



Risultati, Best Practice e Raccomandazioni OTP-Italy

OTP-Italy, un progetto che punta a migliorare la **transizione dagli standard *de facto* a quelli *de jure* dei dati sulla mobilità**, potenziando l'**interoperabilità** e migliorando l'**accessibilità** del trasporto pubblico.

1. Introduzione	2
2. Converter	4
2.1 Report sviluppo dei converter	4
2.2 Report sviluppo dei mapping	7
2.3 Confronto con soluzioni esistenti	10
2.4 Raccomandazioni e sviluppi futuri	11
3. Miglioramento ed ottimizzazione dei dati in formato GTFS e GTFS-RT	18
3.1 Creazione di file shapes.txt	19
3.2 Raggruppamento di fermate tramite parent_station	21
3.3 Implementazione di route-type extension	22
3.4 Transizione al modello Fares V2	24
3.5 Integrazione dati Pathways e Levels	26
3.6 Raccomandazioni sulla gestione dei dati GTFS	28
4. Sviluppo del fork di OpenTripPlanner	30
5. Conclusioni	32
6. Bibliografia	34

1. Introduzione

Questo documento presenta un report dei **risultati finali** ottenuti dal progetto **OTP-Italy**, svolto in collaborazione fra **OpenMove** e **Cefriel**, ed un insieme di **best practice** e **raccomandazioni** raccolte. Il progetto **OTP-Italy** ha avuto come obiettivo lo sviluppo di un **convertitore dati da GTFS¹ a NeTex profilo italiano² e da GTFS-RT³ a SIRI profilo italiano⁴** (di seguito **converter**) e l'implementazione di un **fork del motore di calcolo del percorso OpenTripPlanner (OTP)**.

Durante il progetto, le attività si sono concentrate su tre territori principali, **Trentino, Alto Adige e Sardegna**, con l'obiettivo di potenziare la **qualità** e la **coerenza dei dati** relativi al **trasporto pubblico locale** per i servizi di trip planning. Le attività hanno riguardato in particolare lo sviluppo dei due converter, la loro validazione **con dati reali forniti dagli operatori e arricchiti** tramite diversi approcci, la realizzazione di un fork personalizzato del motore di calcolo **OpenTripPlanner** e la progettazione di un **frontend dedicato** per la visualizzazione dei dati convertiti e la verifica del corretto funzionamento del processo. La Figura 1 riassume il flusso abilitato dalle soluzioni sviluppate e validato con i dati dei territori considerati.

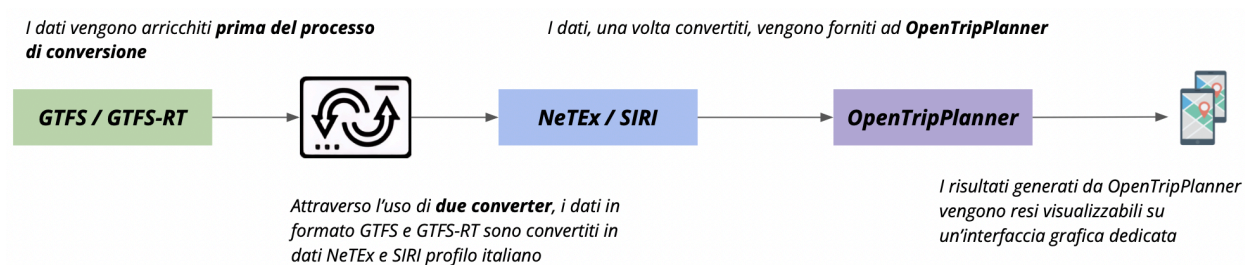


Figura 1: Impostazione logica del flusso tecnico ed operativo

¹ <https://gtfs.org/documentation/schedule/reference/>

² <https://github.com/5TsrI/netex-italian-profile>

³ <https://gtfs.org/documentation/realtime/reference/>

⁴ <https://github.com/5TsrI/siri-italian-profile>

Il **Capitolo 2** illustra in dettaglio il lavoro svolto per lo **sviluppo dei converter**, il **Capitolo 3** approfondisce le attività di **arricchimento e normalizzazione dei dati**, mentre il **Capitolo 4** descrive lo sviluppo e le personalizzazioni apportate al **fork di OpenTripPlanner**. Infine, il **Capitolo 5** presenta le **conclusioni del progetto** e le prospettive di evoluzione futura.

Considerata la rilevanza dei formati dati **NeTEx** e **SIRI** a livello nazionale ed europeo, nonché la diffusione di **OpenTripPlanner** come piattaforma di riferimento per il calcolo dei percorsi multimodali, i risultati del progetto sono stati condivisi con **5T** e **NOI Techpark**, due realtà particolarmente attive e competenti in entrambi gli ambiti.

2. Converter

Questo capitolo fornisce un report del lavoro svolto per lo sviluppo dei converter da GTFS a NeTEx profilo italiano e da GTFS-RT a SIRI profilo italiano, e vengono discussi i mapping implementati confrontandoli con altre soluzioni esistenti. Il capitolo si conclude con un elenco di raccomandazioni sulla base dell'esperienza ottenuta nel progetto OTP-Italy e possibili sviluppi futuri.

2.1 Report sviluppo dei converter

Possono essere adottati diversi approcci per consentire la conversione tra vari formati di dati relativi alla mobilità. La soluzione open-source **Chimera**⁵ [1], basata su **Apache Camel**⁶, consente la trasformazione di dati e schemi attraverso un **approccio low-code alla conversione dell'informazione** [2], che separa la definizione delle regole dal software di conversione, migliorando così la scalabilità e la manutenibilità delle soluzioni di conversione.

