LOGISTICS PERFORMANCE INDICATORS (LPI 2.0) REPRODUCIBILITY VERIFICATION – DETAILED README

OVERVIEW

This reproducibility verification package provides a description of the end-to-end workflow needed to replicate and to verify the results of the new release of the Logistics Performance Indicators (LPI 2.0): underlying dataset and content used in the report (LPI 2.0: Connecting to Compete)[1]. This guide clarifies the steps needed to process, validate, and integrate confidential data from major private sector partners and international organizations (including a large global shipping line, Cargo iQ/IATA, MDS Transmodal, MarineTraffic, and UPU) with public sources inputs used for the report visualizations: World Bank's World Development Indicators; and CEPII's Geodist database. The pipeline is split across two computing environments: a secure, high-performance server for processing data obtained from the global shipping line (i.e., export and import container dwell time indicators, lead time indicators, supply chain initiation and supply chain termination indicators) and a desktop environment for processing the remaining indicators (maritime, aviation, and postal connectivity as well as time-based indicators, followed by consolidation and preparation of visualization materials for the report.

Reproduction proceeds in three stages: 1 – server-side scripts generate confidential intermediate indicators; 2 – desktop scripts process maritime, air, and postal inputs; 3 – final consolidation script merges all outputs, saves them, pulls WDIs, generates report figures. The resulting, publication-ready datasets are stored in the *~/output_database* folder, and charts are placed under *~/report/figures.* All scripts and the results in the report and *~/output_database* folder can be made available publicly. The rest of the files, including intermediate results saved locally and on the server, are confidential and should be verified virtually.

Package was prepared by Daria Ulybina (dulybina@worldbank.org) on February 26, 2026. Updated on March 19th, 2026.

DATA AVAILABILITY

- Most **LPI 2.0 database input data** is shared by the data partners with the World Bank LPI team on a confidential basis and cannot be made publicly available or with WB colleagues outside the LPI team.

Files in the following directories are confidential:

- *~/maritime/** folder including *~/maritime/extra/** *~/maritime/mt inputs/** and *~/maritime/mdst inputs/** files with data from MDS Transmodal and Marine Traffic including country names-country codes mapping file (*MDST Country names mapping.csv*). Last accessed: February 2026.
  - **MDS Transmodal:** Confidential data under licensed agreement without the right to redistribute. For more information, visit: https://www.mdst.co.uk/
  - **Marine Traffic**: Confidential data under licensed agreement without the right to redistribute. For more information, visit: https://www.marinetraffic.com/
- **<<server>>** Strictly confidential data located on the secured server (data provided by a large global shipping line, name cannot be disclosed due to confidentiality agreement). Last accessed: February 2026.
- *~/iata/** - confidential data provided by Cargo iQ under MOU to be used only for LPI project. Locally saved files last accessed: February 2026. For more information on Cargo iQ, visit: https://www.iata.org/en/programs/cargo/cargoiq/

[1] Arvis, J.-F., M. Herrera Dappe, D. Ulybina, and C. Wiederer (forthcoming 2026) Connecting to Compete 2025: The New Logistics Performance Indicators 2.0. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO

- *~/postal/** confidential data generated at the request of LPI team by the Universal Postal Union (UPU) under the terms of MOU agreement. Last accessed: February 2026. For information on data provider, visit: https://www.upu.int/en/home

Most **LPI 2.0 report's input data** (processed and aggregated results from the data sources above) can be published under Creative Common Attribution 4.0 license. These are files located in *~/output_database/* folder and in the *~/report/* folder only.

Additional data used to produce visualizations in the report includes the following:

- **World Development Indicators** accessed via World Bank's V2 Indicators API. Example: https://api.worldbank.org/v2/country/{countries}/indicator/{indicator}?format=json&per_page=10000"
  Indicators downloaded (Indicator.Code: Indicator.Name):
  - NY.GNP.PCAP.PP.CD: PPP GNI per capita (current intl $)
  - NY.GNP.MKTP.PP.CD: GNI PPP, (current intl $)

  Script used: *~/report/scripts/gni_utils.py*
  Downloaded data saved in *~/metadata/GNI_WDIs.parquet*
  License: Creative Common Attribution 4.0 Last accessed: February 2026

- **CEPII GEODIST dataset**, available under Etalab 2.0 license. Download data here: https://www.cepii.fr/cepii/en/bdd_modele/bdd_modele_item.asp?id=6 Download country-specific **(geo_cepii.xls**) and dyadic files (**dist_cepii.xls**). Reference: Mayer, T. & Zignago, S. (2011), Notes on CEPII's distances measures: the GeoDist Database*, CEPII Working Paper 2011-25* Last accessed: October 2025. As of February 2026, the download is available.

