Report on public transport data: availability and homogenisation needs

data.europa.eu
The official portal for European data

This study has been prepared as part of the European Data Portal (data.europa.eu). The European Data Portal is an initiative of the European Commission and is the official portal for European data. The Publications Office of the European Union manages the European Data Portal.

European Commission

Directorate-General for Communications Networks, Content and Technology
Unit G.1 – Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

European Data Portal
Email: info@data.europa.eu

Authors – Universidad Politécnica de Madrid, King's College London
Katherin Alexandra Corredor Páez
Dorsa Soleimani Kourani
Oscar Corcho
Elena Simperl

Last updated: August 2026

Disclaimer

The views set out in this publication are those of the authors and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use that may be made of the information contained therein.

Luxembourg: Publications Office of the European Union, 2026

ISBN 978-92-78-45252-0 doi: 10.2830/9105423 OA-01-26-072-EN-N

Contents

Table index

List of abbreviations

APIs – accessible programming interfaces
bbox – bounding box
EDP – European Data Portal
GTFS – General Transit Feed Specification
HVD – high-value datasets
ITS – intelligent transport system
NAPs – national access points
Napcore – National Access Point Coordination Organisation for Europe
NetEx – Network Timetable Exchange
NUTS – Nomenclature of territorial units for statistics
REST – representational state transfer
UUIDs – universally unique identifiers

Table of Member State country codes

Member State country codes, as shown in this table, are used extensively throughout the report.

Member State code	Member State
BE	Belgium
BG	Bulgaria
CZ	Czechia
DK	Denmark
DE	Germany
EE	Estonia
IE	Ireland
EL	Greece
ES	Spain
FR	France
HR	Croatia
IT	Italy
CY	Cyprus
LV	Latvia
LT	Lithuania
LU	Luxembourg
HU	Hungary
MT	Malta
NL	Netherlands
AT	Austria
PL	Poland
PT	Portugal
RO	Romania
SI	Slovenia
SK	Slovakia
FI	Finland
SE	Sweden

1. Executive summary

This report assesses the availability, coverage and structural consistency of public transport data in General Transit Feed Specification (GTFS) format across EU Member States, with a focus on its suitability for cross-country analysis and policy use. GTFS is an open, machine-readable data standard designed to describe scheduled public transport services and enable their use in journey-planning applications, maps, accessibility analyses and transport research. A GTFS Schedule feed consists of a set of interrelated text files containing information on transport operators, stops and stations, routes, individual trips, stop times and service calendars. It may also include optional information on route geometries, transfers, fares, accessibility and other operational characteristics.

The report finds that while GTFS has been widely adopted and provides a robust common baseline for timetable data, the overall data landscape remains fragmented, unevenly distributed and only partially interoperable. National access points (NAPs) emerge as the primary and most reliable source of GTFS data, whereas the European Data Portal (EDP) largely acts as a partial mirror or metadata layer.

Coverage across the EU is incomplete and highly imbalanced. The analytical corpus contains at least one usable GTFS feed for 18 of 27 Member States, but country representation does not necessarily imply comprehensive territorial or modal coverage. Only 11 show near-national coverage, while others remain partial or entirely absent. The corpus is strongly concentrated in a small number of Member States, particularly France, and important gaps persist in both geographic and modal dimensions, especially for rail data. These disparities limit the ability to perform consistent EU-wide analyses of public transport systems.

At the structural level, the GTFS core specification is implemented almost universally, ensuring a consistent foundation for timetable data. However, optional components such as fares, accessibility and transfer information are weakly and inconsistently adopted, and proprietary extensions are common in more advanced ecosystems. Most importantly, the absence of harmonised identifier conventions, particularly for stops, constitutes the main barrier to interoperability, preventing the straightforward integration of datasets across Member States.

The report concludes that improving EU-wide usability of public transport data requires moving beyond format standardisation towards a stronger harmonisation of identifiers, controlled vocabularies and publication practices. Key priorities include enhancing direct accessibility to GTFS feeds, promoting globally unique stop identifiers, aligning route classification schemes and strengthening the adoption of underused but policy-relevant GTFS components. Together, these steps would significantly improve the potential for cross-border integration, data-driven policymaking and the development of EU-wide mobility services.

2. Context and background

The Open Data Directive (Directive (EU) 2019/1024) establishes a comprehensive legal framework for the reuse of data and information held by public sector bodies, including the principle that research data resulting from publicly funded research must be open access by default. A central element of the directive is the concept of high-value datasets (HVDs), which is associated with important benefits for society and the economy when reused. The directive identified six thematic categories: geospatial, Earth observation and environment, meteorological, statistics, companies and company ownership, and mobility.

The subsequent implementing regulation (2023), referred to as the HVD Implementing Regulation and adopted by the European Commission, laid down a list of those concrete HVDs within the limits of those six categories and the arrangements for their publication by the Member States. More specifically, in the context of mobility, the HVD Implementing Regulation identifies the INSPIRE data theme (as defined in Annex I to Directive 2007/2/EC) of transport networks and discusses the arrangements for the publication and reuse of transport network datasets.

The Intelligent Transport System (ITS) Directive (Directive 2010/40/EU) sets the framework for the deployment of intelligent transport systems in the field of road transport and for interfaces with other modes of transport. Among the delegated acts adopted under the ITS Directive is the Delegated Regulation on Multimodal Travel Information Services (Commission Delegated Regulation (EU) 2017/1926), which establishes requirements for the provision of EU-wide multimodal travel information. Together, these instruments created the obligation for Member States to establish NAPs, dedicated portals designed to facilitate access to and the exchange and reuse of transport-related data, with the overarching goal of enabling interoperable, EU-wide travel and traffic services for end users.

Data on this subject is typically collected by transportation agencies and authorities, transmitted to open data portals, NAPs and national statistical offices, and subsequently catalogued in the EDP and reported to Eurostat. As a result, public transport datasets are now available across several overlapping catalogues (including local and regional data portals, NAPs, national open data portals and the EDP) and in a variety of formats, with GTFS being the most common, followed by Network Timetable Exchange (NeTEx), another standard for exchanging detailed public transport data, and other ad hoc formats.

This fragmented landscape of overlapping sources and heterogeneous formats points to a clear need to assess the availability, coverage and consistency of public transport data across the EU, which is the focus of this report.

A relevant use case is provided by Eurostat's experimental statistics on public transport availability and efficiency. These statistics combine open GTFS data with OpenStreetMap, NUTS (Nomenclature of territorial units for statistics) territorial classification, and population-grid data to produce regional indicators on the number of public transport stops and lines, along with the proportion of territory and population that can be reached by public transport within specified travel times compared with private cars. This initiative demonstrates the potential of GTFS data for producing comparable regional statistics without imposing additional data-collection burdens on national statistical authorities. At the same time, it highlights the importance of having accessible, geographically comprehensive and structurally consistent GTFS feeds, since gaps in territorial coverage, transport modes or data quality directly affect the comparability and completeness of the resulting indicators.

The objective of this research is to explore the availability and potential of datasets related to public transport, with a focus on GTFS and similar data sources (e.g. NeTEx), with the overall goal that this data should enable measurement of key indicators such as the service efficiency, coverage and accessibility of public transport systems.

The research questions that will be addressed are the following.

— **Q1.** What is the degree of overlap between the data available in national access points from all Member States and the European Data Portal, in terms of GTFS (and NeTEx) data?
— **Q2.** Considering both types of data sources, what is the degree of coverage of the EU territory?
— **Q3.** Considering GTFS data sources, which parts of the specification are most commonly used in the datasets that are published, and which ones are less frequent?
— **Q4.** What are the main differences in the provision of identifiers and other types of reference data in the GTFS data sources available? Is there an opportunity for a further homogenisation of values beyond the use of a common data format?

Expected added value

For open data reuse. GTFS and related datasets accessible via accessible programming interfaces (APIs) are essential for mobility applications that display real-time public transport information. While such apps function well in many countries, broader data availability and homogenisation could enable cross-border integration.

For citizens. Improved access to accurate, real-time public transport information may increase public transport usage and enhance user satisfaction.

For policymakers. Insights derived from this data could support the promotion of ecological transport modes, contribute to emission reduction targets and advance the objectives of the European Green Deal.

For local governments. Analysing these datasets can help identify service gaps and inform targeted improvements in transport accessibility and planning.

Related EU initiatives

Two EU-funded projects have previously addressed challenges closely related to those of this research, and their findings are drawn upon at relevant points in this report.

Data4PT (Data for public transport) was a Connecting Europe facility programme support action that supported the coordinated implementation of NeTEx and SIRI (Service interface for real-time information), a European standard for exchanging real-time public transport data, across NAPs in Member States. Among its outputs, the project produced a methodology for comparing transport data standards (Bourée et al., 2021), which defines a structured, three-step process for analysing relationships between data formats – from a high-level structural comparison, through concept-level mapping, to a detailed attribute-level analysis. It also introduces the concept of lexical scope to describe the uniqueness boundary of identifiers, a key concept for understanding cross-feed identifier fragmentation. This methodology and its conceptual framework directly inform the analytical design of this research.

Napcore (National Access Point Coordination Organisation for Europe) is an ongoing Connecting Europe facility programme support action, bringing together public authorities from all Member States, aimed at facilitating the EU-wide coordination of NAPs under the ITS Directive. Within its Task 2.4 on EU identifiers for multimodal access nodes, Napcore published Report D2.4.1 (Napcore, 2025), which maps how transport access nodes are currently identified across 21 Member States through structured interviews with national representatives. Its findings, documenting the wide variety of national identifier structures, the absence of harmonised systems in several Member States and the structural fragmentation of stop identifiers across operators and data providers, provide the empirical governance baseline against which the GTFS data findings of this research are contextualised.

3. Research methodology

3.1. Research scope

This study examines the availability, overlap, coverage and structural characteristics of public transport data published in GTFS Schedule (static) format across Member States. GTFS Realtime and GTFS-Flex are excluded to maintain a consistent analytical scope focused on static timetable data.

The main analytical unit is the GTFS feed, understood as a downloadable and processable GTFS package, either as a ZIP file or as an extracted folder containing GTFS text files. By contrast, datasets and distributions are treated only as metadata layers describing how feeds are catalogued and made accessible through data portals. They are not used as the basis for the analytical results.

The empirical corpus is therefore built from downloaded and validated GTFS feeds collected from NAPs and from the EDP. Metadata counts are used only as contextual information, particularly in cases where a portal lists GTFS-related records that do not resolve to downloadable or usable feeds.

Data collection from the European Data Portal

Using keyword search in the EDP and filtering by the GTFS format, 216 datasets were identified.

Accessible programming interface query

Then, we harvested the datasets' metadata, extracted candidate distributions and retained validated GTFS feeds for analysis, throughout March 2026, using a structured Python-based pipeline. We retrieved the datasets using the EDP search API endpoint:

```
https://data.europa.eu/api/hub/search/search?filters=dataset&aggregationAllFields=false&aggregationFields=format&q=gtfs&page={n}&limit=300
```

The query used *q=gtfs* as a full-text search parameter across all metadata fields (titles, descriptions, tags, format labels). This approach was deliberately broad to ensure no GTFS-relevant datasets were missed due to inconsistent format labelling. A subsequent filtering step was applied to exclude GTFS Realtime feeds and all non-GTFS formats, retaining only static schedule data.

Search of General Transit Feed Specification feeds

The EDP search API was queried using the term 'gtfs' as a full-text search across all metadata fields, returning a total of 1 369 dataset records retrieved across five paginated API calls (pages 0 to 4, limit 300 per page). All pages were combined into a single metadata file for processing.

Each dataset record was then inspected for distributions (the individual downloadable files attached to each dataset) yielding 4 215 distribution records in total. These were filtered to retain only GTFS Schedule candidates, excluding GTFS Realtime feeds, GTFS-Flex extensions and all confirmed non-GTFS formats such as PDF, JSON, XML and NeTEx, along with distributions with no download URL.

Using this procedure, a total of 172 GTFS feeds were downloaded and validated from the EDP for analysis. These feeds were then used in the source-overlap analysis with the NAP corpus.

From these countries we excluded Switzerland and the EU tag. The EU Member State tag on the EDP is being misused in two ways: as a fallback for datasets for which the Member State was not properly assigned, and for genuinely non-transport datasets that appeared in the GTFS search due to unrelated mentions of the word in their metadata.

It should be noted that although the EDP exposes data through dataset and distribution metadata, the substantive unit of analysis in this report is the static GTFS feed, understood as a downloadable GTFS package that can be opened and inspected. Dataset and distribution counts are used only to describe portal representation and metadata accessibility.

3.2. Data collection methodology in national access points

Before downloading any data, the first task was to identify which GTFS datasets were available in each source. This discovery stage consisted of determining what resources existed, where they were published and whether they corresponded to the target of this study: static GTFS timetable feeds rather than real-time services or other transport data formats. Because there is no single central repository for GTFS data in Europe, discovery had to be adapted to the publication model of each NAP and of the EDP.

Four main discovery approaches were used:

- **Querying a machine-readable catalogue**

In countries or portals with structured APIs, the pipeline retrieved dataset metadata programmatically and filtered records based on declared format or transport type. This approach was used for sources such as Germany, Spain, France, Austria and the EDP. Where available, API metadata fields were used to identify candidate GTFS resources, for example by selecting records explicitly labelled as GTFS or public transport data.

Germany uses a platform called Mobilithek (mobilithek.info), which describes itself as a national mobility data marketplace. Its catalogue is only accessible with a registered account. After logging in via a protocol called Keycloak (a standard secure login system used in enterprise platforms), the loader sends a search query specifying that it only wants datasets whose declared data model matches the official GTFS semantic identifier (https://w3id.org/mdp/schema/data_model#GTFS). The platform then returns all matching publications, paginated across multiple pages. In total, this found around 44 GTFS datasets. Without providers' credentials, only a small hardcoded fallback list of known publication IDs would be used.

- **Scraping a file directory**

The second approach was scraping file directories when no formal API catalogue existed. In the Netherlands, the portal exposed a public web directory listing downloadable ZIP files. The directory was parsed automatically, and only timestamped snapshot files were retained, avoiding alias links that could repeatedly point to the same latest version.

- **Known, hardcoded identifiers**

The third approach was the use of manually identified stable resource identifiers. In several Member States, discovery could not be automated through a searchable public catalogue because access depended on registration, approval or prior knowledge of dataset pages. In these cases, resource URLs, content IDs, universally unique identifiers (UUIDs) or API endpoints were first identified manually and then incorporated into the loader scripts as fixed references. This approach was used for cases such as Belgium, Ireland, Croatia, part of Poland and Slovenia.

Belgium (transportdata.be) does not have an API that can be queried for a GTFS catalogue. The portal was browsed manually to find the datasets that existed and whether they had downloadable GTFS files. Two operators were found: TEC (Wallonia) with a direct public URL, and De Lijn (Flanders) accessible via a developer API with a subscription key. SNCB (the national Belgian rail operator) was investigated, but no downloadable GTFS file was found on their portal. The two URLs and one API key were then hardcoded.

Ireland (transportforireland.ie) provides direct download URLs for 22 operator feeds. These were identified by manually exploring the portal and hardcoded.

Croatia (b2b.promet-info.hr) requires manual portal registration and approval for each of the six transport operators that publish GTFS data. Once approved, each operator's dataset is identified by a UUID (a long unique identifier). These six UUIDs are hardcoded.

Poland (dane.gov.pl) is a general government open data portal. GTFS datasets can be found by browsing or searching the portal. Two GTFS-related datasets were identified. Only one – the Wrocław city feed (resource ID 1823) – was directly downloadable as a ZIP. The second (Małopolska region, resource ID 46165) exposes GTFS tables through a representational state transfer (REST) API requiring a registered account, so it was not included.

Slovenia (nap.si) requires registering an account and requesting access to each dataset. Once approved, each dataset has a content ID that forms part of its download URL (https://b2b.nap.si/data/<content_id>). Six such content IDs were hardcoded.

