iTEM Open Data & Harmonized Transport Database

Sonia Yeh, Chalmers University of Technology, Sweden

Paul Kishimoto, international Institute for Applied Systems Analysis (IIASA), Austria

Manuel Pérez Bravo, Universidad Pontificia Comillas - IIT

7th International Transport Energy Modeling (iTEM) Workshop - TDC

18th September 2024 , United Nations Economic Commission for Europe, Palais des Nations, CH-1211 Geneva 10, Switzerland

Vision and aims of iTEM Open Data

Problems

- Incomplete and inconsistent transportation data across different sources.
- Differences in data collection, processing, or interpretation.
- Currently available datasets are too expensive.

Consequences

- Unable to fully understand today's transportation systems and challenges.
- Impediments in the creation of useful knowledge for the transition to sustainable and low-carbon mobility.

Goals

- To create a **transparent**, **open-data**, and **open-code** historical transportation database.
- **Open-Data:** Assembling a comprehensive collection of publicly-available transportation data.
- **Open-Code:** All code and documentation will be publicly accessible and open for modification and extension.

An example

- For the IPCC Sixth Assessment Report (AR6), we need historical data series for the following:

  - 7.1 Direct CO2 emissions from global (and regional) transport scenarios.

  - 7.2 Transport demand/activity trajectories for passenger and freight.

  - 7.3 Mode shares of passenger and freight transport

  - 7.4 Energy intensity of freight and passenger transport

  - 7.5 Fuel shares (freight and passenger transport)

- There is not an easy way to get these data and they essentially don't exist in the public space!

Evolved thinking about the process and technical requirements for high (scientific) quality data

Collection, cleaning, error correction.

Harmonization.

Diagnostics

The International Transport Energy Modeling (iTEM) Open Data & Harmonized Transport Database

A harmonized transport data set of historical values, 1970 – present https://zenodo.org/record/4287423#.Yxm6MuxBzX1

28 public datasets: 240 countries and their subdivisions, 30,000 rows, 86 variables. The dataset is **17.64% complete**

Open Data: The input files https://github.com/transportenergy/metadata/tree/master/historical/input

Open Code: The scripts for cleaning the data https://github.com/transportenergy/database/tree/master/item/historical/scripts

Data Architecture — *Identical to iTEM Modeling Comparison*
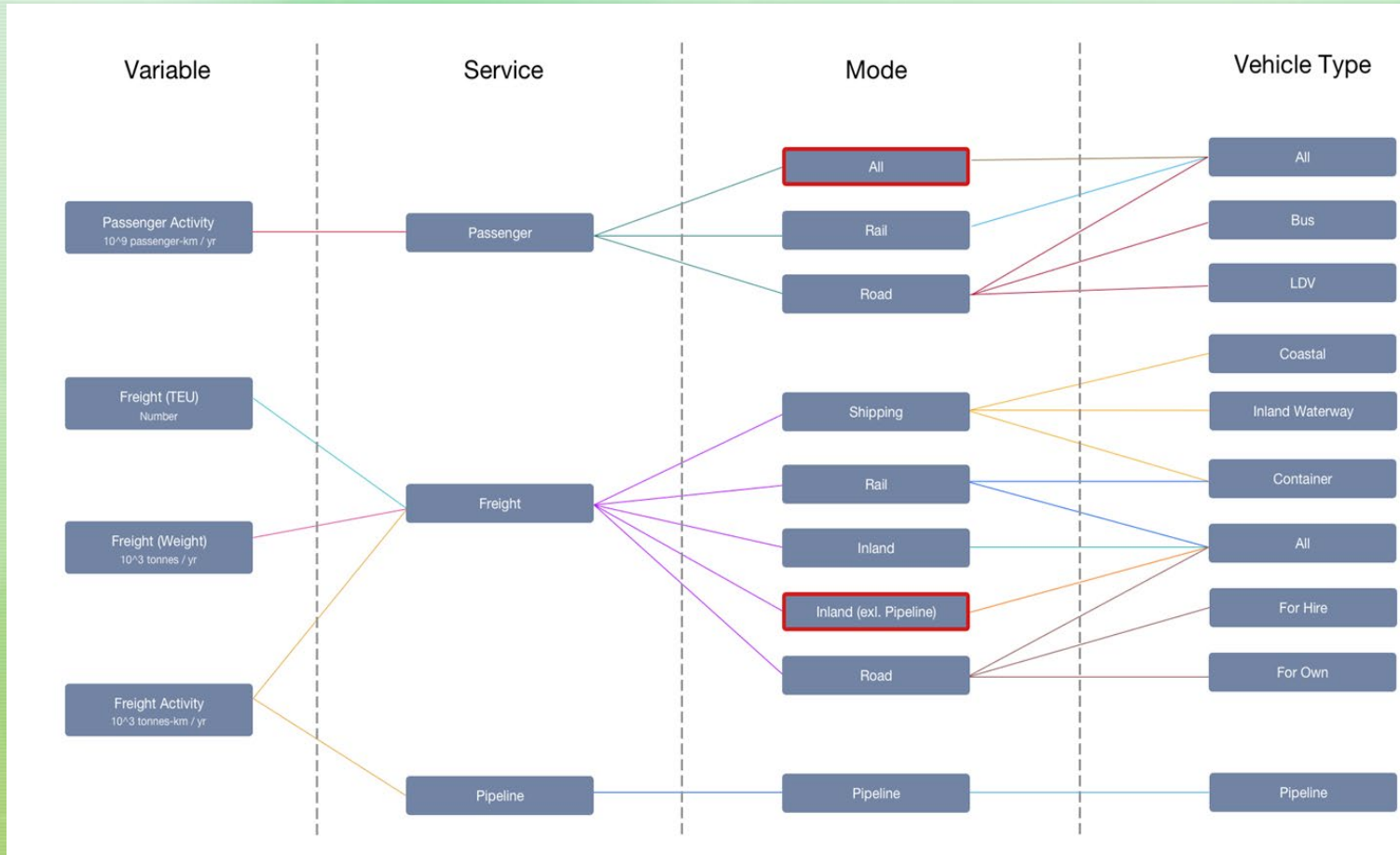
The International Transport Energy Modeling (iTEM) Open Data Project
Rule Book

