iTEM Open Data & a "Transport Data Commons"

Sonya Yeh, Chalmers University of Technology, SE
Paul Kishimoto, International Institute for Applied Systems Analysis (IIASA), AT

8th ITF Transport Statistics Meeting
§9 Towards an Inclusive Transport Statistics Community Globally
19–20 September 2022 @ OECD, Paris

iTEM = International Transport Energy Modeling

- An open group of people and organizations interested in the role of energy in the world's transport system — **https://transportenergy.org**
- Shared goal: to better understand the methods and data that are employed to study this system—especially, models with international or global scope—and through dialogue to improve knowledge of the system, its ongoing evolution, and the policy and technology options for guiding its changes.

Activities:

- Workshops
- Model intercomparison of projections
- **Open data** ← focus for this session

iTEM Open Data — vision & aims

Problems

- Transportation data is available from many sources, each incomplete and inconsistent.
- Differences in data collection, processing, or interpretation; some not available for scrutiny.
- Currently available datasets are too expensive for most researchers.

Consequences

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

Goals

- Create a transparent, open database of historical transportation statistics for iTEM & other users.
- **Open data**: a comprehensive collection of publicly-available transportation data that is itself public.
- **Open code**: code publicly accessible, transparent, documented, open for modification & extension.

One example of the recurring problems

- SY & PNK were co-authors on IPCC Sixth Assessment Report (AR6) WG III Chapter 10: Transport.
- Assessment authors needed historical data for the following basic concepts, in order to benchmark & compare future projections from global models:

  §10.7.1 Direct CO2 emissions from global (and regional) transport scenarios
  §10.7.2 Transport demand/activity trajectories for passenger and freight
  §10.7.3 Mode shares of passenger and freight transport
  §10.7.4 Energy intensity of freight and passenger transport
  §10.7.5 Fuel shares for freight and passenger transport

- Although these are fundamental concepts for transport, **there is no easy way to get these data** —they essentially don't exist in the public space!

Evolved thinking about processes for high-quality data

Learning from other scientific disciplines (e.g. climate science) and professionals (e.g. NSOs) that handle more data, frequently, with high quality standards & scientific validity:

- **Automate** data work tasks that are still, often done in costly, manual ways:
  - Fetch/collect 'upstream' data from sources via known APIs, URLs, and formats.
  - Check for the presence of known errors; apply corrections.
  - Transform to common formats, labels, and identifiers.
  - (Dis)aggregate and derive quantities, fill gaps, etc.
- **Harmonize**
  - Identify the dimensions, categorizations, etc. of input data.
  - Apply transparent, configurable transformations to common structure.
- **Diagnose**
  - Sanity checks and coverage metrics to help iTEM researchers/other data workers improve quality.

iTEM Open Data harmonized transport database

- A harmonized transport database produced in 2020-10 from the first prototype of this process:
  - https://doi.org/10.5281/zenodo.4121179 —citable, **open data**
- Assembled from 15 existing public data sets.

Coverage:

- Annual resolution 1970–2018
- 240 geographical units (countries and subdivisions)
- ~100k observations

Open code: https://github.com/transportenergy/database

- Cleaning & harmonization scripts: https://github.com/transportenergy/database/tree/main/item/historical
- Discussion of data quality with users: https://github.com/transportenergy/database/issues

Towards a “Transport Data Commons”

TDC Initiative or TDC(I) incubated by GIZ on the sidelines of the 2022-05 ITF Forum in Leipzig, DE.

- International development organizations working on low-carbon mobility face many of the same data problems as iTEM modelers & researchers!

Goals:

- Apply, improve, and expand on the process and concepts prototyped by iTEM.
- Substantial and sustainable funding for recurring data work.
- Well-maintained data infrastructure, tools, and standards (SDMX-based) flexible to all data.
- Support a broad variety of users, from researchers (e.g. iTEMs) to NGOs, international development organizations, etc. in retrieving, understanding, visualizing, and using data.
- Support would-be data providers in sharing their data via the Commons.

iTEM would be one subgroup of TDC users/providers, and operate on TDC infrastructure.

Final thoughts on “inclusion”

Incentives and rewards can push people to:

- produce **“[yet] a [nother one-off] database”** so we can claim sole/chief credit/ownership.
- be suspicious of others (‘competitors’) “free-riding” on costly (time-consuming, expensive) data cleaning work and expertise we have developed.
- …thereby exclude other researchers & stakeholders who lack similar resources.

We can and must shift this culture to instead think about **sustainable, community processes** to maintain **data as a public good**.

- Many organizations are pushing in the same direction of supporting mobility & decarbonization through knowledge derived from data & models. Those linked to iTEM/TDC include ADB/ATO, APERC, BP, ICCT, IEA, ITF-OECD, KAPSARC, UN ECE, U.S. DoE & several affiliates, U.S. EPA, SLoCaT, WB incl. SuM4All etc. —plus 20+ academic/research entities and private firms.
- We can achieve much more through shared efforts than by separately duplicating data work & software development; proliferating databases, websites, platforms, etc.

Next steps / building the community

iTEM workshops

- Held annually; organized by ITF and hosted at OECD or other venues → host a meeting.
- Topics & themes always include data → suggest presenters; organize a session.
- Contact mail@transportenergy.org.

Transport Data Commons Initiative

- at GIZ, contact Kirsten Orschulok (kirsten.orschulok@giz.de), Verena Knöll (verena.knoell@giz.de), and Daniel Bongardt (daniel.bongardt@giz.de).
- per iTEM's participation, contact mail@paul.kishimoto.name.

Share with us:

- What are your data needs, and how have you been meeting them thus far?
- What kind of contributions can you make to community processes?