- **No automated discovery**

The fourth approach was a fully manual download. For Luxembourg, Hungary and Portugal, no automated discovery workflow was implemented. Instead, files were downloaded manually from the relevant portals and placed in predefined local folders for later processing by the transformation pipeline.

Overlap methodology

For Q1, overlap between NAP and the EDP feeds was assessed through a two-stage matching procedure. In the first stage, a filename-based comparison was performed using the normalised core portion of each filename, referred to here as the filename **stem**. This stem was obtained by removing the file extension and variable elements such as timestamps and version numbers. For example, gtfs_sweden_2024-11.zip was reduced to gtfs_sweden. Specifically, the NAP stem (zip_stem) was derived from the filename of the ZIP archive published on the NAP, while the EDP stem (dist_stem) was derived from the filename or URL path of the corresponding distributions on the EDP. In DCAT-AP terms, a distribution is a specific downloadable or accessible representation of a dataset. Both stems were converted to lower case before comparison.

In the second stage, candidate matches were evaluated through content comparison, focusing on stop coordinates and route information. To support this step, a reverse index was built from stop coordinates to NAP feeds, and an overlap percentage was calculated for each candidate pair. On this basis, feed pairs were classified into three categories: 'match' when overlap was equal to or greater than 80 %; 'partial' when overlap was between 30 % and 79 %; and 'no match' when overlap was below 30 %. This procedure allowed the identification of exact or near-equivalent feeds across sources while distinguishing them from cases of partial similarity and from feeds that were substantively unrelated.

3.3. Analysis framework

This framework adapts the stepwise mapping process developed by the Data4PT project (Bourée et al., 2021), originally designed for comparing two data standards ([1]), as a progressive data quality and homogeneity assessment methodology for GTFS feeds collected across Member States. The three steps are preserved in their original logic (moving from broad to detailed) but reoriented towards a descriptive analysis of a single standard's implementation in practice.

The framework has three stages, which will help to address research questions: 1. high-level comparison; 2. concept-level consistency assessment; and 3. attribute-level identifier analysis.

1. High-level comparison

Objective: Establish which components of the GTFS specification are present across feeds, country by country.

([1]) NeTEx and GTFS.

The GTFS specification defines a set of files, some required and some optional. At this first level, the analysis consists of scanning each downloaded feed and recording which files are present.

This step directly addresses Research Question Q3, i.e. which parts of the GTFS specification are most used across EU feeds, and which are systematically absent.

2. Concept-level consistency assessment

Objective: Determine whether the values used within shared GTFS fields are consistent and semantically coherent across feeds.

Once the structural presence of files is established, the analysis moves inside those files to examine how key fields are populated. The analysis focuses on:

Tier 1. Required fields within required files as a first priority (see Annex 1, Tier 1).

Tier 2. High-priority optional fields with policy relevance (see Annex 1, Tier 2).

For each field, the analysis records the range of values actually used across feeds and compares them against the specification. Divergences may indicate misuse of the field, use of extended/unofficial values or legitimate national adaptations.

3. Attribute-level identifier analysis

Objective: Conduct a detailed examination of how **identifiers are structured**, formatted and scoped across feeds, assessing whether they allow data from different feeds to be compared, combined or linked consistently, with the aim of identifying concrete opportunities for homogenisation.

The third step is the most granular level of analysis and directly addresses Research Question Q4. The analysis focuses on five key identifier fields: agency_id, stop_id, route_id, trip_id and service_id.

Field name	Description
agency_id	Identifies a transit brand that is often synonymous with a transit agency
stop_id	Identifies a location: stop/platform, station, entrance/exit, generic node or boarding area
route_id	Identifies a route
trip_id	Identifies a trip
service_id	Identifies a set of dates when service is available for one or more routes

Table 1. Description of key fields

Two dimensions are examined systematically: first, formatting conventions, assessing whether identifiers follow numeric, alphanumeric, URI-based or structured naming schemes, and whether patterns reflect national or operator-level practices; second, cross-feed collision potential, evaluating the degree to which identifier values from different feeds would conflict if merged into a unified dataset, and whether foreign key references between files are internally consistent within each feed.

4. Analysis

Overlap between the European Data Portal and national access points (Q1)

The overlap analysis was conducted based on unique GTFS feeds identified after validation and deduplication in each source. A GTFS feed is a downloadable machine-readable package, usually provided as a ZIP file containing a set of structured text files (e.g. stops.txt, routes.txt, trips.txt and stop_times.txt) that together describe a public transport network and its services. A dataset, by contrast, is the catalogue entry published on a data portal that describes and provides access to data resources; a single dataset may contain one GTFS feed, multiple feeds, different versions of the same feed or other related distributions. Therefore, the counts reported here refer to individual validated and downloadable GTFS feeds.

Under this criterion, the corpus comprises 995 NAP feeds and 172 feeds from the EDP. Of these, 153 feeds were found in both sources, meaning that the shared set represents approximately 15 % of the NAP corpus and 89.5 % of the EDP corpus. These figures already indicate a strongly asymmetric relationship between the two sources: while most of the GTFS feeds available through the EDP are also present in NAPs, the NAP corpus is substantially broader and contains a much larger volume of unique material.

The comparison between NAPs and the EDP shows that the degree of overlap in GTFS availability is generally low across the EU. In most cases, the EDP does not provide an additional or equivalent layer of operational GTFS data relative to the national sources. Instead, the evidence suggests that meaningful overlap is confined to a very small number of country cases, while for most Member States the NAP remains the only programmatic source from which GTFS data can be obtained and analysed.

Member State	NAP feeds	EDP feeds
BE	2	0
DE	1	1
EE	25	0
ES	149	0
FR	517	169
LT	53	0

Member State	NAP feeds	EDP feeds
AT	21	0
PT	5	1
FI	172	0
IE, HR, CY, LU, HU, NL, PL, SI, SE	1–22	0

Table 2. Country-level coverage

According to the country-level results, France is the only case of major overlap. Portugal is a formal but not substantive case of overlap, while Belgium is present in the catalogue but not operationally usable from the EDP, and Germany is a case in which both sources exist but do not overlap geographically or institutionally. Thirteen of the seventeen Member States represented in the NAP corpus have no representation at all on the EDP side.

France constitutes by far the main overlap case in the corpus. On the NAP side, France contributes 517 feeds, while 169 feeds were retained from the EDP. Of these, 153 feeds were matched across the two sources, including 53 identical files and 100 additional content matches. A further 7 feeds were classified as partial overlaps, and 9 remained unique on the EDP side. In proportional terms, 90.5 % of the French EDP feeds are duplicates of NAP material. These shared French feeds correspond to 38.7 % of the French NAP corpus and 15.4 % of the total NAP corpus, confirming that the overlap between the two sources is overwhelmingly concentrated in France.

This feed-level pattern is reinforced by the provenance and agency comparison. The GTFS distributions harvested through the EDP point predominantly to **data.gouv.fr**, the same underlying open data environment that supports the French NAP, **transport.data.gouv.fr**. The French NAP contains 517 feeds and 653 unique agencies, whereas the EDP contains 169 feeds and 254 unique agencies. Of the agencies found on the EDP, 234 also appear in the French NAP, corresponding to 92.1 % of the EDP agencies and 35.8 % of the NAP agencies. The agency comparison was performed using agency_name. Taken together, these results indicate that the French GTFS material exposed through the EDP is best understood as a partial republication of the national source rather than as an independent complementary layer.

Metric	Value
NAP (transport.data.gouv.fr)	517 feeds, **653** unique agencies
EDP	171 feeds, **254** unique agencies
Agencies in both	**234** (92.1 % of data.eu; 35.8 % of NAP)
Agencies only in NAP	**419** NAP is far more complete
Agencies only in data.eu	**20** minor additions (seasonal, local operators)

Table 3. France overlap

Portugal appears at first sight to be another case in which both sources are present, but this is not a meaningful overlap in operational GTFS terms. The Portuguese data retrieved through the EDP does not function as a timetable feed. Instead, it contains only 19 478 stop coordinates with stop names and latitude and longitude values for Lisbon and Porto, and does not include routes, trips or stop_times. As a result, it cannot be considered comparable to the operational timetable feeds available through the Portuguese NAP. In this case, formal co-presence across the two sources does not translate into substantive overlap for transport service analysis.

