

# Introduction to the Tax Gap Toolkit

Dr Amina Ebrahim, Research Fellow, UNU-WIDER

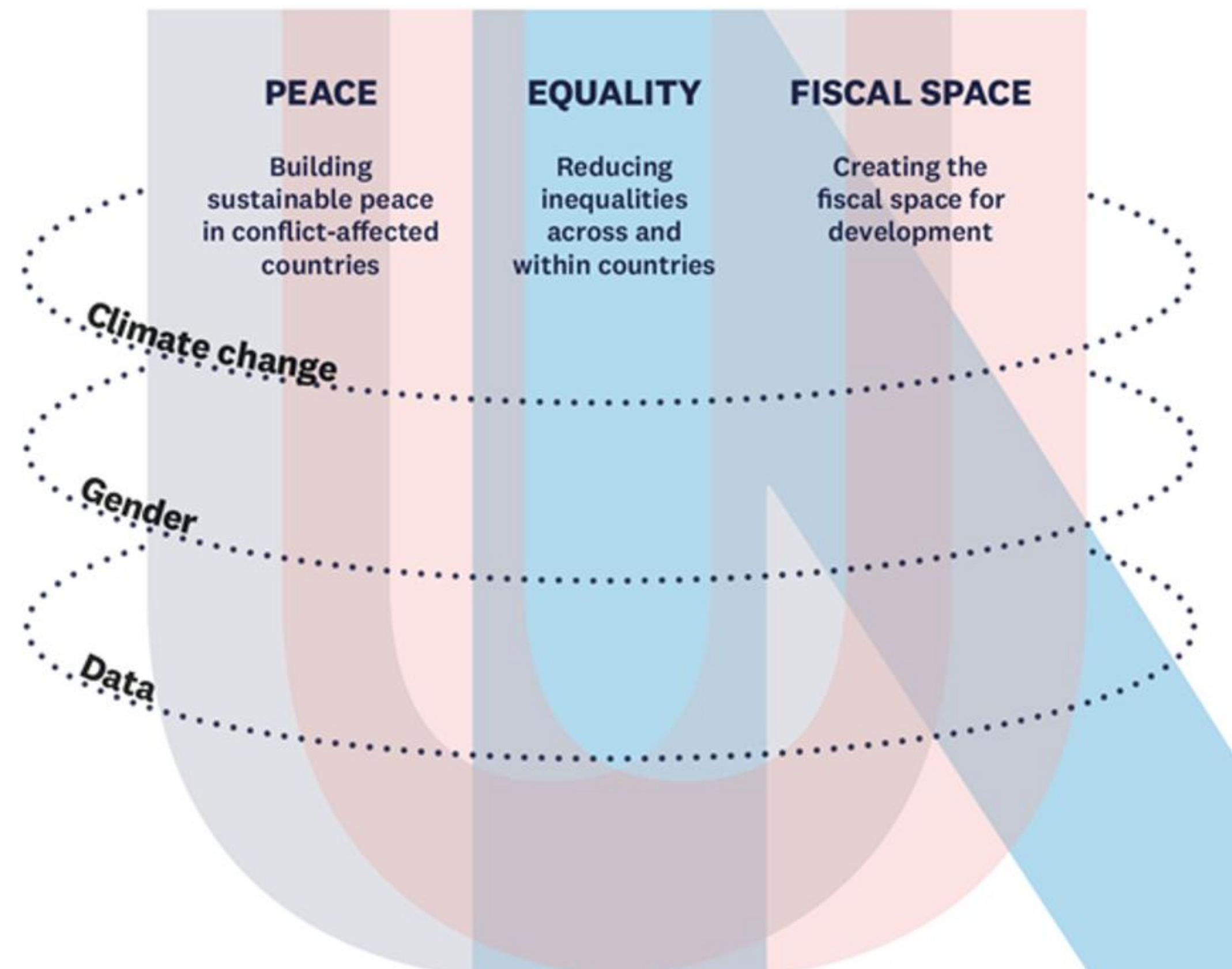
7 May 2025 | Follow-up training



The Tax Gap toolkit was developed with support from the [International Tax Compact \(ITC\)](#), which facilitates the Secretariat of the ATI. The ITC is funded by the [German Federal Ministry of Economic Cooperation and Development](#), co-funded by the [European Union](#), and implemented by the [Deutsche Gesellschaft für Internationale Zusammenarbeit \(GIZ\) GmbH](#).



