

## Authors (alphabetical order)

Cristian Jara Nercasseau  
Maryla Maliszewska  
Claudio Montenegro  
Israel Osorio Rodarte  
Javiera Petersen Muga  
Raimundo Smith Mayer  
Huanjun Zhang

# Gender Disaggregated Labor Database



**WORLD BANK GROUP**  
Macroeconomics, Trade & Investment

April 6<sup>th</sup>, 2020

## OUTLINE

- 1.** Motivation for constructing a new database
- 2.** Methodology
- 3.** Deliverables



# 1

## Motivation



## Motivation

---

Need for a (re)examination of the **distributional** consequences of international trade beyond macroeconomic indicators

---

Nevertheless, detailed comparable data for **global** policy analysis is **scarce, particularly for impacts by gender**

---

Objective: Create a **comparable micro-based** dataset that that can inform macro (CGE) models, providing **detailed** accounts on employment levels **by gender**, occupation, wage, employment status at a **finer level** in economic activity.

---

## Future applications within the World Bank

---

a) **Processed “ready-made” statistics,**

**To be used in the upcoming World Bank (2020) *Trade and Gender Report***

Cross-country stats such as women's labor intensity and gender gap by sector

b) **Statistics in flexible format** to inform other databases.

**To be used as an input to the Global Trade Analysis Project (GTAP) Database**

Disaggregate labor accounts by gender.