Belgium illustrates a different problem: the gap between catalogue presence and effective downloadability. On the NAP side, the available material includes the De Lijn and TEC feeds. On the EDP side, the catalogue lists De Lijn Static, TEC, SNCB, De Lijn Realtime and one unrelated dataset. However, the GTFS-related Belgian entries on the EDP do not resolve to downloadable GTFS packages. Some point only to API documentation pages, while others to directory listings or general service pages, and the only resource with an actual download URL was not GTFS at all but a shapefile related to electricity substations. Belgium therefore cannot be treated as a case of operational overlap between the two systems, even though both sides exhibit some form of metadata presence.

Germany is the only Member State besides France in which both a NAP feed and an EDP feed are available as actual downloadable sources; yet, here again the evidence does not support meaningful overlap. The German NAP side is represented by the Baden-Württemberg feed published through Mobilithek under the NVBW umbrella, while the EDP side consists of a single feed for the Ingolstadt metropolitan area in Bavaria. The two sources have zero agency overlap.

These figures were obtained by parsing the content of each downloaded ZIP file: the NAP feed contains 596 agencies listed in agency.txt, 9 580 routes listed in routes.txt and

136 619 stops listed in stops.txt, whereas the EDP side contains one agency, 129 routes and 4 212 stops in the corresponding files.

These are therefore independent and geographically non-overlapping datasets. The German case is particularly important because it also reveals the limits of access on both sides. Mobilithek reportedly catalogues around 44 GTFS datasets nationally, but only the Baden-Württemberg feed was publicly accessible without operator approval, while the EDP lists five German datasets of which only Ingolstadt was downloadable. Germany thus illustrates that the problem is not simply one of portal duplication, but also of restricted or incomplete public accessibility.

Finland appeared in EDP metadata, but the download_url does not work, and the access_url redirects to another page that does not work either and ends up redirecting to the official website. Czechia does not have GTFS datasets, only shapefiles or a geodatabase. Italy has GTFS datasets that are tagged as ZIP files, but they are not included in the initial analysis since they are not explicitly identified as GTFS. It is the same for Slovakia, but in this case only one dataset in ZIP.

Taken together, these results suggest that the EDP plays a limited role as a source of GTFS data compared with NAPs. In France, it largely mirrors an already richer national source; in Portugal, it contributes only stop coordinate data without a timetable structure; in Belgium, it provides catalogue visibility without usable GTFS downloads; and in Germany, it offers a geographically separate local feed rather than a parallel or overlapping national alternative. For the remaining Member States included in the NAP corpus, no matching feed was identified in the EDP equivalent. The broader implication is that the EDP cannot currently substitute for the national NAP landscape if the goal is to achieve comprehensive and operationally usable EU-wide GTFS coverage. Rather than functioning as a complete supranational layer, it appears mainly as a partial mirror, a selective index or, in some cases, a catalogue without effective downloadability.

Coverage of the EU territory (Q2)

The territorial coverage analysis indicates that the current corpus contains at least some GTFS data for 18 of the 27 Member States, while 9 Member States remain without data in the current collection. For analytical purposes, the Member States were grouped into three categories: Member States with apparent full or national coverage; Member States with partial coverage; and Member States with no data. Under this classification, 11 Member States fall into the full or national coverage group, namely Estonia, Ireland, Spain, France, Cyprus, Luxembourg, the Netherlands, Austria, Slovenia, Finland and Sweden. A further 7 Member States – Belgium, Germany, Croatia, Lithuania, Hungary, Poland and Portugal – were classified as partial because the available collection is known to omit important operators, is geographically limited or was assembled under constrained collection conditions. The remaining 9 Member States – Bulgaria, Czechia, Denmark, , Italy, Latvia, Malta, Romania and Slovakia – are currently unrepresented in the corpus. In this context, full coverage should be understood operationally rather than absolutely: it indicates that a working national collection pipeline was in place and that no systematic national gap was identified during data gathering.

Status	Member States	Count
Full/national coverage	EE, IE, ES, FR, CY, LU, NL, AT, SI, FI, SE	11
Partial coverage	BE, DE, HR, LT, HU, PL, PT	7
No data	BG, CZ, DK, EL, IT, LV, MT, RO, SK	9

Table 4. EU-27 Member State coverage

The distinction between full and partial coverage is important because territorial presence alone does not indicate analytical completeness. In Belgium, for example, the available feeds cover De Lijn and TEC, but omit SNCB and STIB. In Germany, the corpus contains only the Baden-Württemberg feed, while many additional GTFS datasets are catalogued nationally but have not yet been downloaded. In Croatia and Hungary, the collection has relied on manually assembled feed lists rather than a complete national discovery mechanism. Lithuania is classified as partial because its bus data appears broad, but the rail feed is effectively empty. Poland is represented only by a regional feed for the Wrocław area, and Portugal by a small, manually assembled set that excludes major operators such as Carris (Lisbon public surface transport), CP (the Portuguese national rail operator) and Metro Lisboa. These cases show that territorial inclusion and territorial completeness are not equivalent and should not be interpreted as such in the report.

Member State	Reason for partial classification
BE	De Lijn + TEC downloaded, but SNCB (national rail) and STIB (Brussels urban) are absent
DE	Only one feed (Baden-Württemberg); Mobilithek has 44 datasets, most not yet downloaded
HR	Six operators hardcoded manually; no catalogue API to discover others
LT	Bus coverage looks comprehensive, but rail feed contains only four stops and two routes
HU	Four feeds placed manually; national operator registry unknown

Member State	Reason for partial classification
PL	Single regional feed (Wrocław area only)
PT	Five feeds placed manually; major operators (Carris, CP, Metro Lisboa) absent

Table 5. Details for partial classification

In terms of corpus scale, the distribution of feeds, routes and stops is highly uneven across Member States (see Annex 2). The current collection comprises 995 feeds, 131 372 routes and 2 461 010 stops with valid non-zero coordinates. France dominates the corpus by a substantial margin, with 517 feeds, 55 451 routes and 962 226 located stops. Spain follows with 149 feeds and 24 558 routes, while Finland contributes 172 feeds and 11 007 routes. Germany appears with only one feed in this collection, but that single feed contains 9 580 routes and 136 619 located stops, reflecting its large regional scale. The Netherlands presents an especially important structural case: it appears as a single feed, yet that feed contains 3 412 routes and 60 777 located stops, indicating that feed counts alone do not necessarily reflect substantive national coverage. Estonia also stands out, with 384 303 located stops, a figure that is unexpectedly high relative to country size and likely reflects very fine-grained stop modelling or overlap across published feeds. Overall, the territorial picture is therefore shaped not only by the presence or absence of GTFS, but also by a very strong concentration in a small number of countries.

Mode availability shows a similarly uneven pattern. Nine Member States in the corpus contain both bus and rail GTFS feeds: Estonia, Ireland, Spain, France, Croatia, Lithuania, Austria, Portugal and Finland. The remaining nine Member States with data are represented as bus only in the current collection: Belgium, Germany, Cyprus, Luxembourg, Hungary, the Netherlands, Poland, Slovenia and Sweden. No country appears as rail only. This pattern should be interpreted cautiously. The absence of rail GTFS in a country does not necessarily mean that rail services are absent, but rather that separate rail GTFS feeds were not identified in the collected material. Lithuania is a particularly important special case because, although it falls into the bus plus rail category, its rail feed contains only four stops and two routes, making it effectively unusable for substantive rail analysis. The practical conclusion is that multimodal territorial coverage remains uneven even among Member States that are nominally represented.

Mode	Member States
Bus + rail	EE, IE, ES, FR, HR, LT, AT, PT, FI
Bus only	BE, DE, CY, LU, HU, NL, PL, SI, SE
Rail only	–

Table 6. Mode availability

The analysis of stop location quality suggests that the corpus is generally strong in geographic usability, although not entirely free from quality issues. Two distinct error types were identified: missing coordinates, where stop_lat ([2]) or stop_lon ([3]) is empty or non-numeric; and zero coordinates, where stops are assigned the placeholder location (0.0). Spain records 604 stops with missing coordinates; France records 5 missing and 621 zero-coordinate stops; Germany has 100 zero-coordinate stops; and Sweden records 2 zero-coordinate stops. In all four cases, the proportion remains below 0.25 % of total stops, so these problems do not materially undermine the usability of the feeds for broad territorial analysis. However, they remain relevant for downstream harmonisation and automated processing, since missing coordinates and placeholder coordinates require different handling rules.