# About: UNU-WIDER



- Core set of activities:
  1. Undertaking research & policy analysis
  2. Providing a forum for knowledge-sharing & debate
  3. Offering capacity development & training
- The main aim of UNU-WIDER's 2024-27 work programme is to harness research and evidence to address three mega challenges which will define the course of progress towards sustainable development in the decades to come:
  - **PEACE:** Building sustainable peace in conflict-affected countries
  - **EQUALITY:** Reducing persistent inequalities across and within countries
  - **FISCAL SPACE:** Creating the fiscal space for development.

# Fiscal Space Research Area



**Tax research for  
development**



**SOUTHMOD**



**The political economy  
of taxation**



**Reforming the international  
financial architecture**

- Research in the area cuts across the entire Institute's mission:
  - Research and policy analysis
  - Capacity development and training
  - Knowledge sharing, discussion, and debate

# Tax Research for Development

## Main objectives:

1. **Strengthen tax administrators' skills** through research collaboration
2. **Help policymakers** create better development policies with research insights.
3. **Build communities** focused on using tax data in the Global South.

## Tax Gap research

- Start: Understand the Tax Gaps in Zambia, estimating VAT and CIT at disaggregated levels
- Continue: Looking at VAT Gap in Tanzania to improve tax compliance
  - paper published in [CESifo Economic Studies](#)
- Develop: A toolkit that is useful in other contexts for regular estimation
- Ongoing: Estimations in Uganda (CIT), Rwanda (VAT) and South Africa (CIT & VAT)