c) **Added variables to World Bank micro-data surveys**

**To be used in simulations or generation of statistics**

i.e. Impact of trade policy on WB client countries, AfCFTA, CP-TPP, USMCA

# 2

## Methodology

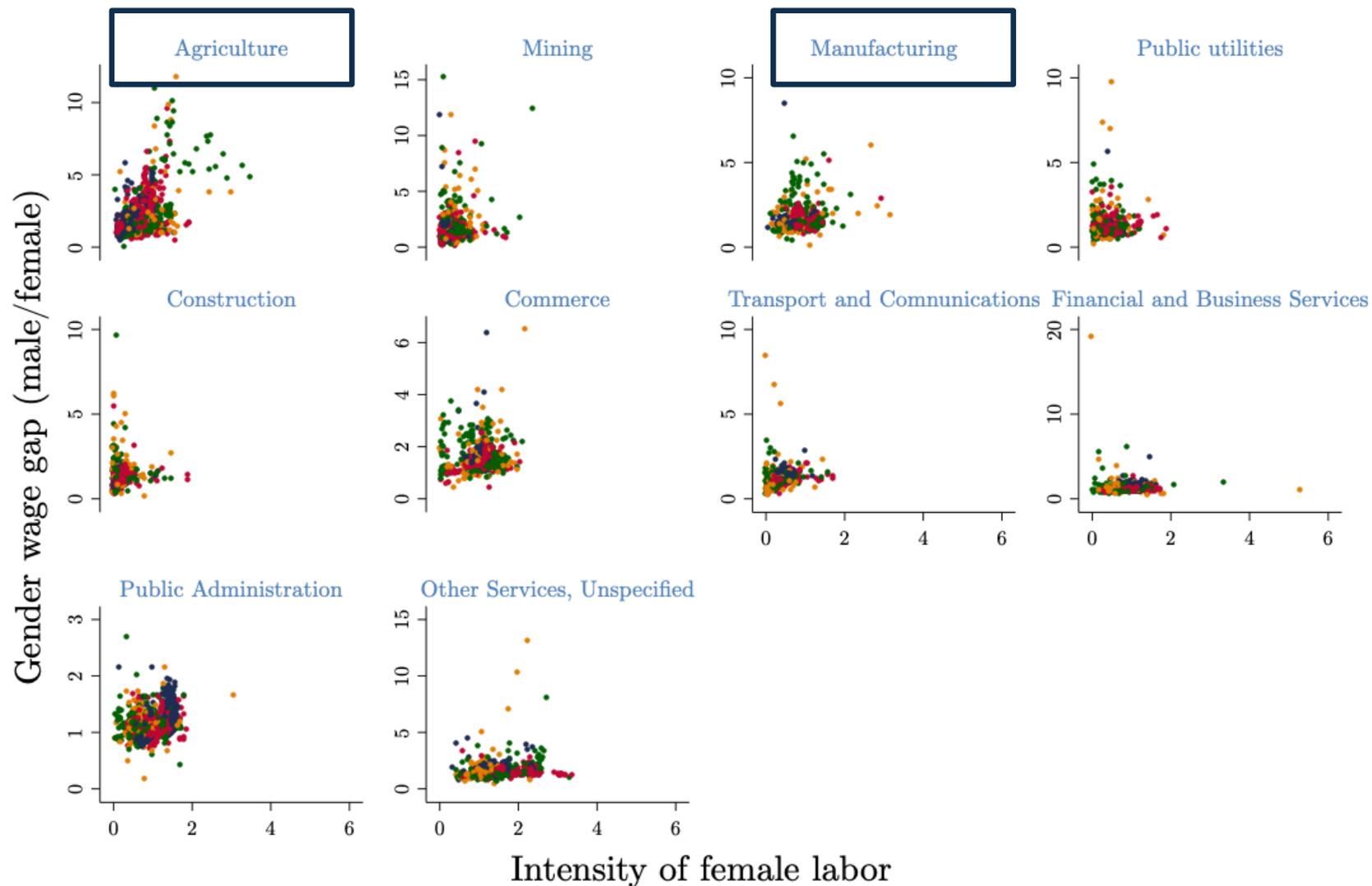


## Starting point

- The World Bank has harmonized data for 138 countries **with broad industry (3 sectors or 10 sectors)** and occupation variables that account for 90% of population and GDP