Il framework Camel supporta un'integrazione semplice dei convertitori configurando pipeline di integrazione dei dati che sfruttano i componenti disponibili, senza richiedere uno sviluppo dedicato. Il progetto OTP-Italy, sulla base dell'analisi preliminare eseguita nei primi mesi di progetto, ha applicato questo approccio per la definizione di due convertitori: (i) **da GTFS al profilo italiano di NeTEx**, e (ii) **da GTFS-RT al profilo italiano di SIRI**. I convertitori sono resi disponibili in maniera open-source su GitHub con il nome **Convergo** al seguente link: <https://github.com/cefriel/convergo>. La licenza applicata è **EUPL v1.2**⁷.

⁵ <https://github.com/cefriel/chimera>

⁶ <https://camel.apache.org/>

⁷ EUPL v1.2 <https://eupl.eu/>

L'**ontologia TANGENT** [3] è stata utilizzata per supportare i mapping concettuali tra i formati considerati. Tuttavia, per garantire prestazioni migliori, abbiamo optato per l'implementazione di pipeline che non serializzino esplicitamente una rappresentazione RDF, ma generino direttamente il formato di output desiderato. L'abilitazione di una rappresentazione RDF intermedia esplicita, come supportato da Chimera, può offrire un vantaggio quando le stesse informazioni devono essere mappate su diversi standard di destinazione e/o quando il grafo RDF viene utilizzato per la fusione di dati provenienti da più sorgenti mappate sullo stesso modello concettuale.

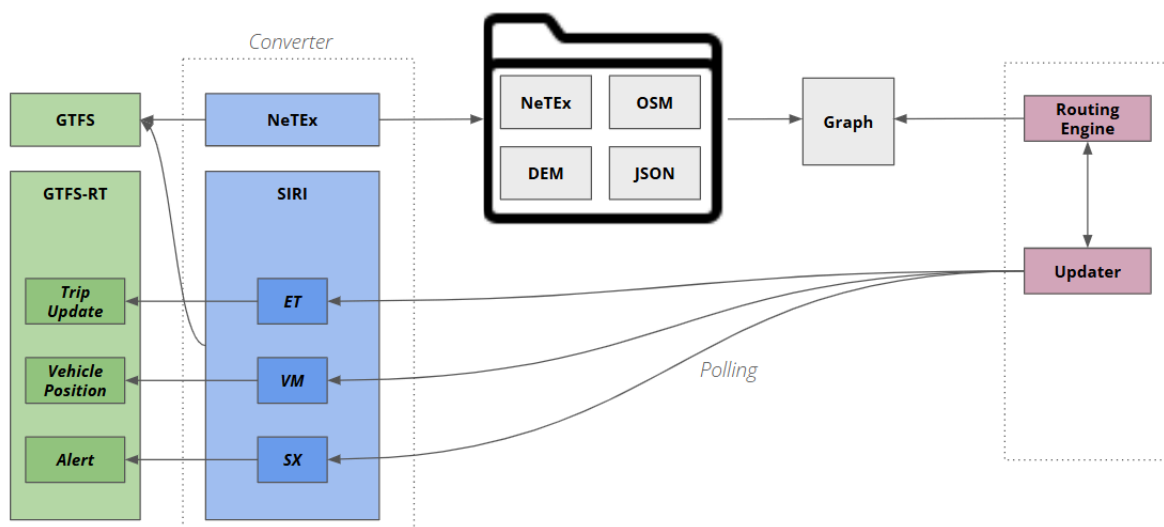
Le **pipeline Chimera** sviluppate **sono rese disponibili come unico componente software eseguibile come OCI container**⁸ tramite un apposito runtime. Il container può essere ottenuto dal codice sorgente o direttamente dal Package Registry di GitHub. **Il componente espone un insieme di API REST** per la conversione e la generazione di dati nei formati NeTEx e SIRI. Le API sono **documentate su GitHub tramite specifica Open API**. Il componente richiede un'opportuna configurazione ma permette di supportare più operatori contemporaneamente. Maggiori indicazioni per l'utilizzo del componente sono fornite nel README della repository GitHub.

Per la conversione NeTEx, che utilizza i feed GTFS come input, è stata implementata un'API asincrona a causa dei tempi di elaborazione elevati. Questa API consente agli utenti di avviare il processo di conversione e di configurare il formato di output (un singolo file contenente tutti i **Frame NeTEx** oppure più file organizzati per tipo di Frame o linea di servizio). Una seconda API consente di monitorare lo stato del processo e, una volta completato, fornisce un pacchetto compresso con i dati NeTEx convertiti.

La conversione SIRI, che utilizza dati GTFS-RT come input, è gestita tramite un'API sincrona, poiché il tempo di esecuzione è adeguato per una trasformazione in tempo reale. È

⁸ <https://opencontainers.org/>

implementata come tre pipeline che forniscono le diverse funzionalità SIRI del profilo italiano: **Situation Exchange (SX)**, **Vehicle Monitoring (VM)**, **Estimated Timetable (ET)**. Inoltre, è possibile richiedere tramite API un output SIRI arricchito considerando i dati statici GTFS. In questo caso, l'output viene generato combinando i dati dinamici dell'endpoint GTFS-RT con quelli del feed statico GTFS. