It is Not the Standard, It is the Feed: Analysis of a Minimum Viable Denominator for Comparing GTFS and NeTEx in European Public Transport Data

Livia Mihaela Miclaus[1], Shahrom Sohi[1]

[1]*Vienna University of Economics and Business, Institute of Data, Process and Knowledge Management, Welthandelsplatz 1, 1020 Vienna, Austria*

Abstract

Cross-border journey planning in Europe requires public transport data that can be reliably integrated across national and operator boundaries. Although the General Transit Feed Specification (GTFS) and Network and Timetable Exchange (NeTEx) represent many of the same concepts and share the Transmodel reference framework, schema alignment alone does not ensure that published feeds are comparable. This study introduces a Minimum Viable Denominator (MVD), a regulation-aligned comparison framework based on three fundamental dimensions: stops and stations, routes and lines, and service calendars. The framework is applied to published GTFS and NeTEx feeds from fifteen European countries. Records are matched using country-specific combinations of identifiers, normalized labels, geographic coordinates, and active-day patterns, and evaluated through two-sided coverage rates. GTFS data were available for all fifteen countries, whereas limitations in NeTEx availability and structure restricted stop, route, and calendar comparisons to ten, nine, and six countries, respectively. The results reveal a consistent hierarchy: stops are the most comparable dimension, routes are less consistent, and calendars show the greatest divergence. France, Switzerland, Luxembourg and Norway achieve high stop and route coverage, while Austria, Finland, Sweden and the Netherlands are strongly asymmetric, and Italy and the Czech Republic show weak or absent overlap, including no route-label matches at all. These differences arise primarily from fragmented publication, inconsistent identifiers, semantic repurposing, provenance gaps, and differences in scope and granularity rather than from the expressive limitations of the standards. The findings show that common standards are necessary but not sufficient for the published data to be comparable in practice. The MVD provides a repeatable instrument for monitoring comparability and highlights the need for persistent cross-standard identifiers, traceable provenance, and stronger validation of published feeds.

Keywords

GTFS, Public Transport Data Interoperability, NeTEx, Cross-Border Mobility, Mobility Data Governance

Introduction and Objectives

Cross-border and multimodal journeys are central to the European Union's transport and climate agenda, yet from a passenger's perspective they remain markedly harder to plan and book than domestic ones [1, 2]. Behind every journey planner, ticketing app, and multimodal information service lies a more basic requirement: the underlying public-transport data; where vehicles stop, which lines serve them, and when they run. This must be discoverable, exchangeable, and combinable across national and operator boundaries. Where that data cannot be reliably joined, the seams between national systems resurface as gaps in information and, ultimately, as friction for the traveler creating potential influence on passenger modal choice [3].

Two open data standards dominate the exchange of scheduled public-transport information in Europe. GTFS (General Transit Feed Specification) is a lightweight, tabular format that originated in the mid-2000s from a collaboration between Google and a public transit agency, and has since become the de facto global standard for feeding journey planners [4]. NeTEx (Network and Timetable Exchange, CEN/TS 16614) is a rich, modular XML standard developed within the European standardisation system to capture network, timetable, and fare information in far greater detail [5]. Both are traceable to the same conceptual foundation of the Transmodel reference data model (EN 12896) [6]. European law

Semantics – 22nd International Conference on Semantic Systems
livia-mihaela.miclaus1@s.wu.ac.at (L. M. Miclaus); shahrom.sohi@wu.ac.at (S. Sohi)
0009-0000-5073-7279 (L. M. Miclaus)

reinforces this convergence: Delegated Regulation (EU) 2017/1926 obliges Member States to make such data available in a standardised format through National Access Points [7], a requirement extended to dynamic data by its 2024 amendment [8].

Taken together, these facts invite a tempting inference: if two standards describe the same concepts, are rooted in the same reference model, are explicitly mapped to one another, and are published under a common legal mandate, then data expressed in one should be readily comparable and combinable with data expressed in the other. This paper argues that the inference does not always hold. Shared conceptual coverage is necessary but not sufficient for comparability. Even where GTFS and NeTEx represent exactly the same real-world object such as a station, a line, a service calendar, the published data are frequently not directly comparable, because comparability depends not only on what a standard can express but on how, and how consistently, data are actually produced and published.

The obstacles we observe are operational rather than representational. Provenance is scattered: the same journey may be described by an authority, an operator, and an aggregator, with no single authoritative source. Identifiers are reused or non-unique, so that an apparent match between two records cannot be trusted. Publishing practice diverges: a field that is populated and reliable in one country is empty, differently encoded, or semantically repurposed in another. The result is that a governance framework designed to produce interoperable data does not, by itself, guarantee data that can be interoperably used. This mirrors a broader pattern in European transport data, in which the technical means to integrate exist but the binding constraints are organisational and institutional [9].

To make this argument precise and measurable, we introduce a Minimum Viable Denominator (MVD): the smallest set of concepts that both GTFS and NeTEx represent and that together suffice to compare published datasets in practice:

- stop/station information,
- route/line information,
- calendar information.

These three dimensions coincide with the minimum data categories required by Level of Service 1 of Regulation 2017/1926, and they answer the three questions any journey planner must resolve before any other: where a passenger can board, where the service goes, and when it runs. Using the MVD as an instrument, we compare published GTFS and NeTEx feeds across fifteen European countries, ten of which could be compared in both formats, and classify, dimension by dimension and country by country, whether and how well the two can be matched.

This paper makes four contributions:

- We define the MVD as a minimal, regulation-aligned basis for comparing public-transport data across structurally different standards, reducing two incompatible formats to three comparable dimensions.
- We conduct a cross-country empirical comparison of published GTFS and NeTEx feeds in fifteen European countries, documenting the concrete matching technique each country required.
- We provide evidence that the decisive barrier to comparing published feeds is operational (provenance, identifiers, publishing practice) rather than a limitation of the standards themselves.

Accordingly, the study is guided by one research question: to what extent are the GTFS and NeTEx feeds that European countries actually publish comparable at the level of stops, routes and calendars, and what accounts for the differences observed? The first part is addressed by the availability findings reported with the dataset table and by the comparison results; the second by the discussion of the barriers observed.

Background and Related Work

GTFS represents a public-transport network as a set of flat, comma-separated tables. In its basic form these include `stops.txt`, `routes.txt`, `trips.txt`, `calendar.txt`, `calendar_dates.txt`, stop_times.txt, and shapes.txt [4, 10]. Its design priorities are simplicity and a low barrier to adoption, which made it easy for agencies to publish and for developers to consume; it is now maintained as a community specification and underpins a large share of the world's journey planners. This includes more than 6000 feeds over 99 countries [11]. These same priorities are also its limits: the format is deliberately shallow and leaves many fields optional. Because much is optional and little is enforced, two conformant feeds can differ widely in completeness and in how identifiers are assigned and reused.

NeTEx (CEN/TS 16614) is a modular XML standard organised into frames, each covering a functional area of public transport such as stops, lines, timetables, or fares [5]. It is an implementation and refinement of the Transmodel reference model, and is complemented by SIRI for real-time information, OpRA for exchange format for operational raw data and OJP, an Open API for distributed journey planning [6, 12]. Its published parts progressively cover network topology, timetables, fares, European Passenger Information Profile, Alternative modes exchange format and a European Passenger Information Accessibility Profile [13]. NeTEx is more expressive and precise than GTFS and is aligned with European regulation by design; the cost of that expressiveness is complexity, heavier tooling, and implementation that is partial and profile-dependent, so that what a national feed actually populates can diverge substantially from what the standard permits.

Both standards trace back to Transmodel (EN 12896), the European reference data model for public transport [6], and the Data4PT initiative has published an explicit mapping between GTFS files and NeTEx elements [10]; ontology-engineering efforts have likewise aligned public-transport vocabularies with Transmodel [14]. This common root is precisely what motivates the assumption we test: if two standards are both Transmodel-derived and are mapped element by element, one would expect data expressed in either to be directly comparable. Much existing tooling reinforces this expectation by treating the problem as one of format conversion. We revisit whether the expectation holds once the object of comparison is not the schemas but the feeds that countries actually publish.