## Starting point: Global

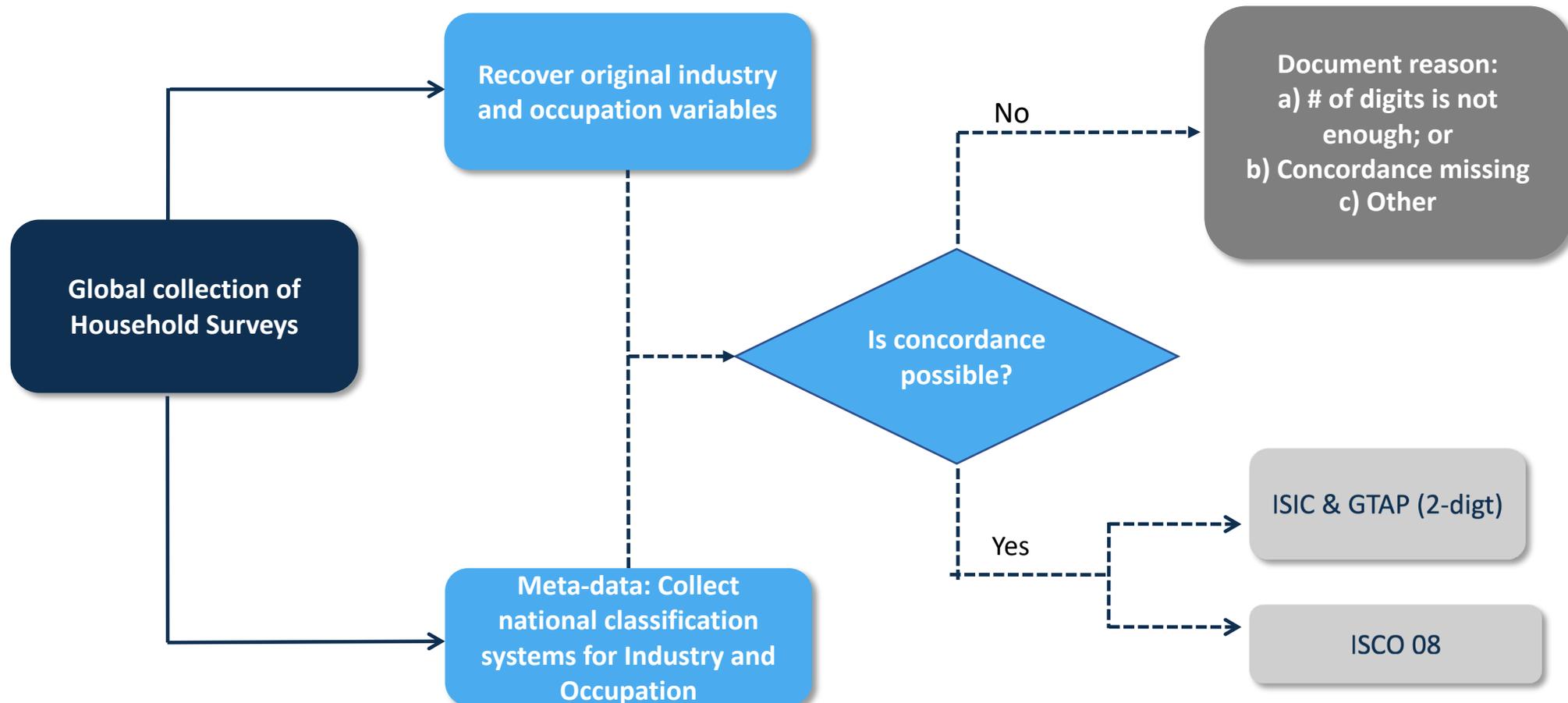


Female labor intensity:

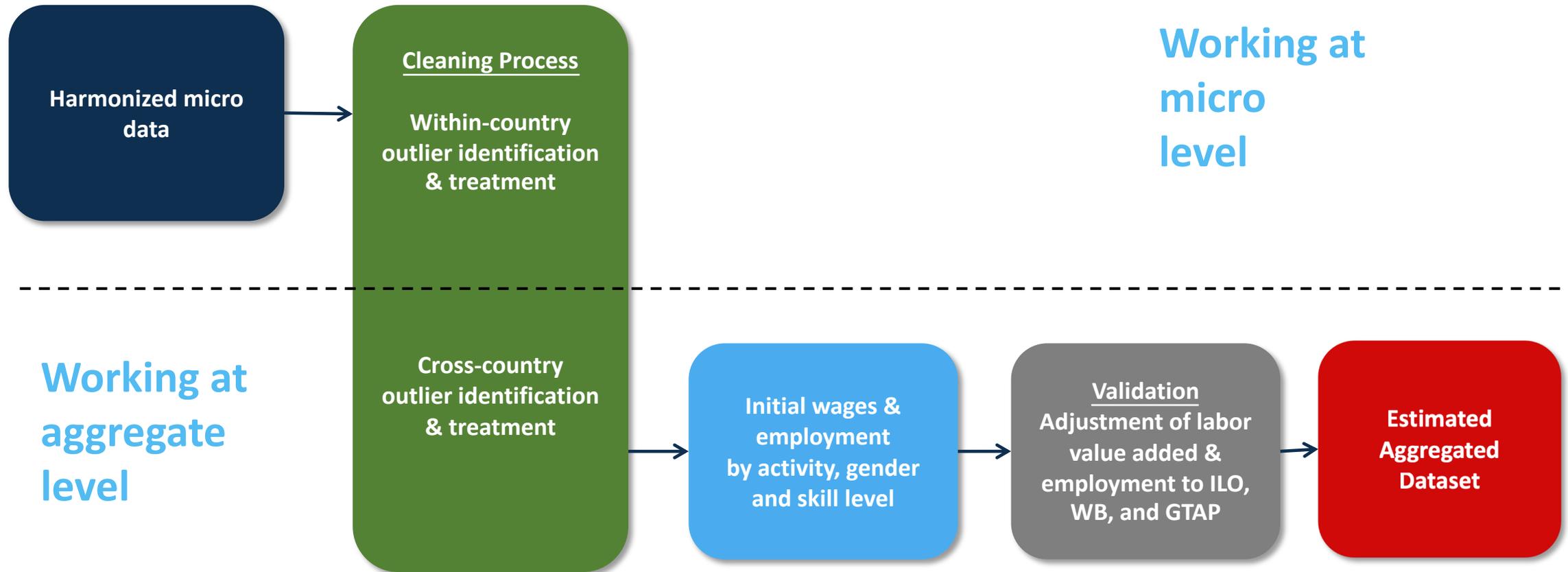
The share of female employment in an economic activity divided by the share of female employment in the country

## Harmonization Process for the Gender Disaggregated Labor Database

Creates a finer level of disaggregation for “industry” and “occupation” harmonized to international standards.