CODE INSTRUCTIONS

1. Server-side processing (Dwell time indicators from large shipping line)

1.1 Go to the server home directory. It should have the following structure:

```
#/home
#---/scripts/
#---/DATA/
#---/rawinput/
#---/interm/
#---/phases/
#---/indicators/
#---/figures/
#----ccodes.csv
#----decoded events.csv
#----lpi_server_env.txt (use this file to replicate environment)
```

1.2 Open the file `<server>>/scripts/runall.py` and change the line # 18 with the actual top-level home directory.

1.3 Execute `runall.py.` The script will automatically call other scripts located in that folder. Once finished, the results will be placed in *<<server>>/figures/** and *<<server>>/indicators/**

folder for final steps of data processing and visualizations. They will be read from the local computer and a different python environment.

2. Desktop processing

Either run in the following order or use the script `gendata.py` located in the home directory to trigger execution of these scripts automatically. No need to adjust paths at the top of each script – they should be resolved automatically except for the final consolidation script.

2.1 IATA[2]

2.1.1 `~/iata/Scripts/BALKANS DATA 2023-24.py`

2.1.2 `~/iata/Scripts/cargoiq processing [repr].py`

2.2 MARITIME[3]

2.2.1 `~/maritime/scripts/maritime_connect [repr].py`

2.2.2 `~/maritime/scripts/marinetraffic.py`

2.3 POSTAL[4]

2.3.1 `~/postal/Script/postal [repr].py`

3. CONSOLIDATION (NO NEED FOR PATH ADJUSTMENTS)

3.1 `~/run all database.py`.

This script consolidates all of the previously generated outputs (on the server and in the desktop environment) as well as triggers the generation of all of the images for the report. Prior to that, you can update/import selected WDIs indicators using World Bank APIs. (The module (*gni_utils*) can be imported and used – instructions on data download are available at the top of this script behind the directory setting options).

`run all database.py` will iterate over the output folders of each data provider (generated using the scripts above) and collect relevant outputs.

This script automatically imports one custom module for downloading relevant WDI indicators and triggers 3-4 additional scripts (also automatically):

- Module for the download of a GNI-related WDI indicators using WB API: `~/report/scripts/gni_utils.py`. These indicators are saved under **~/metadata/** for future usage for reports chart creation.
- Script to compute and create Market Access Connectivity charts: `~/report/scripts/market_access_conn.py`

Script for creating all other visualizations from the report - `~/report/scripts/gen_all_charts.py.` This script references files on the server – adjust lines 20 and 21 if the path is different from *"Z:/figures/"*

[2] *(lpienv) C:\Users\XXX > python "HOME_DIRECTORY\iata\Scripts\BALKANS DATA 2023-24.py"*

(lpienv) C:\Users\XXX > python "HOME_DIRECTORY\iata\Scripts\cargoiq processing [repr].py"

[3] *(lpienv) C:\Users\XXX >python "HOME_DIRECTORY\maritime\scripts\maritime_connect [repr].py"*

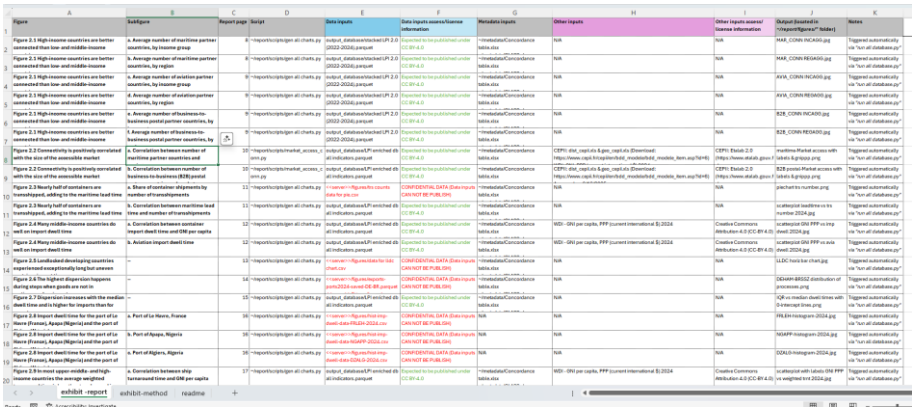
(lpienv) C:\Users\XXX > python "HOME_DIRECTORY\maritime\scripts\marinetraffic.py"

[4] *(lpienv) C:\Users\XXX > python "HOME_DIRECTORY\postal\Script\postal [repr].py"*

- Script for creating two charts with main summary indicators for Egypt and its regional and income group aggregates (part of Methodological Note): `methodology_egypt.py`. Triggered automatically.