Member State	Missing	Zero (0.0)	% of all stops	Nature
DE	0	100	0.07 %	Zero placeholder
ES	604	0	0.25 %	Empty fields – spec violation
FR	5	621	0.07 %	Mix of both types
SE	0	2	< 0.01 %	Zero placeholder

Table 7. Stop location data quality

To assess the territorial extent of the available data beyond simple country presence, the analysis also compared the bounding box (bbox) of stop coordinates in each Member State against the official national bbox retrieved from Nominatim (OpenStreetMap). This produced a span-based indicator of how far the data reaches across a national territory in latitude and longitude. The results are useful for detecting extreme under-coverage, but they must be interpreted with caution. The metric captures edge-to-edge spread rather than true areal coverage, so a sparse set of edge points can yield a deceptively high result. It is also distorted in Member States in which the official bboxes include overseas

([2]) Latitude of the location.
([3]) Longitude of the location.

territories or where feeds include cross-border stops. Even with these caveats, some results are informative. Poland scores only 5 %, confirming that the available material corresponds essentially to a single-city case. Hungary reaches 44 %, supporting the interpretation that the current collection covers only part of the country. Cyprus reaches 51 %, indicating that the available data covers only the southern part of the island. By contrast, the Netherlands appears artificially compressed because the denominator includes Caribbean territories, even though mainland coverage is effectively national. The bbox analysis is therefore best used as a diagnostic complement rather than a stand-alone measure of territorial completeness.

Member State	**Lat. span**	**Lon. span**	**Avg**	**Assessment**
CY	40 %	62 %	51 %	Only southern Cyprus; northern part has no GTFS data
HU	48 %	41 %	44 %	Confirmed partial: western Hungary only (Budapest, Kaposvár, Paks)
NL	30 %	27 %	28 %	Nominatim bbox includes Caribbean territories (Aruba, Curaçao), inflating denominator; mainland coverage is actually full
PL	4 %	5 %	5 %	Single city (Wrocław): confirmed
PT	25 %	4 %	14 %	Confirmed partial: Nominatim bbox includes Azores/Madeira; even within mainland, only central/northern corridor covered

Table 8. Geographic extent: bbox span coverage

When the EDP is examined as a comparator rather than as the main basis for territorial coverage, its contribution appears limited. Only three Member States (Germany, France and Portugal) plus one non-EU country are represented by imported GTFS material on the EDP side. Moreover, the scale of that data is markedly smaller than what is available through the NAPs. In Germany, the EDP contains only the Ingolstadt feed, whereas the NAP side contains the much larger Baden-Württemberg feed. In France, the EDP contains 171 feeds, 13 183 routes, and 215 462 stops, compared with 517 feeds, 55 451 routes, and 962 226 stops in the French NAP. In Portugal, the EDP side contains only stop coordinates and no timetable structure, making it unsuitable for timetable-based territorial analysis. On this basis, the EDP does not add meaningful geographic coverage beyond the national sources already included in the corpus.

Country	Modes	Feeds	Routes	Located stops	Notes
DE	Bus	1	129	4 212	Ingolstadt city only
FR	Bus + rail	171	13 183	215 462	Subset of French NAP
PT	Bus	2	0	19 478	Stop coordinates only no timetable
UA (Ukraine)	Bus	1	6	152	Non-EU; Chornomorsk city

Table 9. European Data Europa coverage

Key takeaways

1. In the corpus, 18 of 27 Member States have at least some GTFS data; 9 have none (BG, CZ, DK, EL, IT, LV, MT, RO, SK).
2. Eleven Member States have what appears to be national coverage; the remaining seven with data are confirmed or likely partial.
3. France dominates the corpus in absolute terms (517 feeds, 962 000 stops, 55 000 routes), reflecting its mature open data ecosystem.
4. Rail data is sparse: only nine Member States publish separate bus and rail GTFS feeds, and Lithuania's rail feed is effectively empty. Several Member States with active national rail services (NL, SI, SE) publish no separate rail GTFS.
5. Stop location quality is generally high: only four Member States have stops without a valid location, and in all cases the proportion is below 0.25 %. The two failure modes empty fields and (0.0) placeholders are distinct and require different handling in a data pipeline.
6. The bbox span metric confirms the most extreme coverage gaps (PL at 5 %, HU at 44 %) but is unreliable as a general coverage percentage due to cross-border feeds and overseas territory effects on the denominator. The combination of bbox analysis, agency count and known missing operators gives a more complete picture.
7. The EDP adds no meaningful geographic coverage beyond what the national NAPs already provide, and its data is significantly smaller in scale.

Adoption of General Transit Feed Specification components across published feeds (Q3)

The analysis of GTFS specification usage was conducted on a corpus of 995 feeds collected from 18 EU NAPs, covering Belgium, Germany, Estonia, Ireland, Spain, France, Croatia, Cyprus, Lithuania, Luxembourg, Hungary, the Netherlands, Austria, Poland, Portugal, Slovenia, Finland and Sweden. At the broadest structural level, the results show that the GTFS core is highly stable across the corpus. The five required GTFS files, `agency.txt`, `stops.txt`, `routes.txt`, `trips.txt`, and `stop_times.txt`, are present in 98.2 to 98.4 % of feeds. This means that the basic timetable structure of GTFS has been adopted almost universally across the countries included in the corpus.

File	Feeds	%
routes.txt	979	98.4 %
stops.txt	979	98.4 %
agency.txt	978	98.3 %
stop_times.txt	977	98.2 %
trips.txt	977	98.2 %

Table 10. Required files usage

Calendar modelling is also relatively consistent, although not fully uniform. GTFS requires at least one of `calendar.txt` or `calendar_dates.txt`, and the dominant pattern in the corpus is to use both files together. This combined model appears in 656 feeds, or 65.9 % of the total, indicating that many publishers rely on a hybrid representation of regular weekly schedules plus date-specific exceptions. A further 18.9 % of feeds use only `calendar_dates.txt`, while 13.3 % use only `calendar.txt`. Only 19 feeds, or 1.9 %, contain neither file, which constitutes a specification violation. These cases are concentrated mainly in Finland and, to a much smaller extent, in France. Overall, the evidence suggests that service calendars are usually modelled in a way that supports operational realism, but also that a small number of feeds remain structurally non-compliant.

Pattern	Feeds	%
Both calendar + calendar_dates	656	65.9 %
Calendar_dates only	188	18.9 %
Calendar only	132	13.3 %

Pattern	Feeds	%
Neither (spec violation)	19	1.9 %

Table 11. Service calendar patterns

Beyond the required core, adoption drops sharply. Only two optional GTFS files are present in more than one quarter of the corpus: `shapes.txt`, included in 75.5 % of feeds, and `feed_info.txt`, included in 57.2 %. All other optional files fall below 25 % usage: `transfers.txt` appears in 21 % of feeds, while the classic fare files `fare_attributes.txt` and `fare_rules.txt` appear in only 8.1 % and 6.5 %, respectively. Files related to accessibility, such as `pathways.txt` and `levels.txt`, are even rarer, at 3.1 % and 2.7 %. Files associated with service frequency, demand-responsive transport, grouping structures and the newer fare framework are largely marginal in the corpus. At the same time, no optional GTFS file is completely absent, which indicates that the full specification is being used somewhere in the EU NAP landscape, but only in very limited and uneven ways.