## From micro-based statistics to national aggregates matching national accounts (GTAP)

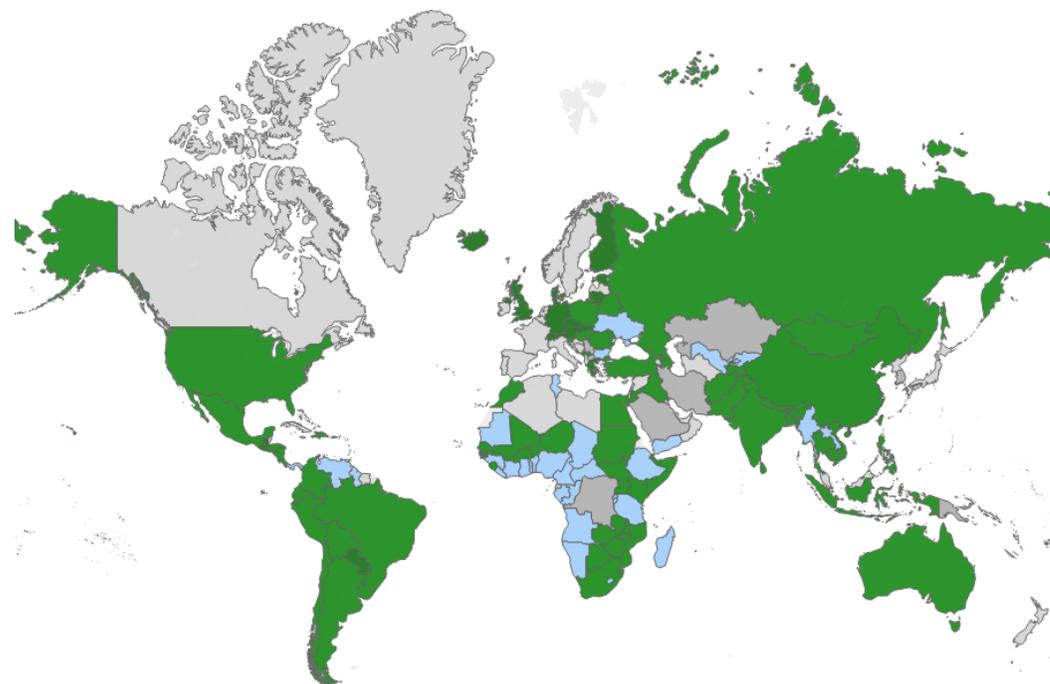


## Building a Disaggregated Labor Database, limitations

- On the microeconomic side, it requires a re-harmonization of industry and occupation data relying on a sub-sample of the World Bank collection of household surveys
- The strategy has been to find initial documentation for a cross-section of countries
- Finding disaggregated data at industry level in household surveys is not easy, because of:
  - **Focus.** The World Bank efforts to harmonize household survey data are concentrated on collecting data for monitoring “global poverty”. This has created a gap in survey coverage for **a.** data on high-income countries, and **b.** an emphasis in poverty, rather than in the production of labor statistics.
  - **Survey design.** Either when the classification system present in the household survey cannot be further disaggregated. Due to sampling, most surveys are not representative at the highest disaggregation level
  - **Data quality:** Even if data can be disaggregated, there are some internal inconsistencies, inherent to data collection in difficult environments and with teams with low-statistical capacity
  - **Inconsistent meta-data:** Comparable international information about national industry and occupation classifications doesn't exist and country information is not always easily available

## Building a Disaggregated Labor Database, progress on meta-data collection

- 219 countries-surveys were revised. When possible, information was collected about national industry and occupation classification, including number of digits in the survey



■ processed (micro data)  
■ processed (LIS)

■ not processed (micro data is not available)  
■ not processed (variable is not included in micro data)

■ not processed (missing documentation or miscoded variab..)

## Building a Disaggregated Labor Database, progress on performing conversion

- a. ISIC Rev 4 and GTAP v10 codes

<i>Industry</i>	East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	North America	South Asia	Sub-Saharan Africa	Total
Harmonized in individual-level format	11	12	15	7	1	8	22	76
Harmonized from Luxembourg Income Study		13	2					15
Not harmonized, concordance missing	4	4	3	1			12	24
Not harmonized, not enough digits	4	11	3	3			11	32
Not harmonized, raw Data before 2000	1		5					6
Microdata is not available in WB domain	8	8	1	8			2	27
Microdata do not exist (territories/small states)	7	9	13		1			30
Microdata do not contain industry variable	3	2		2	1		1	9
<b>Total</b>	<b>38</b>	<b>59</b>	<b>42</b>	<b>21</b>	<b>3</b>	<b>8</b>	<b>48</b>	<b>219</b>

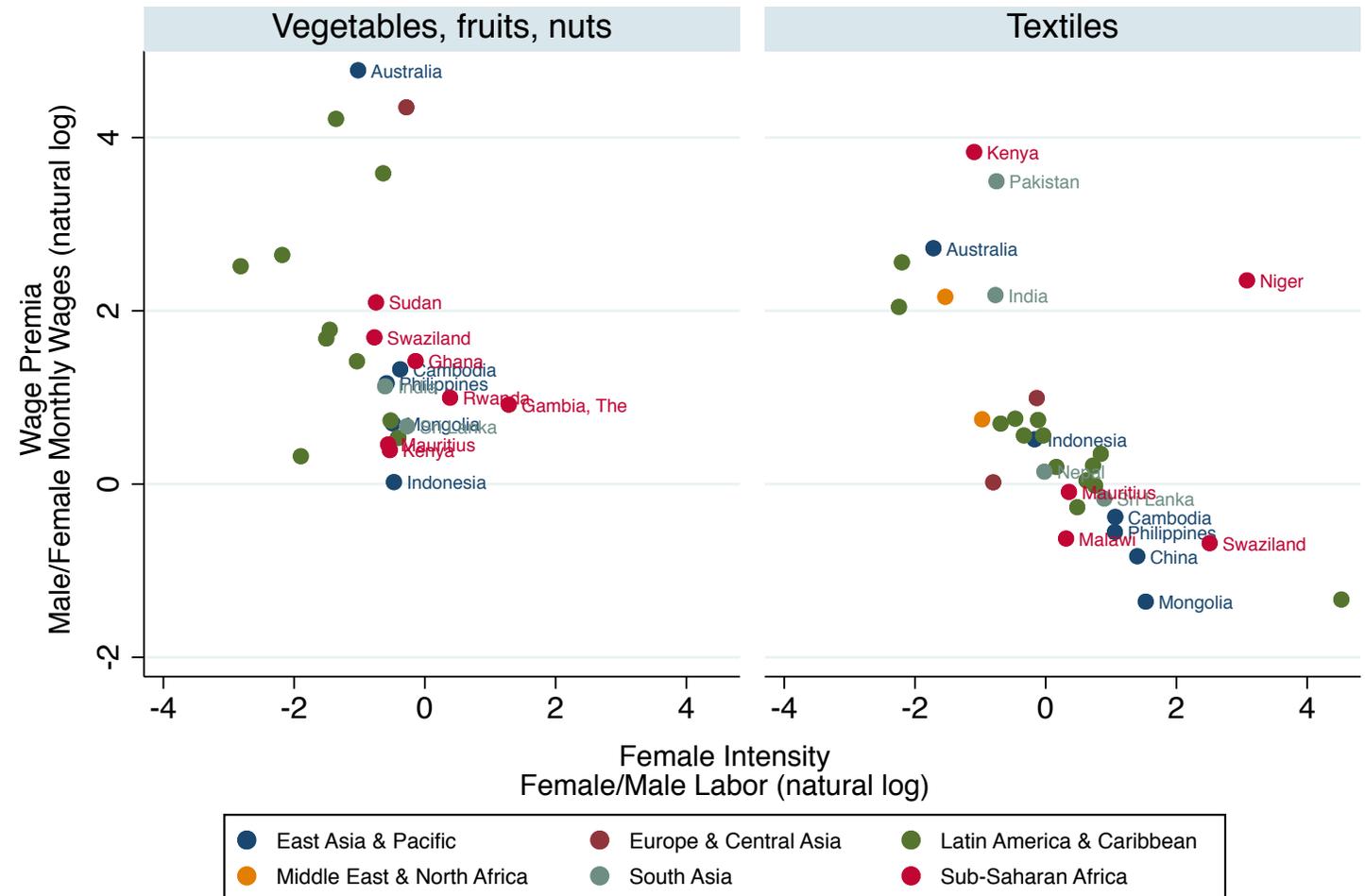
## Building a Disaggregated Labor Database, progress on performing conversion

