The state of open data in Czech public transport

David Koňařík

FOSDEM 2026

About me

Currently a computer science student at MFF CUNI in Prague.
Lifelong fan of trains (and other public transport).

I've been working on making *usable* public transport data for the whole Czech Republic since 2018.

https://dvdkon.ggu.cz, dvdkon@konarici.cz

Talk outline

- Czech public transport crash course
- Rough history of central timetable database
- What data is available today
- Issues with international formats
- FLOSS projects and hopes for the future

Czech public transport

Trains, underground (“metro”) trains, trams, buses, trolleybuses, ferries, funiculars…

Ran by either private or public agencies.
Mostly ordered by a public institution: a city, a region, or the transport ministry.

Multiple possible tariffs:

- Agency’s own
- City-specific
- Region-specific
- OneTicket

What data *exists*?

- Timetables
  - Regular
  - Diversions
- Real-time positions, delays
- Train composition, used vehicles
- Reasons for disruptions
- Positions of stops
- Trajectories of lines

Whether and how this data is available to the public is another matter. The data being held by public institutions is basis for requesting them under freedom of information law.

The birth of CIS JŘ

The Ministry of Transport runs CIS JŘ, a central database of timetables for the whole country[1].

The first system was deployed in 1998.
The contractor offered to run it *for free*, then got into financial trouble and wanted to cancel the contract.

Another company, CHAPS, offered to run CIS JŘ.
Still for free, but they claimed all the data as their own.
They use the data as basis for their commercial trip planner, IDOS.

[1] Only inter-city buses first, local city transport and trains (incl. trains and *trolleybuses*) were added later

Fight for timetable data

In 2011, Seznam[1] started a legal fight with CHAPS for timetable data based on the freedom of information law.

This resulted in multiple requests and lawsuits at every level of the Czech judicial system, resulting in novel judicial opinions.

In the meantime, the Ministry changed their regulation to make releasing timetable data compulsory and started paying CHAPS for running CIS JŘ.

Through multiple requests and lawsuits, Seznam managed to force CHAPS to release the timetable data in 2015, except…

[1] A large Czech web services company, offering web search, news, email hosting, ads, Mapy.com…

Fight for timetable data, round 2

…the data was released per-line and stops weren't uniquely identifiable.

Seznam (presumably) worked around this by mapping every stop and getting some timetables directly from transport agencies.

I decided to continue in trying to get the full data.
Many years later, after tens of official documents sent and two parallel lawsuits in the same thing, got *slightly better* data.

Current state of CIS JŘ data

I managed to get the Ministry of Transport to order CHAPS to release data with uniquely identified stops.
They only release *new* timetables like this.

In the meantime I managed to implement a workaround for the old incomplete data.

This uncovered other problems with the data, like missing detour timetables and remaining long-cancelled lines.

It's a good base, but not the 100% correct database it should be by law.

Compulsory release of NeTeX

The European Commission has mandated the release of a bunch of transport data, including timetables in NeTeX.

The deadline (December 2023) has long passed.
Some countries provide usable data, some provide nothing.
Look if your country does this!

The Czech Republic provides *broken* data – it passes XSD validation, but is semantically broken and doesn't follow the spec.

Other timetable sources

Some public institutions publish their own timetables (typically in GTFS):

- PID – Prague + Central Bohemia region
- IDS JMK – South Moravia region
- DÚK – Ústecký region
- Liberec, Jablonec nad Nisou
- Olomouc

and some smaller institutions.

Other open data

GTFS timetables contain stop positions,
regions generally publish separate datasets as well (some open).

Real-time positions of vehicles are available for all trains and most government-paid trips.
Only ~three systems provide open APIs, some even provide details on the specific vehicle.


98 → SPOJOVACÍ
Aktuální zpoždění 1 min.
Na základě poslední zprávy z vozidla před 98 s. (00:34:09)
Výchozí zastávka: Sídliště Řepy 00:13
Poslední zastávka: Újezd 00:34:09
Příští zastávka: Národní divadlo 00:36:09
Dopravce: Dopravní podnik hl. m. Prahy
Typ vozu: ČKD Tatra T3R.P
Vůz, oběh: #8465 na 98/52
Informace o vozidle:
ECO
Karta a fotografie vozidla
Jízdní řád
JOSEFOV
STARÉ MĚSTO
Národní divadlo
Národní třída
Lazarská
Lazarská
Vodičkov
NOVÉ MĚS
97
97
98
Zastávky
Úseky se zpožděním 1
linky, pořadí, ev. čísla...
Filtrovat
?
Čas
© Pražská integrovaná doprava. © OpenStreetMap contributors. © CARTO.

State of railway data

Sadly the railway manager, Správa železnic, is not open-data friendly. They release hardly any and they deny information requests based on spurious reasons, they added (laughable) anti-scraping measures. They *do* publish timetables in a custom XML format, as required by law[1].

They have two web pages with real-time passenger train positions, but refuse to provide a proper API.

They also keep, but don't publish in any way:

- historical delay data
- reasons for delays

[1] But I have reports that some of the data isn't correct and SŽ doesn't care to force the agencies to fix it

State of railway data

- passenger train composition
- data for freight and other trains

Formats for timetable data

Some datasets are straightforward enough (stop positions, delays…), timetables are nuanced and hard to capture.

- **JDF** - Native Czech format for non-rail transport. Basically a CSV transcription of the law from 1998.
- **GTFS** - International standard originally by Google.
  Similar in structure to JDF, but can't express e.g. multiple tariff zones per stop, stops without positions, forbidden travel within a group of stops…
- **NeTeX** - European standard. Very broad "standard for making standards". Nobody supports the full format, but theoretically covers everything.