The Delegated Regulation (EU) 2017/1926, supplementing the ITS Directive 2010/40/EU, requires Member States to make static travel and traffic data accessible in a standardised format through National Access Points (NAPs), organised into three ascending Levels of Service [7]; its 2024 amendment (Regulation (EU) 2024/490) extends these obligations to dynamic data [8]. In parallel, the Commission is building a Common European Mobility Data Space (EMDS) [15], now entering deployment through initiatives such as deployEMDS [16]. This emphasises the governance backdrop about the supply of standardised, discoverable data and how these data published are actually interoperable in use.

Research on public-transport data interoperability in Europe has followed interoperability itself by transformation: it uses Semantic Web technologies to lift heterogeneous feeds into a shared model and to convert authoritative data into EU-compliant NeTEx and SIRI, assembling multimodal transport knowledge graphs that can be queried uniformly [17], and it engineers public-transport ontologies explicitly aligned with Transmodel [14]. Beyond the exchange formats themselves, recent work targets the discoverability and governance layers of this ecosystem: mobilityDCAT-AP standardises metadata so that mobility datasets are findable across National Access Points and borders [18], while governance models such as TANGENT address how mobility-data sharing is organised and trusted between parties. Here interoperability is an output: given source data and a target schema, a pipeline yields integrated, standard-conformant data [17].

Methods

Data

Fifteen European countries were analysed using independently sourced GTFS and NeTEx feeds, prioritising official National Access Points, public-data portals, and operator sources. GTFS was available for all fifteen countries, while NeTEx availability was more limited: Poland, Romania, and Spain publish no NeTEx; Belgium's feed could not be obtained and Germany's fragmented per-line publication did not support a country-wide comparison. Ten countries therefore entered the cross-standard comparison.

Table 1 summarises the datasets used, including source, publisher, scope, validity, access date, and exclusions. Full file names, licence details, and access URLs are provided in the accompanying repository [19].

Table 1: GTFS and NeTEx feeds analysed. Where the two formats differ, values are prefixed **G** (GTFS) and **N** (NeTEx); otherwise the value applies to both. "Acc." is the download date. Validity is the window declared by the feed itself, reported as declared. Full file names, licences and access URLs are in the accompanying repository [19].

Country	**Source / version**	**Publisher (provenance)**	**Scope / modes**	**Declared validity**	**Acc.**	**Dimensions compared / exclusions**
Both formats available and compared						
Austria	G: data.oebb.at N: data.mobilitaetsverbuende.at ver. Fahrplan 2026 / EVU 2026	ÖBB (operator); NeTEx via the Mobilitätsverbünde portal	National; rail	2025-12-14 – 2026-12-12	2026-03-04/05	Stops, routes, calendar
Switzerland	opentransportdata.swiss ver. G 2026-04-01 / N 2026-03-31	BAV (NAP)	National; all PT	2025-12-14 – 2026-12-12	2026-04-08	Stops, routes
France	transport.data.gouv.fr ver. SNCF open-data export, 2026-04	SNCF Voyageurs (operator via NAP)	National; rail	G: 2026-04-26 – 2026-10-31 N: 2026-04-26 – 2026-09-23	2026-04-27	Stops, routes, calendar
Luxembourg	data.public.lu ver. 20260408–20260430	Verkéiersverbond (authority)	National; all PT	2026-04-08 – 2026-04-30	2026-04-22	Stops, routes, calendar
Netherlands	G: gtfs.ovapi.nl N: data.ndovloket.nl ver. G NL-20260505 / N epiap 2026-05-05	DOVA (authority); GTFS via OVapi	National; all PT	G: 2026-05-04 – 2026-12-12 N: from 2026-05-05	2026-05-05	Stops only; NeTEx contains no lines or calendars
Finland	fintraffic.fi ver. static export, 2026-05	Fintraffic (NAP aggregator)	National; all PT	2014-07-01 – 2080-12-31 (declared; identical in both formats)	2026-05-12	Stops, routes
Norway	developer.entur.org ver. national aggregate, 2026-05	Entur AS (NAP)	National; all PT	G: 2019-04-13 – 2027-03-31 N: 2019-04-13 – 2031-12-31 (declared)	2026-05-08	Stops, routes, calendar; shared `NSR:StopPlace` identifiers
Sweden	trafiklab.se ver. daily build, 2026-05-11	Trafiklab / Samtrafiken (NAP)	National; all PT; NeTEx covers only part of the operators	Updated daily	2026-05-11	Stops, routes, calendar; API key required
Italy	G: dati.toscana.it N: cciss.it ver. G trenitalia.gtfs / N Trenitalia L1	Trenitalia (operator); NeTEx via NAP	G: **Tuscany only**; rail N: **national**; rail	G: 2026-01-15 – 2026-12-12 N: from 2025-12-14	2026-05-16/17	Stops, routes, calendar; NeTEx restricted to Tuscany before stop matching

Country	Source / version	Publisher (provenance)	Scope / modes	Declared validity	Acc.	Dimensions compared / exclusions
Czech Rep.	G: spojenka.cz N: portal.cisjr.cz ver. G daily build / N GVD 2026, 2026-05	G: Spojenka (community aggregator) N: CIS JŘ / Ministry of Transport	G: countrywide; **rail and integrated transport** N: **rail only**	G: 2023-01-27 – 2030-09-29 (declared) N: GVD 2026	2026-05-18	Stops, routes; no calendar comparison possible
Not carried through to the cross-standard comparison						
Germany	opendata-oepnv.de ver. 20260413 export	DELFI e.V. (national aggregator)	National; all PT	2026-03-28 – 2026-12-12	2026-04-13	None: NeTEx distributed as a large number of per-line files rather than one national dataset; sample examined, indicative only
Belgium	G: gtfs.irail.be N: belgiantrain.be ver. G 2026-05-31	G: iRail (community, from SNCB/NMBS) N: SNCB (operator)	National; rail	G: 2026-03-30 – 2026-12-12	2026-05-31	None: NeTEx published but not obtained; licence request unanswered
Spain	nap.transportes.gob.es ver. 20260517	RENFE Operadora (operator via NAP)	National; rail (MD, LD, AVE)	2026-05-16 – 2026-11-03	2026-05-18	None: no NeTEx published
Poland	mkuran.pl/gtfs ver. 2026-07 build	M. Kuranowski (community); schedules from PKP PLK	National; rail, 13 operators	2025-12-14 – 2026-08-29	2026-07-13	None: no NeTEx published
Romania	github.com/vasile/data.gov.ro-gtfs-exporter ver. master branch	Community conversion of CFR Călători data	National; rail	2025-12-14 – 2026-12-12	2026-05-18	None: no NeTEx published

The Minimum Viable Denominator

GTFS and NeTEx differ in structure and level of detail, so their published feeds cannot be compared directly. We therefore define a *Minimum Viable Denominator* (MVD) based on three concepts represented in both standards: stops/stations, routes/lines, and service calendars.

The MVD measures *cross-standard comparability*: whether corresponding information in independently published GTFS and NeTEx feeds can be related using the information available in the feeds. Dual publication is used as the evaluation setting because, where both formats cover the same transport context, their agreement can be tested directly. It is not assumed to be the only way to achieve interoperability. The MVD is therefore distinct from data availability and coverage and should be understood as one part of a broader monitoring framework that may also consider freshness, identifier consistency, provenance, and transformation fidelity.

The three dimensions correspond to the minimum data categories of Level of Service 1 of Regulation 2017/1926 [7] and represent the basic information needed to determine where a service stops, where it goes, and when it operates. Detailed mappings and country-specific extraction procedures are provided in the accompanying repository [10, 19].

Matching

Matching varied by country because publishers populate different fields. Stops were matched using shared or cleaned identifiers where possible, and otherwise using normalised names or geographic coordinates. Country-specific fields and preprocessing rules are documented in the accompanying repository [19].

Routes were compared at the level of unique public labels rather than technical rows, since both standards may create several records for different directions, operators, or journey patterns of the same passenger-facing line. Where the published labels could not be meaningfully aligned, as in Italy and the Czech Republic, no valid route-label comparison was possible.

Calendars were converted on both sides into active-day patterns over their shared validity window. Each day was encoded as active or inactive, and two patterns were considered equal only when their complete sequences matched. Detailed extraction paths and country-specific matching procedures are provided in the repository [19].

Procedures and validation

For proximity-based stop matching, thresholds were 100 m in Italy and Sweden and 50 m in France and the Netherlands. These thresholds are conventions rather than derived values; sensitivity tests in Sweden from 10 m to 1000 m did not change the overall conclusion. Matching was many-to-one, with ties occurring only for duplicate coordinates. Normalisation was deliberately light: stop names were lowercased, trimmed and reduced to single spaces, with accents and punctuation removed where needed; identifiers were reduced to their numeric core when prefixes differed; and route labels were trimmed, with country-specific uppercasing or suffix removal where required. NeTEx records were deduplicated by identifier, while calendar patterns were compared as unique sets. Full country-specific rules are documented in the repository.

Matching methods were also validated using independent evidence. Identifier and name matches were checked geographically, while matched line labels were compared using their served-stop sets. These checks support identifier- and name-based matching but show that label and calendar-pattern matching require more caution: a shared line label does not necessarily imply the same itinerary, and an identical calendar pattern indicates the same active-day sequence rather than the same service.

Main Findings

Results are reported per country and per dimension in Table 2. Each comparison yields two rates rather than one, because the two feeds rarely contain the same number of records: the GTFS-side rate expresses matched units against the GTFS total, the NeTEx-side rate against the NeTEx total. Reporting both is necessary, since a feed can be fully contained in a larger counterpart without the two disagreeing about anything. Proportional overlap diagrams for each country and dimension are available in the accompanying repository [19].

Stop information was compared for ten countries and is the most consistent dimension. France, Switzerland, Luxembourg and Norway reach near-total agreement on both sides, and in each of these the two feeds are published by the same organisation. Finland, Sweden, Austria and the Netherlands are strongly asymmetric: GTFS-side rates above 92% against NeTEx-side rates between 57% and 70%, reflecting NeTEx stop sets that are substantially larger than their GTFS counterparts. The two lowest results have identifiable causes in scope rather than in publication practice alone. The Czech GTFS feed covers rail together with integrated regional transport while its NeTEx feed covers rail only, so the 2.27% GTFS-side rate is largely a consequence of comparing a broad feed against a narrow one. The Italian GTFS feed covers Tuscany while its NeTEx feed is national; restricting the NeTEx set to the GTFS area before matching mitigates but does not remove this asymmetry.

Country	Matched	GTFS	NeTEx	GTFS %	NeTEx %
Stops / stations					
France	3,506	3,506	3,506	100.00	100.00
Switzerland	35,256	35,267	35,324	99.97	99.81
Luxembourg	2,796	2,796	2,798	100.00	99.93
Norway	51,529	51,529	57,957	100.00	88.91
Finland	88,640	89,547	137,227	98.99	69.58
Sweden	48,137	48,778	71,650	98.69	67.18
Austria	1,084	1,119	1,669	96.96	65.01
Netherlands	14,536	15,658	25,380	92.80	57.27
Italy	276	313	1,040	88.18	26.54
Czech Rep.	771	34,006	2,209	2.27	34.90
Routes / lines					
Switzerland	1,789	1,789	1,789	100.00	100.00
France	728	728	728	100.00	100.00
Luxembourg	690	690	690	100.00	100.00
Norway	2,460	2,460	2,461	100.00	99.96
Austria	167	173	227	96.53	73.57
Sweden	1,580	2,904	3,431	54.41	46.05
Finland	1,865	3,512	1,875	53.10	99.47
Italy	0	14	15	0.00	0.00
Czech Rep.	0	447	2,427	0.00	0.00
Service calendars					
France	11,203	11,203	11,427	100.00	98.04
Luxembourg	81	100	115	81.00	70.43
Sweden	4,027	6,892	7,398	58.43	54.43
Norway	2,789	4,791	3,057	58.21	91.23
Austria	764	1,338	4,111	57.10	18.58
Italy	44	87	5,431	50.57	0.81

Table 2: Comparison results by country and dimension. “Matched” counts units present on both sides; the two rates express this against the GTFS and NeTEx totals respectively. Countries are ordered by GTFS-side rate within each dimension.

Route information was compared for nine countries and is less consistent. Switzerland, France, Luxembourg and Norway again reach complete or near-complete agreement, and Austria reaches 96.53% on the GTFS side. Finland and Sweden fall to roughly 53% and 54%, but for different reasons: in Finland the NeTEx side reaches 99.47%, indicating that the NeTEx label set is essentially contained in a larger GTFS set inflated by descriptive fallback labels, whereas Sweden is partially asymmetric on both sides. Italy and the Czech Republic return no matches at all, because the label fields hold different kinds of information in the two formats rather than because the underlying lines differ.

Calendar information was compared for six countries and is the weakest and most asymmetric dimension. Only France approaches agreement on both sides. Luxembourg, Sweden and Norway fall between 54% and 91%, while Austria and Italy show the largest gaps in the study, with NeTEx-side rates of 18.58% and 0.81% against GTFS-side rates near 57% and 51%.

These asymmetries carry information that a single symmetric score would conceal. In Austria, a GTFS stop rate of 96.96% against a NeTEx rate of 65.01% indicates that the GTFS stop set is largely contained within a richer NeTEx set, not that the two disagree. In Finland, a GTFS route rate of 53.10% against a NeTEx rate of 99.47% indicates that GTFS over-splits lines relative to NeTEx. The direction of the imbalance is therefore as informative as its magnitude.

Two qualifications apply throughout. The sample sizes differ by orders of magnitude, from 313 Italian stops to 89,547 Finnish ones, so percentages are not equally precise across countries. And, as established in the Methods, the three dimensions do not demonstrate the same thing: a stop match evidences a shared location, a route-label match a shared public label, and a calendar-pattern match a shared active-day sequence.

Discussion and Conclusion

Across fifteen countries and three dimensions, representational equivalence between GTFS and NeTEx rarely translated into direct comparability of the published data. Both standards are explicitly mapped to one another, yet observed match rates ranged from complete, symmetric agreement (Switzerland and France at or near 100% on stops and routes) to total failure (0% on routes in Italy and the Czech Republic). Shared conceptual coverage is therefore necessary but not sufficient: what a standard can express matters far less than how, how completely, and how consistently data are actually published.

The first barrier is not agreement but supply. A usable GTFS feed was found for all fifteen countries, but a NeTEx feed suitable for the calendar comparison could be carried through for just six. Three countries (Poland, Romania, Spain) publish no NeTEx at all, Belgium publishes it but did not release it on request, and Germany publishes it in a form that did not support a country-wide comparison. The legal mandate has produced GTFS broadly as open data but NeTEx unevenly. Where both feeds existed, comparability fell along a clear dimensional gradient: stops matched best, routes next, calendars worst. Stop sets, anchored by names and coordinates, reached GTFS-side rates above 88% in every country except the Czech Republic; route matching collapsed wherever labelling systems diverged; and calendar information which NeTEx distributes across `DayType`, `DayTypeAssignment`, `OperatingPeriod`, and `UicOperatingPeriod` was the weakest and most asymmetric dimension, with only France agreeing on both sides. The more structurally elaborate the concept, the less its published representations agreed. A second pattern is organisational rather than geographic. The four countries with near-symmetric agreement on stops and routes publish both formats through the same organisation, while the weakest results arise where the two feeds differ in scope: a countrywide GTFS feed against a rail-only NeTEx feed in the Czech Republic, and a regional GTFS feed against a national one in Italy.

Furthermore, in Austria, a GTFS stop coverage of 96.96% against a NeTEx coverage of 65.01% means GTFS is largely a subset of a richer NeTEx, not that the two disagree; in Finland, a GTFS route coverage of 53.10% against a NeTEx coverage of 99.47% means GTFS over-splits lines relative to NeTEx. A single symmetric score would have hidden these relationships.

The failures we observed are operational, not linguistic, and they recur: (i) unavailability or fragmentation of NeTEx; (ii) identifier practice, shared identifiers yielded confirmed matches (Switzerland, France), while reused or non-unique identifiers made matches untrustworthy (Czech Republic); (iii) semantic repurposing of the same field, where Italy encoded a physical route and the Czech Republic an internal train number in fields that NeTEx used for a service category, producing route matches of 0% that no cleaning could repair; and (iv) differences of scope and granularity, such as Italy's regional GTFS feed against a national NeTEx file. None of these is a limitation of GTFS or NeTEx as languages.

Hence, the obligation to publish standardised data through National Access Points [7, 8] has succeeded in producing feeds, but not consistently comparable ones. Our evidence indicates that the decisive gains lie in the operational layer: persistent, resolvable, cross-standard identifiers; documented provenance and lineage; and validation of published feeds. This is precisely the layer targeted by cross-border metadata specifications such as mobilityDCAT-AP [18] and by mobility-data governance models, and it is where the Common European Mobility Data Space [15, 16] can most directly improve the comparability we measured.

The study is a single snapshot of feeds that change over time. The Southern Europe and Central-Eastern Europe figures each rest on a single country (Italy and the Czech Republic), so they should be read as cases rather than robust regional averages. Name- and coordinate-based stop matching admits both false positives and false negatives. Finally, the MVD deliberately covers only stops, routes, and calendars; fares, accessibility, and real-time data are out of scope and may behave differently.

Overall, common data standards are necessary but not sufficient for the published data to be comparable in practice. On the feeds that Member States actually publish, GTFS–NeTEx comparability is uneven but predictable: high where identifiers are shared and data ecosystems are mature, and collapsing where data are missing, semantically repurposed, or differently scoped. This unevenness is not merely a technical inconvenience: because a cross-border journey must combine several national feeds, the weakest link governs the result, and the gaps we measured directly constrain the cross-border data integration on which passenger-facing services depend. Open data is a principal driver of innovation in this domain, and applications such as journey planners, ticketing, and Mobility-as-a-Service platforms can only reach across borders as far as the underlying data can be reliably joined; where comparability collapses, so does the reach of these services, regardless of how open or standardised each national feed is on its own. The binding constraints are therefore organisational rather than representational [9, 2], and the returns to seamless cross-border travel lie in governing identifiers, provenance, and publishing practice.

The Minimum Viable Denominator offers a minimal, regulation-aligned, and repeatable instrument for measuring that progress—one that could be re-run periodically across National Access Points to track whether investments in the European mobility data space are actually closing the comparability gap.

Acknowledgments

The contribution of this paper was supported by the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI) under the endowed professorship for "Data-Driven Knowledge Generation: Climate Action".

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT and Claude in order to: check grammar and spelling, assist with code debugging, extract and document methodological details from the analysis notebooks, and draft and revise passages of the manuscript. After using these tools, the author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] C. Fetting, The european green deal, ESDN report, December 2 (2020) 53.

[2] S. Sohi, S. Harikrishnan, G. Labanca, P. Kovac, M. Macko, B. Scepka, A. Anjomshoaa, A. Polleres, Inter-railways data sharing for seamless travelling applications in europe., in: SEMANTiCS (Posters, Demos, Workshops & Tutorials), 2025.

[3] S. Peer, F. Tomori, B. Wagner, T. Winkler, The potential impact of google maps on mode choices: Evidence from a stated preference experiment, Transportation Research Interdisciplinary Perspectives 33 (2025) 101560.

[4] MobilityData, Overview – general transit feed specification (gtfs), https://gtfs.org/documentation/overview/, n.d. Accessed: 2026-07-15.

[5] Open Transport Data Switzerland, Timetable cookbook – netex, https://opentransportdata.swiss/en/cookbook/timetable-cookbook/netex/, n.d. Accessed: 2026-06-05.

[6] CEN, Transmodel, 2024. URL: https://transmodel-cen.eu/.

[7] European Commission, Commission delegated regulation (eu) 2017/1926 of 31 may 2017 supplementing directive 2010/40/eu with regard to the provision of eu-wide multimodal travel information services, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R1926, 2017. Accessed: 2026-06-02.

[8] European Commission, Commission delegated regulation (EU) 2024/490 of 29 november 2023 amending delegated regulation (EU) 2017/1926 with regard to the provision of EU-wide multimodal travel information services, Official Journal of the European Union, 2024. URL: https://eur-lex.europa.eu/eli/reg_del/2024/490/oj/eng, accessed 2026-07-16.

[9] S. Sohi, S. Harikrishnan, A. Anjomshoaa, A. Polleres, Be open or competitive? exploring the differences in the adoption of (open) data sharing in European railway public transport, Transportation Research Interdisciplinary Perspectives 34 (2025) 101698. doi:10.1016/j.trip.2025.101698.

[10] N. Knowles, Outline Comparison and Mapping Between NeTEx and GTFS, Technical Report, Data4PT Project, European Commission DG MOVE, 2024. URL: https://data4pt-project.eu/wp-content/uploads/2024/07/2024-June_DATA4PT_GTFS-NeTEx-Mapping_vf.pdf, originally issued March 2021, updated June 2024. Accessed: 2026-06-02.

[11] MobilityData, Mobilitydatabase, 2026. URL: https://mobilitydatabase.org/.

[12] CEN, Public Transport — Service Interface for Real-Time Information (SIRI), Technical Report EN 15531 / CEN/TS 15531, European Committee for Standardization, 2015.

[13] CEN, Public Transport — Network and Timetable Exchange (NeTEx), Technical Specification CEN/TS 16614 (Parts 1–4), European Committee for Standardization, 2021. Based on Transmodel (EN 12896), IFOPT and SIRI.

[14] E. Ruckhaus, A. Anton-Bravo, M. Scrocca, O. Corcho, et al., Applying the LOT methodology to a public bus transport ontology aligned with Transmodel: Challenges and results, Semantic Web (2023). doi:10.3233/SW-210451.

[15] European Commission, Creation of a common European mobility data space, Communication COM(2023) 751 final, 2023. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2023:751:FIN, accessed 2026-07-16.

[16] deployEMDS Consortium, deployEMDS: Deployment of a common European mobility data space, Digital Europe Programme project (2023–2026), 2024. URL: https://deployemds.eu/, accessed 2026-07-16.

[17] M. Scrocca, M. Comerio, A. Carenini, I. Celino, Turning transport data to comply with eu standards while enabling a multimodal transport knowledge graph, in: International Semantic Web Conference, Springer, 2020, pp. 411–429.

[18] M. Scrocca, L. Molinas Comet, B. Witsch, D. M. Mustafa, C. Lange, M. Comerio, P. Lubrich, mobilitydcat-ap: a metadata specification for enhanced cross-border mobility data sharing, in: European Semantic Web Conference, Springer, 2025, pp. 226–243.

[19] L. M. Miclaus, Minimum viable denominator: data, code and supplementary material for the GTFS–NeTEx comparison, GitHub repository, 2026. URL: https://github.com/DataFan2/gtfs-netex-mvd-comparison, accessed: 2026-08-08.

[20] M. Scrocca, M. Grassi, M. Comerio, V. A. Carriero, T. Delgado Dias, A. Vieira Da Silva, I. Celino, Intelligent urban traffic management via semantic interoperability across multiple heterogeneous mobility data sources, in: International Semantic Web Conference, Springer, 2024, pp. 218–235.

[21] CEN, Public Transport — Reference Data Model (Transmodel), European Standard EN 12896, European Committee for Standardization, 2019.

[22] S. Sohi, Semantic technologies for harmonising European railway, in: The Semantic Web: ESWC 2025 Satellite Events, Lecture Notes in Computer Science, Springer, 2025, pp. 195–204. doi:10.1007/978-3-031-99554-5_34.

[23] S. Sohi, G. Wutz, R. Hrivnák, F. Reiter, D. Pichler, A. Anjomshoaa, A. Polleres, Enhancing rail transit accessibility: A data-centric approach to park and ride, Transportation Research Procedia (2025). doi:10.1016/j.trpro.2025.04.051.