File	Feeds	Percentage
shapes.txt	751	75.5 %
feed_info.txt	569	57.2 %
transfers.txt	209	21.0 %
fare_attributes.txt	81	8.1 %
fare_rules.txt	65	6.5 %
translations.txt	38	3.8 %
attributions.txt	35	3.5 %
pathways.txt	31	3.1 %
levels.txt	27	2.7 %
frequencies.txt	25	2.5 %
booking_rules.txt	14	1.4 %
location_groups.txt	13	1.3 %
areas.txt	8	0.8 %

File	Feeds	Percentage
All remaining optional files	< 8	< 0.8 %

Table 12. Optional file presence

Two optional components stand out as de facto extensions of the GTFS core in current publishing practice. First, `shapes.txt` is present in roughly three quarters of feeds, indicating that route geometry is widely treated as an operational necessity even though it remains optional in the specification. This is important for downstream uses such as mapping, route tracing and geographic service analysis. However, adoption is not fully homogeneous. Luxembourg and Poland publish no shapes at all, while most other Member States report inclusion rates between 80 and 100 %. Second, `feed_info.txt` appears in more than half of the corpus, suggesting that publisher and feed metadata is also increasingly treated as standard practice. Here, too, national variation is substantial: fourteen of the eighteen Member States use `feed_info.txt` in 83 % to 100 % of feeds, while France, Spain and Portugal show much lower rates, at 46 %, 53 % and 40 %, respectively.

Other optional components reveal much more fragmented implementation patterns. Transfer information is present in only 21 % of feeds overall, although some Member States, notably Germany, Luxembourg, the Netherlands and Sweden, use `transfers.txt` systematically in all feeds. Fare data is largely absent from the corpus. Classic GTFS fare files appear only in a small minority of feeds and are concentrated in a few Member States, especially Estonia, Spain, France, Cyprus and Portugal. Accessibility-related files are also rare across the corpus, but Austria is a clear outlier: 95 % of Austrian feeds include `pathways.txt`, and 100 % include `levels.txt`. This suggests that accessibility-related stop and station structure is not yet part of the common European GTFS baseline, but that some national ecosystems are already implementing it at high levels. Similarly, files related to flexible or demand-responsive services, such as `booking_rules.txt` and `location_groups.txt`, appear only in 1.4 % and 1.3 % of feeds, respectively, limited mainly to France, Austria and Sweden. In practical terms, this indicates that GTFS-Flex-related publishing remains incipient in the current corpus.

The newer GTFS fare framework is even less mature. Files such as `fare_media.txt`, `fare_products.txt`, `fare_leg_rules.txt`, `fare_transfer_rules.txt`, `rider_categories.txt`, `networks.txt`, `route_networks.txt`, `areas.txt`, `stop_areas.txt` and `timeframes.txt` all appear in less than 1 % of feeds. Only Spain and France show any meaningful experimentation with this model. This is a significant finding for the broader objective of harmonisation, because it suggests that fare modelling remains one of the least standardised and least mature parts of GTFS implementation across the EU. At present, the published NAP corpus is still overwhelmingly oriented towards timetable and basic route structure, not towards integrated fare representation.

A further important result is the prevalence of non-standard extension files. Across the corpus, 56 distinct non-standard tables were identified. These are especially concentrated in France and Finland. In France, the extensions are mainly operator-specific and include files such as trips_extensions, thermometre, ticketing_deep_links, object_codes_extension, route_directions and timetables. In Finland, the extensions are largely associated with the Digitransit ecosystem and include contracts, extended_contracts, digitransit_fare_attributes, digitransit_fare_rules, tdm_blocks and emissions. Additional isolated extensions appear in Spain, Hungary, Poland and Portugal. These extensions cover operational functions that are not yet well represented in the standard GTFS model, including vehicle fleet management, commercial contract structures, block scheduling, dead-run tracking and emissions data. Their presence therefore reveals not simply local inconsistency, but also areas in which publishers are adapting GTFS to meet needs that the official specification does not yet fully accommodate. For a full summary, see Annex 3.

Key takeaways

1. Core files are quasi-universal (~ 98 % adoption) but not perfectly uniform; a small number of malformed feeds are missing required files.
2. Shapes and feed_info are the de facto standard optional files, used by the majority of NAP feeds.
3. Fare data is largely absent from NAP feeds; only 8 % include classic fare files, and GTFS v2 fare extension adoption is below 1 %.
4. Service modelling is predominantly hybrid (calendar + calendar_dates combined, 65.9 %), with a meaningful share using calendar_dates exclusively (18.9 %), reflecting exception-heavy schedule patterns.
5. Accessibility data (pathways, levels) is rare across the EU; Austria is the clear outlier with near-universal pathway data.
6. Proprietary extensions are widespread in France and Finland, indicating that the GTFS spec does not yet cover all operational needs of these data publishers.

Identifier provision and homogenisation opportunities (Q4)

The analysis of identifier structures across the GTFS corpus reveals a high degree of fragmentation in both format and scope, with important implications for cross-feed interoperability. The corpus comprises **995 feeds from 18 Member States**, and the results show that while identifiers are consistently present within feeds, they are not standardised across them. As a consequence, GTFS feeds are internally coherent but remain difficult to integrate or compare across operators, regions or countries without additional transformation layers.

The most critical case is the stop_id, which functions as the primary technical key for stops and underpins multiple relational links across the GTFS model. Across the corpus, four main formatting patterns are observed. Numeric identifiers are the most widespread and appear in the majority of Member States, including Belgium, Estonia, Spain, Cyprus, Lithuania, Luxembourg, Hungary, the Netherlands, Poland, Slovenia, Finland and Sweden. These identifiers are typically short numeric strings with no embedded structure. A second group of Member States, including Germany, Croatia and Austria, uses prefixed identifiers following a NeTEx-like convention, where the identifier encodes country and regional information, for example DE:08235:3040 or at:vvt:10:. A third group, including Ireland and Portugal, uses opaque alphanumeric strings without a clear semantic structure. Finally, France combines numeric identifiers with UUID-like values.

From a homogenisation perspective, the key distinction is not between numeric and alphanumeric formats, but between locally scoped identifiers and globally structured identifiers. Numeric identifiers, although simple, are strictly local to each feed. The same value may refer to completely different stops in different feeds, making cross-feed merging ambiguous. By contrast, the prefixed identifiers used in Germany, Croatia and Austria embed namespace information and are therefore designed to be globally unique. These are the only identifier schemes in the corpus that can support cross-feed interoperability without additional mapping. The absence of a common EU-wide convention for stop_id is therefore the single most significant structural barrier to harmonisation identified in this analysis.

Format	**Member States**
Numeric (digits only)	BE, EE, ES, CY, LT, LU, HU, NL, PL, SI, FI, SE
Prefixed with colon (NeTEx-style, e.g. DE:08235:3040, at:vvt:10:)	DE, HR, AT
Opaque string (e.g. 5524_126901)	IE, PT
UUID (mixed with numeric)	FR

Table 13. stop_id format

This fragmentation is further confirmed in the cross-source comparison between NAPs and the EDP. In the three Member States where both sources are present, Germany, France and Portugal, the stop_id formats are incompatible in all cases. Germany uses prefixed identifiers in the NAP feed and numeric identifiers in the EDP feed. France combines numeric and UUID-based identifiers in the NAP, while the EDP uses opaque strings. Portugal shows opaque identifiers in the NAP and numeric identifiers in the EDP. Only in France is there a measurable intersection, where approximately 54.4 % of sampled stop_id values from the EDP are found in the NAP corpus, reflecting the fact that many of the EDP feeds are direct re-publications of the French NAP. In Germany and Portugal, no meaningful identifier overlap is observed.

The route_id field exhibits similar diversity in formatting but has lower cross-feed impact. Numeric identifiers dominate in many Member States, including Belgium, France, Cyprus, Lithuania, Luxembourg, Hungary, the Netherlands, Poland, Slovenia, Finland and Sweden. Opaque string identifiers appear in Germany, Estonia, Ireland, Spain and Portugal, while prefixed formats are used in Croatia and Austria. However, unlike stop_id, route identifiers are primarily used within a single feed and are rarely intended for cross-feed comparison. As a result, their lack of standardisation does not pose the same level of interoperability constraint, although it still prevents the automated cross-country aggregation of route-level data without additional processing.