Formats for timetable data

Most FLOSS projects support GTFS, some support a variant of NeTeX. This is an issue for Czech data, since important information is either mangled or lost.

A route planner might not need zone data, but will find invalid trips if it doesn't honour restrictions.

NeTeX *seems* like the solution, but it's complex and the most common subsets don't cover our needs anyway.

Existing FLOSS projects

Many past attempts, most have failed to reach maturity.
Current relevant projects:

- **JrUtil** – My project, produces mostly-correct JDF and GTFS timetables from CIS JŘ data.
- **Transitous** – International project, has Czech GTFS data partly based on my own, partly on official sources.
- **Spojenka** – Promising new “full-stack” route planner focusing on cost optimisation.
  Doesn’t attempt to go beyond Czechia for now, but works very well with the data it has.

Existing FLOSS projects

- **JrUtil RtCollect/RtView** – Part of JrUtil, scraper & viewer of delay data, currently set up for Czech trains.

From: 12 / 01 / 2025 To: 01 / 31 / 2026 Set

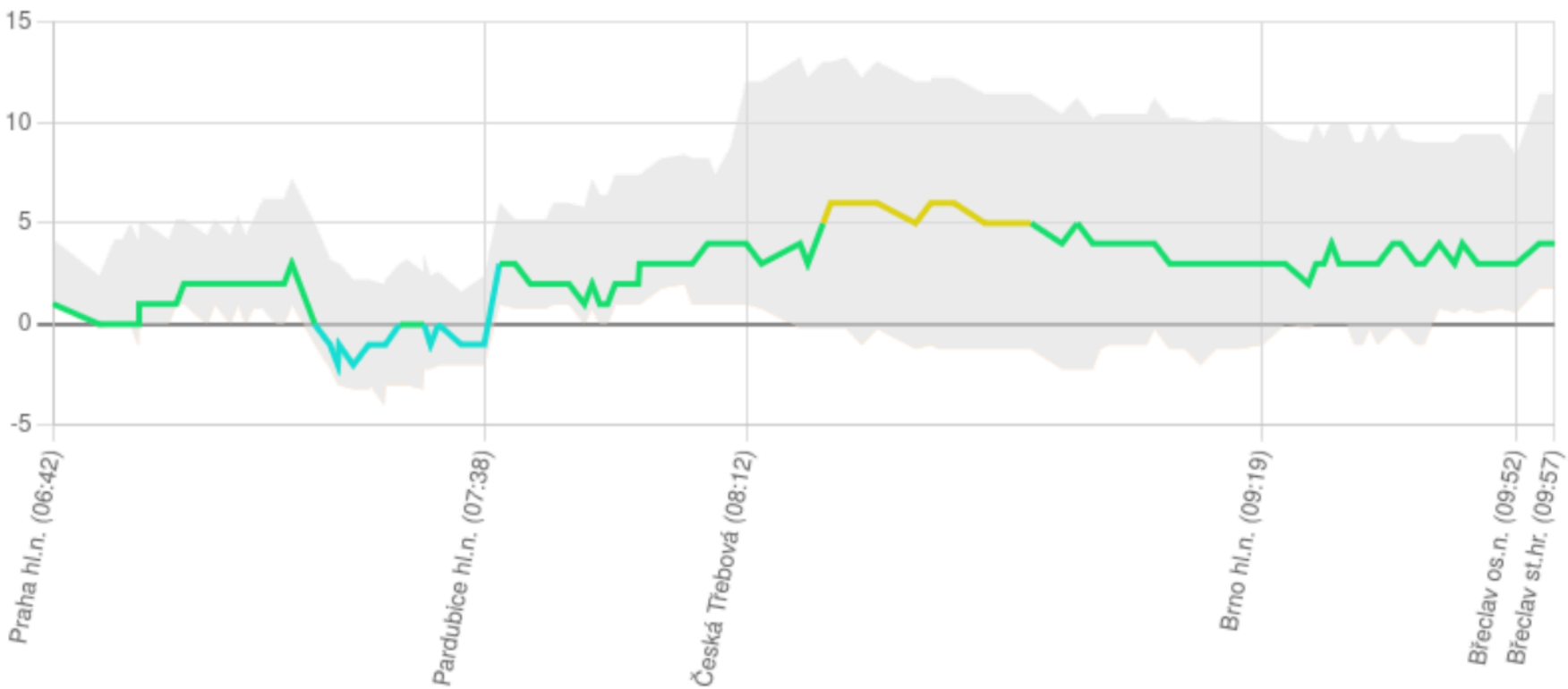
Chart type: Median and percentile bounds

RtView at rt.jr.ggu.cz

Show table

Trip rj 73 Vindobona (SSG 2025-12-01)

▸ Trips in SSG



Further reading

All in Czech:

- I run a forum on this topic, with a useful list of known data sources (open or otherwise).
- I gave a talk on the history of CIS JŘ at OpenAlt 2023.
- I also gave a talk on train statistics based on RtCollect data.
- Spojenka's developer publishes a summary of data used and an open data tierlist.

Wishes for the future

To Czech public institutions: *Please stop fighting our information requests; it's cheaper to just publish the data!*

To Czech transport data enthusiasts: *We've achieved a lot by solo work, but getting to 99.9…% reliability needs more than on person. Let's get together and form a common plan!*

To international projects and consuming software: *Bare GTFS is not enough. Please support either GTFS extensions or a NeTeX profile that supports the various countries' specifics.*