



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



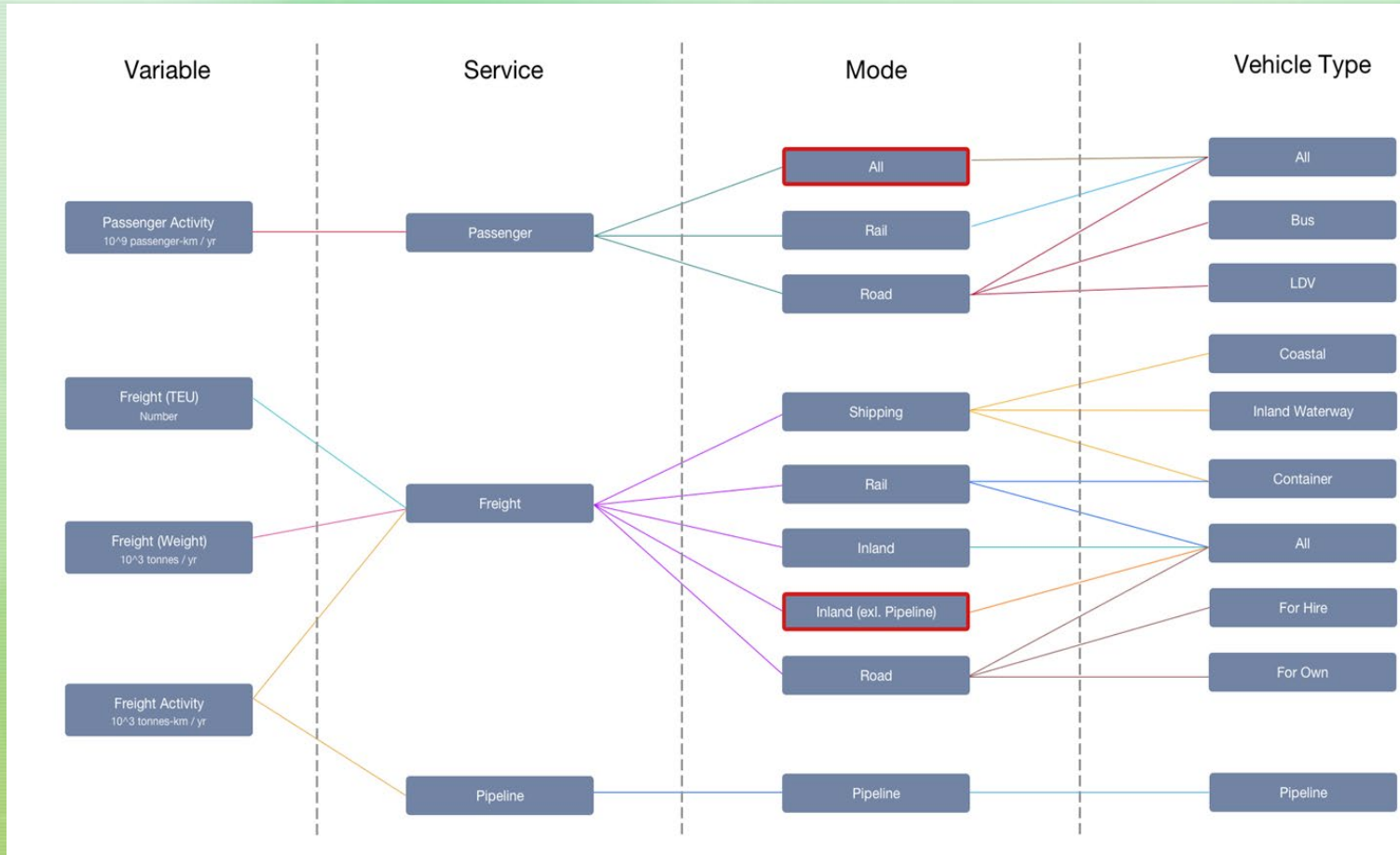
Open Data: The input files
<https://github.com/transportenergy/metadata/tree/master/historical/input>