LIST OF EXHIBITS

List of exhibits is available in an Excel file distributed together with this guideline. File name: ***report-feb26-figures-exhibit.xlsx***. There is an exhibit table worksheet as well as a readme worksheet with clarifications on the headers and color-coding conventions. Screenshot of the table:



Additionally, some information stated in the report is printed out during the execution of the scripts and generation of the charts. Specific text being referenced is located

1. Page 15 (paragraph starting with “In well-functioning ports with seamless clearance systems…”)
2. Page 14 - Figure 2.6 Note

REQUIREMENTS

Environments:

LPIENV Environment of Windows 11 Desktop

Processor	**Intel(R) Core (TM) Ultra 7 165U 2.10 GHz**
Installed RAM	**32.0 GB (31.5 GB usable)**
System type	**64-bit operating system, x64-based processor**
Python environment requirements (libraries)	**lpienv.txt**
Python version	**3.12.4**

LPI_SERVER_ENV Server environment

Processor	**Intel(R) Xeon(R) Platinum 8462Y+ 2.80 GHz (4 processors)**
Installed RAM	**64.0 GB**
System type	**64-bit operating system, x64-based processor**
Python environment requirements (libraries)	**lpi_server_env.txt**

Commented [CW1]: Could you make this a shared file with Matias and me? Won’t make changes unless prompted.

Commented [DU2R1]: It is shared with you under LPI2024 folder: “Reproduced results (shared package)”. I will share with Matias as well in a few.

Python version	3.11.10

HIGHLIGHTS

Server scripts rely on package polars ==1.27.1. Additionally, you will need pandas==2.2 , NumPy==1.26, pyarrow==18.1,

The rest of the processing, aggregation and visualizations rely on standard packages: Pandas==2.3, NumPy==2.3, Seaborn==0.13, Matplotlib==3.10. Make sure that your pandas can read parquet files by installing pyarrow==21.0 package as well. Additional libraries needed for reading excel files are openpyxl==3.1 and xlrd==2.0 (for reading older Excel files with .xls extension). Additional non-native package needed for visualizations: cycler==0.12.

The processing on the server takes 57 minutes (about an hour)

The rest of the scripts (run on local computer) takes less than 10 minutes altogether.

CODE DESCRIPTION

SERVER-SIDE PROCESSING (DWELL TIME INDICATORS FROM LARGE SHIPPING LINE)

INPUTS

Everything related to this dataset and its inputs should be run directly on the server (access using Remote Desktop) using python's environment available (specs provided above under Requirements) in **lpi_server_env.txt** file that is separate and different from the local environment (desktop environment).

Commented [CW3]: Please review; sentence unclear.

Server based scripts (indicators related to dwell time sourced from large shipping line) can be generated using a single script: `runall.py` or one-by-one in the following order:

```
1) preprocess [repr].py
2) import phase [repr].py
3) export phase [repr].py
4) shipping phase [repr].py
5) empty positioning [repr].py
6) empty returns [repr].py
7) import indicators [repr].py
8) export indicators [repr].py
9) leadtime indicators [repr].py
10) empty positioning indicators [repr].py
11) empty returns indicators [repr].py
```

Additional script (`lldc inidcators.py`) located in the scripts folder is triggered automatically via the `import indicators [repr].py` script.

Raw data inputs used for dwell time indicators are in the **~/DATA/** folder. This folder along with intermediate results (folders: **rawinput, interm, phases**) stored on the server are strictly confidential and cannot be copied or moved.

Commented [CW4]: Please make bold and/or highlighted in yellow/red.

Two metadata files are included in the home directory along with environment specification (**lpi_server_env.txt**):

- **ccodes.csv** (country-level code, name, region, income group mapping including classification of the countries according to their logistics type: coastal, landlocked or small island)
- **decoded events.csv** (mapping of track-and-trace event codes with names as well as their classifications into phases of the process: imports & exports)

OUTPUTS

Data for the database

Files located in the **~/indicators** folder (all capitalized names with "…**FINAL (repr).csv"** at the end of the file name):

- IMPORT_DWELL_FINAL (repr).csv
- EXPORT_DWELL_FINAL (repr).csv
- LEADTIME_INDICATORS_FINAL (repr).csv
- LLDCs_FINAL (repr).csv
- SC_END_FINAL (repr).csv
- SC_START_FINAL (repr).csv

After being generated and saved, they will be pulled from the aggregation script located in the package directory in the local environment.

Data for figures in the report

Additional files in the **~/figures** directory are related to the generation of the report figures and showcasing the numbers referenced in text of the report.