- and/or **b.** ISCO 08 codes have been processed for:

<i>Industry</i>	East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	North America	South Asia	Sub-Saharan Africa	Total
Harmonized in individual-level format	11	7	13	5	1	7	19	63
Harmonized from Luxembourg Income Study		13	2					15
Not harmonized, concordance missing	4	9	7	2		1	14	37
Not harmonized, not enough digits	3	10	1	4			3	21
Not harmonized, raw Data before 2000	1		5					6
Microdata is not available in WB domain	8	8	1	8			2	27
Microdata do not exist (territories/small states)	7	9	13		1			30
Microdata do not contain industry variable	4	3		2	1		10	20
<b>Total</b>	<b>38</b>	<b>59</b>	<b>42</b>	<b>21</b>	<b>3</b>	<b>8</b>	<b>48</b>	<b>219</b>

## An application example for finer level industry

- Two prominent options for diversification in Africa are a. Vegetables and Fruits and b. Textiles
- Vegetables and fruits have higher value added than traditional agricultural products such as crops, coffee, cocoa
- Textiles, on the other hand, have been the base of manufacturing expansion in East Asia
- Due to high intensity of female labor and gender wage premia, both sectors can have different implications in job creation, by gender



Graphs by GTAP(65) industry classification

# 3

## Deliverables



## Building a Disaggregated Labor Database, what has been done

The initial deliverables are:

### 1. Harmonized micro data (household survey)

- with labor module including employment status, labor status, demographic information, wage and income, ISCI Rev 4 (2-4 digit) and GTAP10, ISCO 08 (2-4 digit)
- Only open for the users who have the access to the WB household survey data

### 2. Harmonization Tool (in Stata)

- Allows user to harmonize a new household survey between national and international classifications. This is critical for future scaling-up this project.

### 3. Global statistics on wages and employment by economic activity, skill level and gender

- This global statistics match GTAP v10 value-added data

### 4. Meta-data on national classification systems

### 5. Document and Visualization – to be released in March 2020\*

- World Bank Website
- World Bank Open Data Bank
- World Bank Data Catalog

## Dataset: Harmonized micro data for internal use only

- The GDLD database complements global and regional harmonization efforts within the World Bank,
  - Particularly, economic activities and occupation variables were re-harmonized to match international classifications systems (ISIC Rev 4/ISCO 08)
  - Additionally, a proxy variable of “wage per month” was created
  - Because it is based on record-level data, the dissemination of the dataset is restricted to only within the WBG. Nevertheless, the underlying code can be shared without restriction
  - At the global level, the harmonized micro-database accounts for 71% of GDP and 78% of population



## Dataset: Regional coverage of harmonized micro data

Gender Disaggregated Labor Database: Micro-based coverage (economic activity)					
Regions	Population		GDP		Surveys
	millions	% Covered	US\$Billions	% Covered	
<i>East Asia and the Pacific</i>	2,329	85	26,058	66	11
<i>Europe and Central Asia</i>	921	57	23,161	55	25
<i>Latin America and the Caribbean</i>	610	96	5,215	96	17
<i>Middle East and North Africa</i>	449	45	3,674	20	7
<i>North America</i>	364	90	22,283	92	1
<i>South Asia</i>	1,814	100	3,454	100	8
<i>Sub-Saharan Africa</i>	1,087	47	1,706	53	22
<b>Total</b>	<b>7,574</b>	<b>78</b>	<b>85,550</b>	<b>71</b>	<b>91</b>

Notes: Includes surveys from World Bank repositories and aggregates from Luxembourg Income Study  
 Based on World population and GDP in 2018