Format	**Member States**
Numeric	BE, FR, CY, LT, LU, HU, NL, PL, SI, FI, SE
Opaque string	DE, EE, IE, ES, PT
Prefixed with colon	HR, AT

Table 14. route_id format

The agency_id field presents a different pattern. Its fill rate is high, with 16 out of the 18 Member States achieving between 99 % and 100 % completeness, and only Spain and France showing slightly lower rates. This indicates that feed publishers consistently link routes to operating agencies. However, the identifier itself is not standardised. Numeric identifiers are used in several Member States, while others rely on opaque strings, and Croatia again uses a prefixed format. As with stop_id, the identifiers are local to each feed and do not correspond to any shared registry. Consequently, the same operator may be represented by different agency_id values across feeds and countries. This limits the ability to track operators across jurisdictions or to aggregate services at the European level.

A related aspect is the declaration of language in feed metadata. Language codes are present in 17 out of the 18 Member States, and where used they are generally consistent with ISO 639-1 standards. Only Luxembourg does not declare a language code. While this is a relatively minor issue compared to identifier fragmentation, it highlights that even simple metadata fields are not yet uniformly implemented across the corpus.

Format	Member States
Numeric	EE, IE, ES, CY, LT, LU, AT, PL, SI, FI, SE
Opaque string	BE, DE, FR, HU, NL, PT
Prefixed (NeTEx)	HR

Table 15. agency_id format

The route_type field illustrates a different kind of inconsistency, namely divergence in controlled vocabularies. GTFS defines a standard set of route types using values from 0 to 12, and an extended set of values above 100 for more detailed classification. In the corpus, 11 Member States use only the standard values, while 7 Member States use extended values. Sweden stands out as a special case, using only extended values and no standard ones. This creates a structural incompatibility: systems that interpret only the standard GTFS values cannot correctly classify Swedish routes. The most common extended values found in the corpus include categories such as school bus services and long-distance rail, which are not distinguishable within the basic GTFS set. As a result, cross-country modal comparisons require explicit mapping between standard and extended values, and there is currently no consistent approach across Member States.

Extended type	Meaning	Routes
712	School bus	4 113
713	School and public service bus	4 028
700	Bus	3 676
102	Long-distance rail	1 204
1104	Helicopter service	1 047

Table 16. route_type

The optional stop_code field offers a potential mechanism for improving cross-feed comparability, but its use is highly uneven. In two Member States, Austria and Sweden, the field is not present at all. In four Member States, Germany, Luxembourg, the Netherlands and Slovenia, the field exists but is completely empty. Several Member States show low fill rates, including Belgium, Croatia, Lithuania and Finland. By contrast, a smaller group of Member States, including Estonia, Ireland, Spain, France, Cyprus, Hungary, Poland and Portugal, show moderate to high adoption, with some achieving near-complete coverage. In cases such as Estonia and Ireland, stop codes are assigned systematically at the national level, making them suitable for cross-operator identification within the country. However, because there is no EU-wide standard for stop_code, its usefulness remains limited to national contexts.

Status	Member States
No stop_code column at all	AT, SE
Column present but empty (0 %)	DE, LU, NL, SI
Low fill rate (< 50 %)	BE (49 %), HR (42 %), LT (6 %), FI (29 %)
Good fill rate (≥ 50 %)	EE (99 %), IE (100 %), ES (73 %), FR (56 %), CY (100 %), HU (96 %), PL (100 %), PT (65 %)

Table 17. Stop code

Key homogenisation opportunities

Several concrete opportunities for further homogenisation emerge from the analysis. The most critical is the introduction of a globally scoped identifier scheme for stops. The prefixed model already used in Germany, Croatia and Austria demonstrates that such an approach is technically feasible and compatible with existing standards such as NeTEx. Adopting a similar convention across the EU would allow the unambiguous identification of stops across feeds without requiring external mapping tables. A second high-impact opportunity lies in harmonising the use of route_type, either by converging on the standard GTFS set with a defined mapping from extended values or by adopting the extended set consistently across Member States. Medium-impact opportunities include the standardisation of stop_code as a nationally assigned and consistently used identifier, and the linkage of agency_id values to a shared operator registry. Finally, lower-cost improvements, such as ensuring that all feeds declare a language code, could enhance usability with minimal implementation effort.

Overall, the Q4 results confirm that the main limitation to cross-country integration of GTFS data is not the absence of identifiers, but the absence of shared conventions governing their structure and scope. GTFS successfully enforces internal relational consistency within feeds, but it does not provide a framework for interoperability across feeds. As a result, achieving EU-wide comparability requires either the introduction of common identifier schemes or the systematic use of mapping layers that reconcile existing local identifiers.

5. Conclusions and recommendations

Conclusions

- NAPs are the operative backbone of the current EU GTFS landscape, while the EDP plays only a limited supplementary role.
- The EU territorial coverage remains incomplete and uneven; 40 % of the EU has national or near-national coverage.
- The available GTFS corpus is highly concentrated in a small number of Member States, especially France.
- The core GTFS timetable model is highly consolidated, but the optional layers of the specification remain weakly and unevenly adopted.
- Fare representation is one of the weakest areas of standardised GTFS implementation.
- Non-standard extensions are an important signal of unmet specification needs (the corpus includes 56 distinct non-standard tables).
- The main obstacle to cross-country integration is not the absence of identifiers, but the absence of shared identifier conventions.
- Controlled values also lack harmonised implementation, especially in route_type. This means that even apparently simple cross-country comparisons, such as counting bus or rail services, require prior normalisation. In other words, interoperability problems affect not only identifiers but also core semantic classifications.

Recommendations for the European Data Portal

R1. Distinguish between metadata presence and operationally usable feed availability. The country-level analysis consistently shows that catalogue presence on the EDP does not guarantee that a feed is directly downloadable or usable. The EDP should introduce an explicit quality indicator, separate from general dataset metadata completeness, that flags whether a GTFS distribution is directly accessible without operator approval, registration or broken links. Belgium and Germany are cases where this distinction would have immediate practical value.

R2. Actively harvest from NAPs rather than waiting for publisher registration. For the majority of Member States, no EDP equivalent of the NAP feed was identified. This is not necessarily because the data does not exist nationally, but because it has not been registered on the EDP. The EDP should explore systematic harvesting or federated indexing arrangements with NAPs operating under Commission Delegated Regulation (EU) 2017/1926, so that the EDP progressively reflects the national landscape rather than a subset of it.

R3. Use separate feed-level indicators in EU mobility-data monitoring and benchmarking. EDP reporting should not use dataset-record counts as a proxy for effective GTFS availability or territorial coverage. Monitoring exercises should report separately: (a) the number of catalogue records mentioning or declaring GTFS; (b) the number of distributions that are accessible; (c) the number of feeds that can be downloaded and validated as structurally compliant GTFS Schedule packages; and (d) the extent to which those feeds provide substantive geographic, modal and operator coverage. This separation would prevent metadata volume from being interpreted as evidence of operational data availability or EU-wide analytical coverage.

Recommendations for Member States and national access points

R4. Ensure unrestricted public accessibility to GTFS feeds, not just catalogue registration. Germany illustrates that even a well-structured national portal (Mobilithek) may publish dozens of dataset records while making only one feed publicly accessible without operator approval. Member States should review their NAP publication policies to ensure that feeds required under the HVD Implementing Regulation are genuinely open and directly downloadable, consistent with the open data obligations of the regulation.

R5. Adopt a common, globally scoped stop identifier model. This is the highest-impact homogenisation opportunity identified across Q4. A prefixed scheme, structured as [country_code]:[authority_code]:[local_id], along the lines of those already used in Germany (DE:08235:3040), Croatia and Austria (at:vvt:10:), would allow cross-feed stop comparison without requiring external mapping tables for every integration task. Member States and NAP operators should coordinate on a shared prefixing convention, potentially building on existing NeTEx or INSPIRE identifiers where these already exist nationally.

R6. Harmonise route_type **usage across Member States.** The current corpus shows that even a basic cross-country comparison, counting bus or rail services, requires prior normalisation because Member States use both the standard GTFS set and extended values inconsistently. Member States should either use only the standard GTFS route_type values, converting extended values through an agreed common mapping, or use the extended values consistently across Member States with clear published documentation.