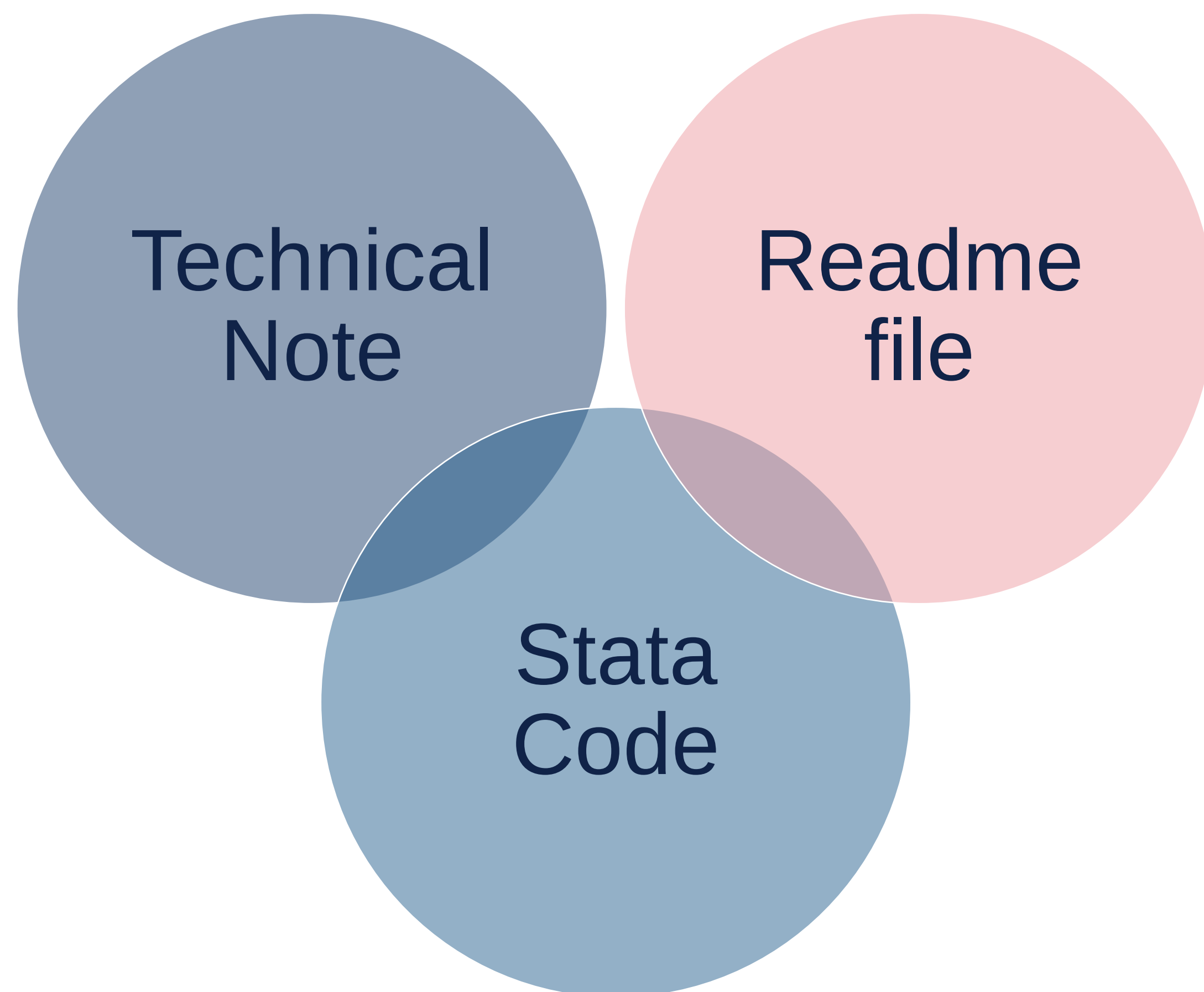
# Recap - Definition

The tax gap is the difference between the *actual* and *potential* tax collected under full compliance with the tax code.

## Terminology:

- From a policy perspective: distinguish between the **compliance gap** and the **policy gap**.
  - **Compliance gap**: additional revenues raised if tax legislation perfectly enforced.
  - **Policy gap**: revenue that could be obtained (assuming perfect enforcement) if the base were taxed at the current standard (basic, average) tax rate.

# What is the Tax Gap Toolkit?



A package of **adaptable materials** to implement tax compliance gap **estimation** based on the bottom-up approach using **machine learning**



# How does the Toolkit work?

## Part 2 Machine Learning Estimation

- The cleaned and prepared data are used to estimate the tax gap.
- For audited firms, we gather the unpaid/recovered tax information.
- The *challenge* with Tax Gap estimation that we don't know the potential tax for **unaudited firms** and unaudited periods.
- Toolkit uses info from audited firms to **estimate** the unpaid/potential recovered tax for unaudited firms and unaudited periods.
- Estimation is done using Machine Learning.

| TPIN   | AUDIT STATUS | RECOVERY ACTUAL | RECOVERY POTENTIAL |
|--------|--------------|-----------------|--------------------|
| 108469 | Audited      | 17,158.95       |                    |
| 105476 | Audited      | 14,921.44       |                    |
| 876560 | Audited      | 14,118.49       |                    |
| 858323 | Audited      | 16,138.63       |                    |
| 684517 | Unaudited    |                 | -                  |
| 124278 | Unaudited    |                 | -                  |
| 368696 | Unaudited    |                 | -                  |
| 335143 | Unaudited    |                 | -                  |
| 368696 | Unaudited    |                 | -                  |
| 212796 | Unaudited    |                 | -                  |
| 166410 | Unaudited    |                 | -                  |

# How does the Toolkit work?

## Part 2 Machine Learning Estimation

- The cleaned and prepared data are used to estimate the tax gap.
- For audited firms, we gather the unpaid/recovered tax information.
- The *challenge* with Tax Gap estimation that we don't know the potential tax for **unaudited firms** and unaudited periods.
- Toolkit uses info from audited firms to **estimate** the unpaid/potential recovered tax for unaudited firms and unaudited periods.
- Estimation is done using Machine Learning.

| TPIN   | AUDIT STATUS | RECOVERY ACTUAL | RECOVERY POTENTIAL |
|--------|--------------|-----------------|--------------------|
| 108469 | Audited      | 17,158.95       |                    |
| 105476 | Audited      | 14,921.44       |                    |
| 876560 | Audited      | 14,118.49       |                    |
| 858323 | Audited      | 16,138.63       |                    |
| 684517 | Unaudited    |                 | 6,517.63           |
| 124278 | Unaudited    |                 | 33,765.42          |
| 368696 | Unaudited    |                 | 22,256.54          |
| 335143 | Unaudited    |                 | 36,984.14          |
| 368696 | Unaudited    |                 | 16,175.93          |
| 212796 | Unaudited    |                 | 23,144.84          |
| 166410 | Unaudited    |                 | 13,103.30          |

# Why machine learning?

**Machine learning** uses both *audited* and *unaudited* taxpayer data to identify patterns and learn the most relevant variables, then predicts the tax potential revenue.

## Strengths of Machine Learning:

- Captures complex, non-linear relationships without strict distributional assumptions, and adapts to complex data structures.
- Improved prediction power over traditional statistical methods in predicting tax evasion.
- Addresses selection bias from observable factors by including all known variables influencing audit selection.
- Advanced ML techniques can incorporate methods, like the Heckman correction, to account for unobserved factors influencing selection bias.