## Tool: Stata ado-files (re)-run the harmonization

- **gdld\_class**
- Provides a framework to harmonize national industrial or occupational classifications to their ISIC Rev 4, GTAP v10 or ISCO-08 counterpart
- formed by a set of STATA programs that runs on top of individual-level micro data
- It helps users to replicate the harmonization done in this project and run their own harmonization based on the local classification information
- It is a critical piece in scaling this project and sharing it with units inside and outside the World Bank

## Dataset: Estimated wages and employment volume (for CGE modeling)

- Estimated statistics based on the harmonized household survey dataset
  - It provides the estimated wages and employment volumes disaggregated by economic activity, skill level and gender
    - Particularly useful for the GTAP 10 database with 65 sectors and 141 regions
  - This database has been adjusted to match macro statistics on labor force and wage bill (available in World Development Indicators and ILOSTAT)
- It includes WB and GTAP regional mappings

## Metadata: national classification systems for economic activity and occupation

- Metadata collected throughout this project.
- It includes information for economic activity and occupation variables present in World Bank household surveys
- The metadata was collected from extensive search, including the national statistical offices, survey documents, external websites, and other sources
- It helps to identify national classifications used by the surveys, in order to create the harmonization process from local classification to international classifications

## Platform: GDLD Website

- GDLD website provides a glance for this project, to download necessary data and files, visualize the database for global comparison, and link to applications such as Trade and Gender Report and AfCFTA project.
  - The internal review link is <http://datatopicsqa.worldbank.org/gess/>
  - The final public URL would be <http://datatopics.worldbank.org/gdld/>



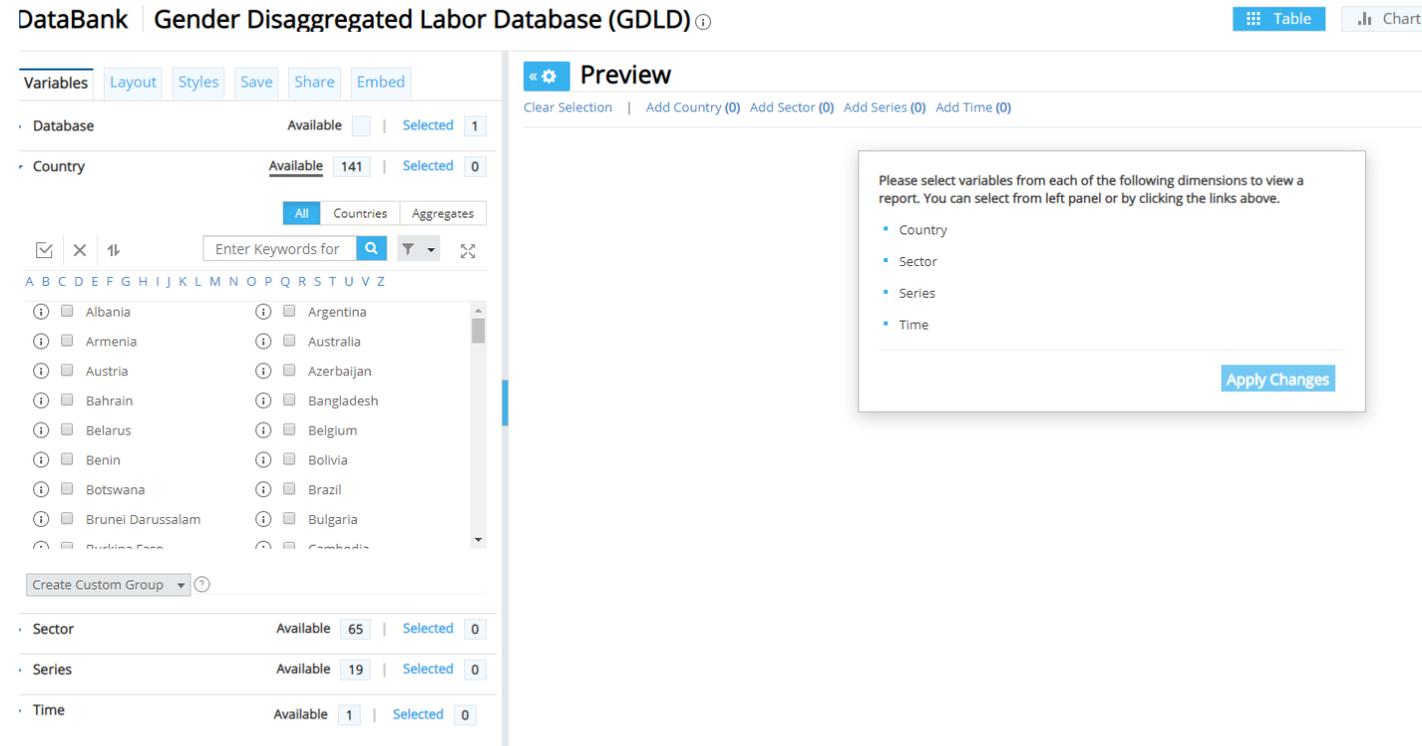
**GENDER DISAGGREGATED LABOR DATABASE**  
Data, Tools and Guidance