Humberto Linero, Sonia Yeh, Paul Kishimoto

April 15, 2020

Contents

Data Sources and Documentation

- International Organization of Motor Vehicle Manufacturers

- Joint Research Center

- United Nations

- International Transport Forum

- iTEM

- United Nations Economic Commission for Europe

- Eurostat

- ATO 2023

- GFEI 2023

Using GitHub for open collaboration

Current and future relationships/roles

TDCI is best poised to:

- Build, set up, and operate data infrastructure (CKAN, SDMX web services, databases, etc.).

- Develop general-purpose tools for data and metadata management.

- Set and evolve community standards for data and metadata.

- Monitor and update upstream data from various sources.

iTEM is best poised to:

- Apply transport research knowledge to assess data quality.

- Develop methods to adjust data for scientific research.

- Clarify input data needs for models and data outputs.

- Provide leadership in data and metadata concepts, structures, and codes.

- Focus on quality checks, metadata clarity, and preparing data for broader audience consumption.

A first step towards an iTEM Open Data process built on TDC tolos & infrastructure

Each part of the pipeline maintained only once

Data sources

Monitor and update

TDCI

Data Tools

Structured Data

SDMX file

Transport Tools

Source data
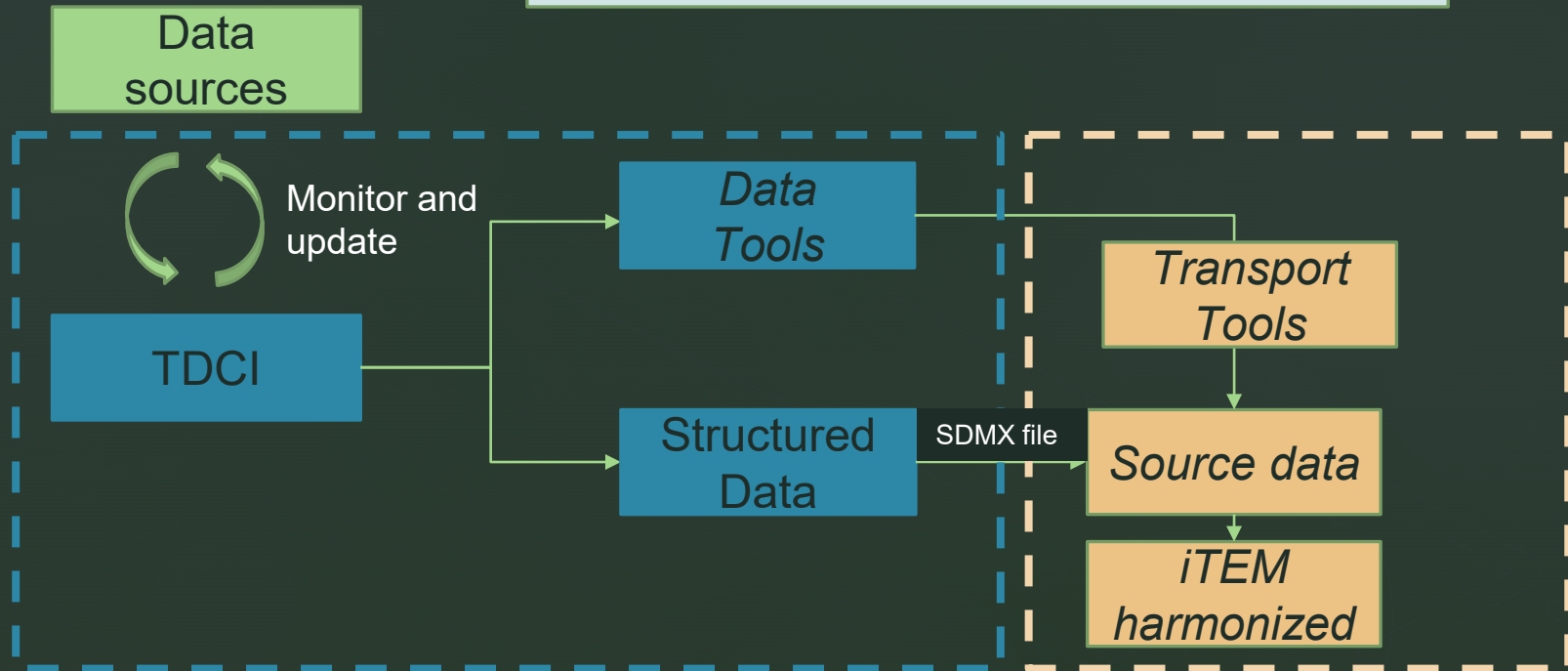
iTEM harmonized

New 13 datasets from the ATO collection

4 adult education students from disadvantaged background

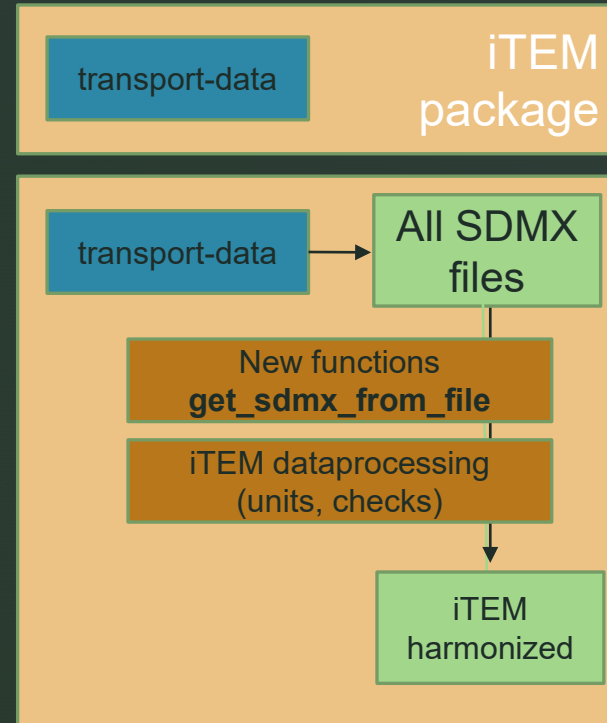13 different scripts (T011-T024)

Each **fetching** and **converting** data

iTEM package

transport-data

transport-data → All SDMX files

New functions **get_sdmx_from_file**

iTEM dataprocessing (units, checks)

iTEM harmonized