- data for lldc chart.csv
- exports-ports2024-saved-DE-BR.parquet
- imports-ports2024-saved-DE-BR.parquet
- shipping-ports2024-saved-DE-BR.parquet
- hist-imp-dwell-data-DZALG-2024.csv
- hist-imp-dwell-data-FRLEH-2024.csv
- hist-imp-dwell-data-NGAPP-2024.csv
- trs counts data for pie.csv

DESKTOP PROCESSING

INPUTS - DATABASE

All the intermediate and raw input datasets are confidential and cannot be shared publicly. To generate these files, run the script `gendata.py` first. Finish by running `run all database.py` which produces final results needed for Data360 and for the report.

IATA

All data files in the folder **~/iata/Inputs/*.** This folder contains raw data received from data provider, metadata files with mapping of airport codes into country codes as well as files with list of countries that have to be excluded from the publication since they don't meet a threshold requirement for publishing (see report for more details). First, run **the ~/iata/Scripts/BALKANS DATA 2023-24.py** to generate a subset of data relevant to Balkan and some Caribbean countries (these needed to be generated manually since they had been aggregated in the original source files shared as Serbia (which included Montenegro/Kosovo/Serbia) and Netherland Antilles (combined BES (Dutch Caribbean Islands) and Curacao/St. Martin). The output is saved under **~/iata/Outputs/Balkans data 2023-2024 out.csv** and used in the subsequent script. Second, run ~/iata/Scripts/cargoiq processing [repr].py file that will generate an output: **Outputs\AVIA INDICATORS FINAL.csv**.

MARITIME

- **~/maritime/mdst inputs/*:** 3 files with data from MDS Transmodal (MDST) including country names-country codes mapping file (MDST Country names mapping.csv)
- **~/maritime/mt inputs/*:** 163 csv.gz files (weekly records) shared by MarineTraffic via SFTP server connection. Only LPI-necessary files are provided in this folder.
- **~/maritime/extra/*:** ports.csv file for processing MarineTraffic dataset to have mapping of PORT_ID (port identifier specific to the data provider) to UN/LOCODES and ISO-2 codes.

POSTAL

Two files located in the directory: ~/**postal/Input/*** generated at the request of LPI team by the data partner**.** To run, use the only script located in **~/postal/Script** directory. The output is saved in Output folder.

INPUTS – REPORT

LPI 2.0 DATASET

- *stacked LPI 2.0 (2022-2024).parquet* (same data as data360 file without number of observations group) Used in *gen all charts.py* script for generating regional and income group aggregates for 2 years (2023, 2024)
- *LPI enriched db all indicators.parquet* – main results file used for generating majority of the charts in the report.

CEPII

GEODIST database, available under **Etalab 2.0 license** (https://www.etalab.gouv.fr/wp-content/uploads/2018/11/open-licence.pdf)

Use the following hyperlink to access the page of the database with further instructions to download/methodology/description/contact point: https://www.cepii.fr/cepii/en/bdd_modele/bdd_modele_item.asp?id=6

Download country-specific **(geo_cepii.xls**) and dyadic files (**dist_cepii.xls**). Files under the same name were place in **~/metadata/** directory

Reference: Mayer, T. & Zignago, S. (2011), Notes on CEPII's distances measures : the GeoDist Database, *CEPII Working Paper 2011-25*

Last accessed in October 2025. As of February 2026, the download is available.

WOLRD DEVELOPMENT INDICATORS

Filename: ~/metadata/GNI_WDIs.parquet

Retrieved via the World Bank V2 Indicators API:

- https://api.worldbank.org/v2/country/all/indicator/NY.GNP.PCAP.PP.CD?format=json&per_page=10000
  (local retrieval: `import gni_utils as gu; gu.get_meta("ppp", 2024)`*)*
- https://api.worldbank.org/v2/country/all/indicator/NY.GNP.MKTP.PP.CD?format=json&per_page=10000
  (local retrieval: `import gni_utils as gu; gu.get_meta("gni", 2024)`*)*

World Bank, World Development Indicators —License: CC BY 4.0

Filename: *"~/metadata/GNI_WDIs.parquet"*, last accessed in February 2026.

METADATA

- ~/metadata/CLASS.xlsx
- ~/metadata/Concordances table
- ~/metadata/ccodes.csv (same file as on the server)

OUTPUTS: DATABASE & REPORT

Files located in this directory are the results that will become available via Data360 and LPI website. These data can be made publicly available under Creative Common Attribution 4.0 license.

- output_database/data360/*
- output_database/* (data inputs used for the report)
- report/figures/*