage Statistics (OEWS) can be downloaded from BLS: <https://www.bls.gov/oes/>. In particular, we used the national data for 2018 (`oesm18nat`, accessed 2025-09), the sectoral data for 2023 (`oesm23in4`, accessed 2025-09), and the sectoral data for 2011 (`oesm11in4`, accessed 2025-06), as well as additional agricultural data for 2011 (`ag_data_2011.xlsx`, accessed 2025-06) – We additionally used education data, which can be downloaded here: <https://www.bls.gov/emp/documentation/education-training-system.htm>, accessed 2024-11 – The ISCO – SOC harmonization table (`isco_soc_crosswalk.xls`) can be downloaded from here: , accessed 2025-06
- Brazil data (restricted: see Data Availability) in Data/brazil
- European Skills, Competences, Qualifications and Occupations (ESCO) data in Data/ESCO – were obtained via the European Union’s ESCO (European Skills, Competences, and Occupations) website: <https://esco.ec.europa.eu/en> – Version used: 1.2.0 – Language: English – Access year-month: 2024-06
- Green occupations data by Granata & Posadas in Data/granata\_posadas – Green skills per occupation; data from Granata, Julia & Posadas, Josefina. 2024. Why Look at Tasks When Designing Skills Policy for

the Green Transition? A Methodological Note on How to Identify Green Occupations and the Skills They Require. Available here: <https://openknowledge.worldbank.org/entities/publication/0fbe9c25-c37b-48a5-82d8-e2d4536fd494> – Access year-month: 2023-11

- International Standard Classification of Occupations (ISCO) data in Data/isco – ISCO occupational similarity scores from Background note 8 to the Argentina CCDR published in 2022 (Transition from Brown to Green Jobs: its Potential, Poverty and Distributional Impacts in Argentina), available here: <https://documents.worldbank.org/pt/publication/documents-reports/documentdetail/099250112052211003> – Access year-month: 2024-03
- Manual data created by the authors in Data/Manual – ISCO\_agriculture.xlsx: Authors’ manual classification of agricultural occupations – onet\_bls\_crosswalk.csv: Authors’ manual adjustment for the harmonization of ONET and BLS SOC data – soc\_missing\_isco\_manual.csv: Authors’ manual adjustment for harmonization of ISCO and BLS SOC data
- ONET data in Data/onet – Data on ‘green’ skills and occupations is available in ONET database version 24, available here: <https://www.onetcenter.org/dictionary/24.0/excel/> — Released: 2019 — Access year-month: 2025-04 – Data on the latest occupations and skills classification is available from ONET database version 30, available here: <https://www.onetcenter.org/dictionary/30.0/excel/> — Released: 2025 (latest before finalising paper) — Access year: 2025-08 — Data from database version 30 (as opposed to 24) can be recognised in Data/onet/work activities by their ‘underscore 30’ ending

## Statement about Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

## Instructions for Replicators

Ensure all required software and dependencies are installed as listed in the Requirements section.

The code runs without the restricted data – it will impute the final numbers for the figures if the restricted data is not available.

New users should follow these steps to run the package successfully: In Anaconda prompt, navigate to the main reproduction package directory, then: 1) create a new environment called ‘envir\_bg’: “conda env create –file=envir\_bg.yml” 2) activate environment: “conda activate envir\_bg” 3) make environment available for papermill: “python -m ipykernel install –user –name envir\_bg” 4) run main script: “python main.py”

The provided code reproduces selected tables and figures in the paper, as explained and justified below

| No. | Exhibit name              | Output filename                    | Script  | Note |
|-----|---------------------------|------------------------------------|---------|------|
| 1   | Figure 1                  | Out/figure_1.png                   | main.py | —    |
| 2   | Figure 2                  | Out/figure_2.png                   | main.py | —    |
| 3   | Text for Figure B1        | Out/figure_B1.txt                  | main.py | —    |
| 3   | Figure 4                  | Out/Figure 4.png                   | main.py | —    |
| 3   | Figure 6 (left)           | Out/Figure 6 (left)                | main.py | —    |
| 3   | Figure 6 (right)          | Out/Figure 6 (right)               | main.py | —    |
| 3   | Figure 7 (oil&gas)        | Out/Figure_7_oil_gas.png           | main.py | —    |
| 3   | Figure 7 (vehicle manuf.) | Out/Figure_7_car_manufacturing.png | main.py | —    |

Figures 6 and 7 consist of two components each that are created separately. They were merged in the final manuscript.

## Software Requirements

- Anaconda (Python distribution and package manager): <https://anaconda.org/>