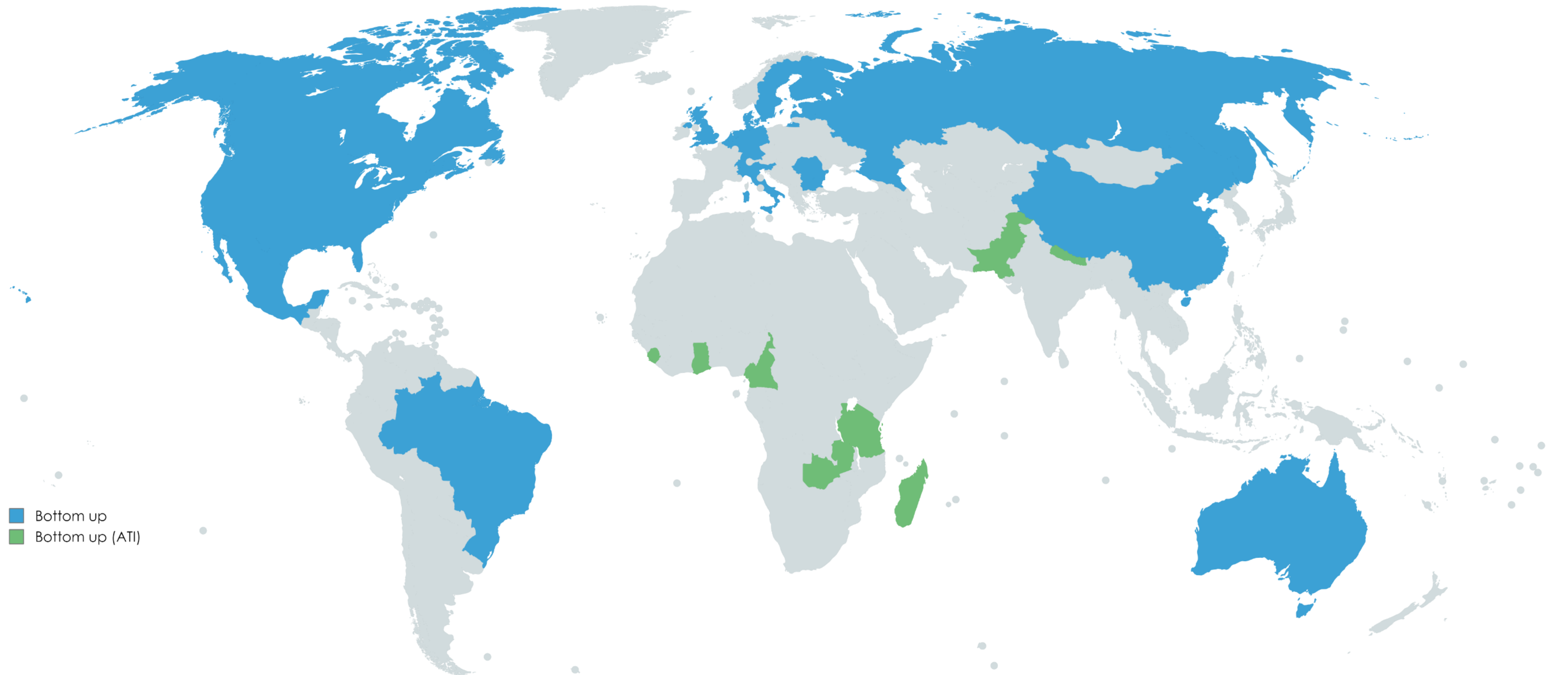


# Key problems the Toolkit solves

- Guides the process
- Code to prepare the data
- Replicability
  - Country contexts
  - Tax types
- Easy to use
- Promotes staff learning
- Possible to institutionalize

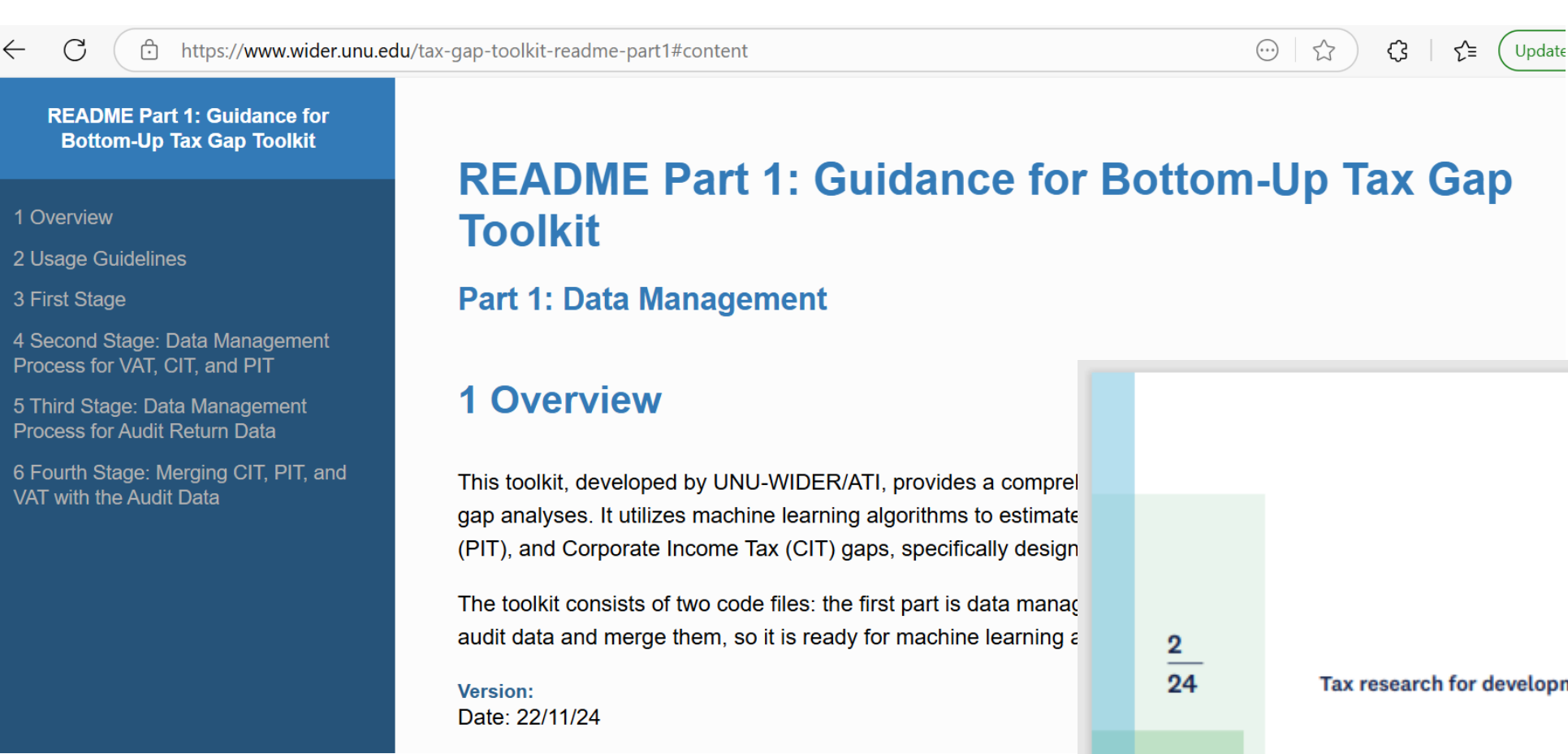


# An achievable reality? Map of countries using a bottom-up approach



# Toolkit materials – available for free

Tax Gap Toolkit [Technical Note](#), [Readme files](#) and example [Stata codes](#) now available on the UNU-WIDER’s webpage. (Available in French, Spanish and Portuguese).



The Tax Gap toolkit was developed with support from the **International Tax Compact (ITC)**, which facilitates the Secretariat of the ATI. The ITC is funded by the **German Federal Ministry of Economic Cooperation and Development**, co-funded by the **European Union**, and implemented by the **Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH**.



**UNU  
WIDER**



Facilitated by



More information available online

**wider.unu.edu**

Follow us on LinkedIn, Facebook, and Twitter @UNUWIDER

# TAX GAP ESTIMATION TOOLKIT FOLLOW UP TRAINING

## Agenda – 7 May 2025

Helsinki time

|                      |                                             |
|----------------------|---------------------------------------------|
| 14:00 – 14:15        | Introduction to the Tax Gap Toolkit         |
| <b>14:15 – 15:30</b> | <b>Session 1 Data preparation practical</b> |
| 15:30 – 16:00        | Break (30 mins)                             |
| <b>16:00 – 17:30</b> | <b>Session 2 Gap estimation practical</b>   |
| 17:30 – 18:00        | Q&A and Conclusion                          |