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

<u>General Information</u>	1
<u>How To</u>	2
<u>T000 Passenger transport: Inland passenger transport</u>	3
<u>T001 Coastal Transport</u>	6
<u>T002 Container Transport</u>	8
<u>T003 Inland Freight Transport</u>	10
<u>T004 New Road Vehicle Registrations by Vehicle Category and Fuel Type</u>	13
<u>T005 Direct CO2 emissions from global (and regional) transport scenarios</u>	16
<u>T006 Modal split of freight transport</u>	19
<u>T007 Modal split of passenger transport</u>	21
<u>T008 Passenger Road Vehicle Fleet and rate per thousand inhabitants by Vehicle Category</u>	23
<u>T009 Passenger Car</u>	25
<u>T010 Commercial Vehicle</u>	27
<u>S012 De facto population (both sexes) in a country</u>	29

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

16 Open ✓ 19 Closed

Author ▾ Label ▾

- [Adjust for new OECD data endpoint](#) help wanted historical
#88 opened on Jul 9 by khaeru
- [Fetch data if not present locally](#) bug historical
#75 opened on Jul 26, 2021 by khaeru
- [T000 and T002 lack CN aviation data and recent years](#) help wanted historical
#70 opened on Apr 21, 2021 by francescolovat
- [Generate SDMX data structures for the model comparison](#) model
#65 opened on Feb 19, 2021 by khaeru
- [T003: Missing data related to the mode "Road" and mode "Total Inland FT"](#)
#61 opened on Feb 9, 2021 by hlnero
- [T000: Missing data for the 22 countries](#)
#60 opened on Feb 9, 2021 by hlnero
- [T008 data issue - Canada](#) historical
#58 opened on Feb 3, 2021 by sorlayeh
- [T008 data error for US, 2006–2008](#) historical
#55 opened on Jan 22, 2021 by sorlayeh
- [Regenerate daily](#) enh historical
#54 opened on Jan 7, 2021 by khaeru
- [Indicate observation-level data provenance](#) enh historical
#53 opened on Jan 7, 2021 by khaeru
- [Process metadata associated with input data](#) enh historical
#52 opened on Jan 7, 2021 by khaeru
- [Establish consistent information model/vocab](#) enh historical
#51 opened on Jan 7, 2021 by khaeru 1 of 2 tasks
- [Allow multi-step processing](#) enh historical
#50 opened on Jan 7, 2021 by khaeru
- [T000 India and Iceland PKT do not seem to be reasonable](#) historical
#47 opened on Dec 1, 2020 by sorlayeh
- [Automate draft release / historical database upload](#) ci enh
#45 opened on Nov 24, 2020 by khaeru
- [Add nightly CI workflow](#) ci
#44 opened on Nov 23, 2020 by khaeru



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.

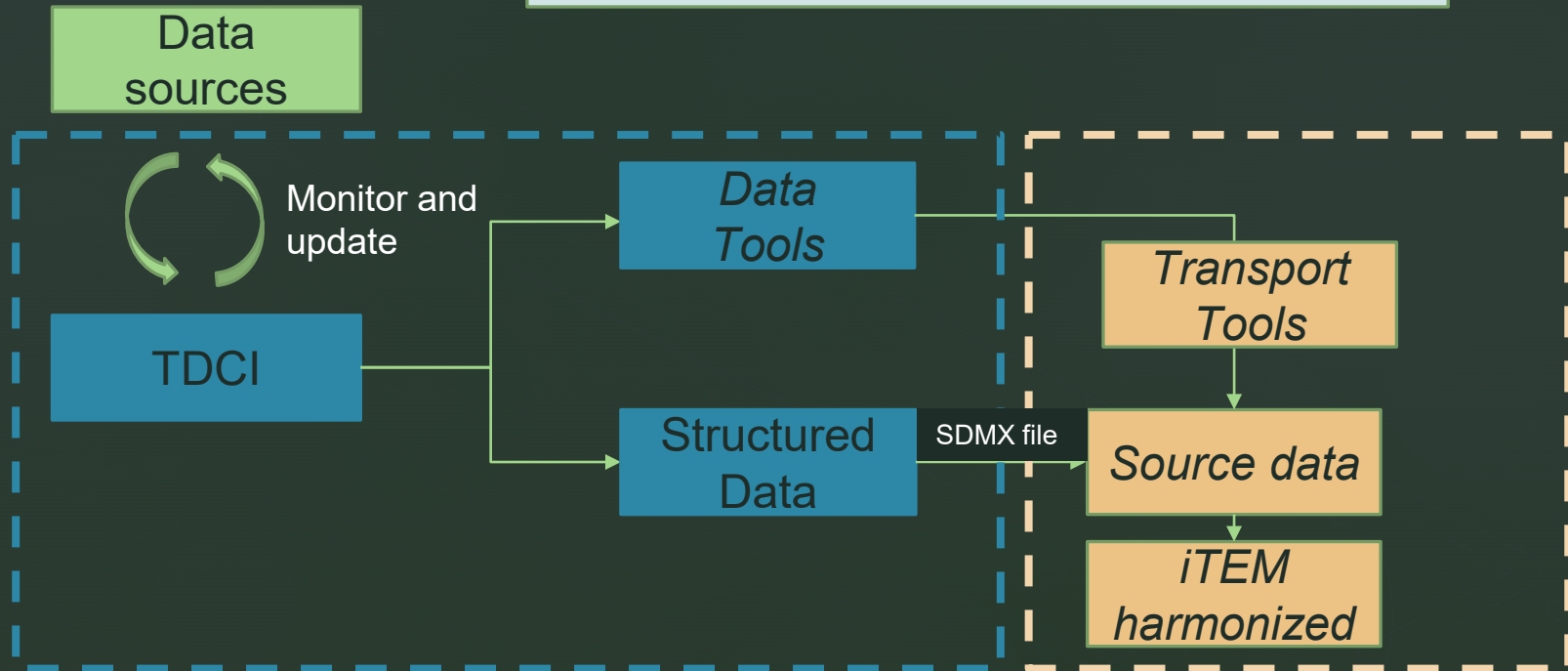
iTEM is best poised to:

- 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

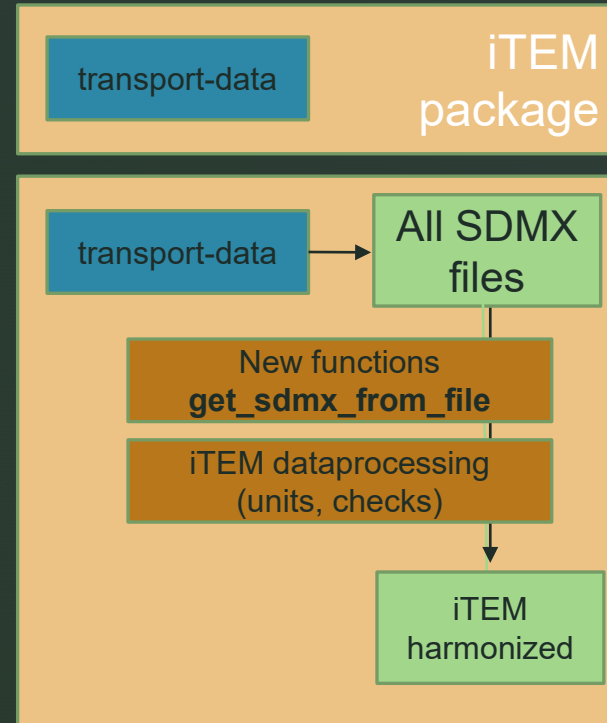


New 13 datasets from the ATO collection

4 adult education students from
disadvantaged background

13 different scripts (T011-T024)

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



<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, Humberto¹; Yeh, Sonia¹ ; Kishimoto, Paul² ; Cazzola, Pierpaolo³; Fulton, Lewis⁴;
 McCollum, David⁵; Miller, Joshua⁶; Kyle, Page; Pérez Bravo, Manuel⁷

[Show affiliations](#)2K
VIEWS1K
DOWNLOADS[▶ Show more details](#)

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