The Gender Disaggregated Labor Database (GDLD) is a global micro labor force database based on the World Bank's household survey collection and other public resources. This database includes harmonized economic activities and occupation categories from local classification to international comparable classifications in detailed levels. It fills an important information gap in global gender statistics by providing detailed accounts on education, employment levels, wages, labor income, and employment status at a very disaggregated (2-digit ISIC level) economic activity level and occupation category than is commonly available.

[READ MORE](#)

## Platform: Data Bank Page

- Data Bank page ([https://databankqa.worldbank.org/source/gender-disaggregated-labor-database-\(gldd\)](https://databankqa.worldbank.org/source/gender-disaggregated-labor-database-(gldd)))
- It is an interactive platform allowing users to select, preview and download necessary summary data. GDLDData Bank page is 4-dimension for country, sector, series and time. In current stage, it is in internal review status.



**DataBank** | Gender Disaggregated Labor Database (GDLDD) ⓘ

Table | Chart

Variables | Layout | Styles | Save | Share | Embed

Database Available 1 | Selected 1

Country Available 141 | Selected 0

All | Countries | Aggregates

Enter Keywords for 🔍

A B C D E F G H I J K L M N O P Q R S T U V Z

Albania  Argentina  
 Armenia  Australia  
 Austria  Azerbaijan  
 Bahrain  Bangladesh  
 Belarus  Belgium  
 Benin  Bolivia  
 Botswana  Brazil  
 Brunei Darussalam  Bulgaria  
 Cambodia

Create Custom Group ⓘ

Sector Available 65 | Selected 0

Series Available 19 | Selected 0

Time Available 1 | Selected 0

**Preview**

Clear Selection | Add Country (0) | Add Sector (0) | Add Series (0) | Add Time (0)

Please select variables from each of the following dimensions to view a report. You can select from left panel or by clicking the links above.

- Country
- Sector
- Series
- Time

Apply Changes

## Platform: Data Catalog Page

- The Data Catalog page (<https://datacatalog.worldbank.org/ddhworkbench/workbench-datasets>) is the place to save all materials, which will be shared to public after the review.

Gender Disaggregated Labor Database (GDLD)

Edit

Go

Type: Time Series Data Classification: Official Use Only Created by: Mohammed Omar Hadi

Created on: December 20, 2019 Updated by: Huanjun Zhang Updated on: December 20, 2019

TTL(s): Maryla Maliszewska Collaborator(s): Huanjun Zhang, Israel Osorio-Rodarte

Published Version Not Available Current Version: [1344436 \(New Dataset\)](#)

Estimated wage and volume for CGE modeler

[Edit Resource](#)

Data Classification: Official Use Only Exception: 9. Deliberative Information Format: EXCEL

Resource Type: Download Updated on: January 30, 2020

Published Version: Not Available Current Version: [1344441 \(New Dataset\)](#)

Metadata

[Edit Resource](#)

Data Classification: Official Use Only Exception: 9. Deliberative Information Format: xlsx

Resource Type: Download Updated on: January 30, 2020

Published Version: Not Available Current Version: [1344446 \(New Dataset\)](#)



**WORLD BANK GROUP**  
Macroeconomics, Trade & Investment

[www.worldbank.org](http://www.worldbank.org)