R7. Strengthen the publication of fare, accessibility and transfer-related GTFS components. These are currently among the weakest areas of the corpus, yet they are central for user-facing multimodal analysis and for policy objectives related to accessibility and integrated ticketing. Austria demonstrates that higher adoption is achievable, as 95 % of its feeds include `pathways.txt` and 100 % include `levels.txt`. Member States with advanced publishing ecosystems should share implementation guidance with those at earlier stages.

R8. Standardise stop_code **nationally and link** agency_id **to shared registries.** More systematic use of nationally consistent stop codes would support cross-operator reference within Member States. A registry-backed operator identifier (linked to national business or transport authority registers) would improve cross-jurisdictional comparability and reduce the current situation where agency identifiers are locally scoped and non-comparable across feeds.

Recommendations for standardisation bodies (MobilityData / GTFS specification)

R9. Treat non-standard extensions as evidence for future specification development. The corpus includes 56 distinct non-standard tables, concentrated in advanced publishing ecosystems such as France. Rather than treating these as local anomalies, standardisation bodies should analyse them systematically as signals of unmet needs within the current specification, particularly in areas such as commercial contracts, fleet operations, emissions data and integrated fare logic, and consider whether future GTFS versions should formalise some of these patterns.

References

Bourée, K. and Knowles, N. (2021) Methodology for comparing data standards. Data4PT. Available at: transmodel-cen.eu (Accessed: 6 August 2026).

NAPCORE (2025) Mapping of the access nodes identifiers in the Member States (Report No. D2.4.1). National Access Point Coordination Organisation for Europe. napcore.eu

Annexes

Annex 1. Analytical framework: detailed field inventory

Tier 1. Required fields within required files

agency.txt	
Field	**Type**
agency_name	Text
agency_url	URL
agency_timezone	Timezone

routes.txt	
Field	**Type**
route_id	Unique ID
route_type	Enumerated

trips.txt	
Field	**Type**
route_id	Foreign ID → routes.route_id
service_id	Foreign ID → calendar.service_id or calendar_dates.service_id
trip_id	Unique ID

stop_times.txt	
Field	**Type**
trip_id	Foreign ID → trips.trip_id
stop_sequence	Non-negative integer

Tier 2. Fields by policy theme

Theme 1: Accessibility of public transport systems

`wheelchair_accessible` in trips.txt

`wheelchair_boarding` in stops.txt (Are stops systematically assessed for accessibility?)

These fields are directly relevant to EU disability policy and the European Accessibility Act. The analytical challenge is that value 0 means 'no information', which means a feed full of zeros says nothing about actual accessibility provision.

Theme 2: Geographic coverage

`location_type` in stops.txt

This field defines the stop hierarchy, whether a record represents a platform, a station, an entrance or a pathway node. (Is the stop hierarchy consistently represented?)

`parent_station` in stops.txt

This field links platforms to their parent stations. Without it, the stop hierarchy cannot be reconstructed even if `location_type` is correctly populated. Its absence makes station-level coverage analysis impossible.

Theme 3: Service patterns and frequency

`headway_secs` in frequencies.txt

This field defines the time between successive trips on a route, essentially the service frequency. It is one of the most direct proxies for service quality available in GTFS. A route with a 10-minute headway serves passengers very differently from one with a 60-minute headway, even if both appear equally in a simple route count.

However, frequencies.txt itself is an optional file, so it is important to check in Step 1 how many feeds include it at all. If very few EU feeds use it, that finding is itself significant; it means frequency-based service quality comparison across Member States is currently impossible from GTFS data alone.

`timepoint` in stop_times.txt

This binary field distinguishes between exact scheduled times and approximate times. Its absence or universal default to 1 (exact) across feeds may mask the fact that some stop times are estimates rather than firm schedules, which affects the reliability of any service efficiency analysis that is performed. (Are scheduled times exact or approximate?)

Annex 2. Distribution of feeds, routes and stops across member states

Member State	Modes	Feeds	Routes	Located stops	Missing	Zero (0.0)
BE*	bus	2	2 940	61 611	0	0
DE*	bus	1	9 580	136 619	0	100
EE	bus+rail	25	2 327	384 303	0	0
IE	bus+rail	22	847	19 087	0	0
ES	bus+rail	149	24 558	241 637	604	0
FR	bus+rail	517	55 451	962 226	5	621
HR*	bus+rail	7	1 062	12 059	0	0
CY	bus	7	860	5 591	0	0
LT*	bus+rail	53	1 760	19 089	0	0
LU	bus	1	702	2 798	0	0
HU*	bus	4	437	6 866	0	0
NL	bus	1	3 412	60 777	0	0
AT	bus+rail	21	7 933	183 187	0	0
PL*	bus	1	114	2 080	0	0
PT*	bus+rail	5	277	5 320	0	0
SI	bus	6	2 347	9 874	0	0
FI	bus+rail	172	11 007	167 098	0	0
SE	bus	1	5 758	180 788	0	2
Total		**995**	**131 372**	**2 461 010**	**609**	**723**

Annex 3. GTFS coverage across Member States (full summary)

File	BE	DE	EE	IE	ES	FR	HR	CY	LT	LU	HU	NL	AT	PL	PT	SI	FI	SE
feed_info.txt	100 %	100 %	100 %	100 %	53 %	46 %	100 %	100 %	-	-	100 %	100 %	100 %	100 %	40 %	83 %	90 %	100 %
shapes.txt	100 %	100 %	100 %	100 %	79 %	80 %	43 %	100 %	100 %	-	100 %	100 %	100 %	-	100 %	100 %	41 %	100 %
transfers.txt	50 %	100 %	-	-	9 %	29 %	-	-	-	100 %	-	100 %	57 %	-	-	-	18 %	100 %
frequencies.txt	-	-	-	-	11 %	2 %	-	-	-	-	-	-	-	-	-	-	-	-
fare_attributes.txt	-	-	8 %	-	17 %	8 %	-	100 %	-	-	-	-	-	-	60 %	-	2 %	-
fare_rules.txt	-	-	8 %	-	17 %	5 %	-	100 %	-	-	-	-	-	-	60 %	-	2 %	-
pathways.txt	-	-	-	-	4 %	1 %	-	-	-	-	25 %	-	95 %	-	-	-	-	-
levels.txt	-	-	-	-	4 %	-	-	-	-	-	-	-	100 %	-	-	-	-	-
fare_media.txt	-	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	-	-
fare_products.txt	-	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	-	-
fare_leg_rules.txt	-	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	1 %	-
fare_transfer_rules.txt	-	-	-	-	-	0 %	-	-	-	-	-	-	-	-	-	-	-	-
rider_categories.txt	-	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	1 %	-
areas.txt	50 %	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	1 %	100 %
stop_areas.txt	50 %	-	-	-	2 %	-	-	-	-	-	-	-	-	-	-	-	1 %	100 %
networks.txt	-	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	-	-
route_networks.txt	-	-	-	-	2 %	0 %	-	-	-	-	-	-	-	-	-	-	-	-

File	BE	DE	EE	IE	ES	FR	HR	CY	LT	LU	HU	NL	AT	PL	PT	SI	FI	SE
timeframes.txt	-	-	-	-	1 %	-	-	-	-	-	-	-	-	-	-	-	-	-
booking_rules.txt	-	-	-	-	-	2 %	-	-	-	-	-	-	19 %	-	-	-	-	100 %
location_groups.txt	-	-	-	-	-	2 %	-	-	-	-	-	-	19 %	-	-	-	-	-
location_group_stops.txt	-	-	-	-	-	1 %	-	-	-	-	-	-	19 %	-	-	-	-	-
translations.txt	-	-	-	5 %	1 %	0 %	-	100 %	-	-	25 %	-	-	-	-	-	15 %	-
attributions.txt	-	-	-	-	2 %	6 %	-	-	-	-	-	-	-	-	-	-	-	100 %
feeds (N)	2	1	25	22	149	517	7	7	53	1	4	1	21	1	5	6	172	1

Publications Office
of the European Union

ISBN 978-92-78-45252-0

doi: 10.2830/9105423

OA-01-26-072-EN-N