https://github.com/transport energy/database/pull/90/

iTEM Open Data & Harmonized Transport Database

Published September 11, 2024 | Version 11-09-2024

Dataset | 🔒 Open

Edit

New version

Share

The International Transport Energy Modeling (iTEM) Open Data & Harmonized Transport Database

Linero, Humberto[1]; Yeh, Sonia[1] (iD); Kishimoto, Paul[2] (iD); Cazzola, Pierpaolo[3]; Fulton, Lewis[4]; McCollum, David[5]; Miller, Joshua[6]; Kyle, Page; Pérez Bravo, Manuel[7] (iD)

Show affiliations

This dataset and documentation contains detailed information of the iTEM Open Database, a harmonized transport data set of historical values, 1970 - present. It aims to create transparency through two key features:

- Open-Data: Assembling a comprehensive collection of publicly-available transportation data
- Open-Code: All code and documentation will be publicly accessible and open for modification and extension. https://github.com/transportenergy

The iTEM Open Database is comprised of individual datasets collected from public sources. Each dataset is downloaded, cleaned, and harmonised to the common region and technology definitions defined by the iTEM consortium https://transportenergy.org. For each dataset, we describe the name of the dataset, the web link to the original source, the web link to the cleaning script (in python), variables, and explain the data cleaning steps (which explains the data cleaning script in plain English).

Shall you find any problems with the dataset, please report the issues here https://github.com/transportenergy/database/issues.

Versions

Version 11-09-2024	Sep 11, 2024
10.5281/zenodo.13749361	
Version 15-04-2020	Oct 23, 2020
10.5281/zenodo.4287423	
Version 15-04-2020	Oct 23, 2020
10.5281/zenodo.4121180	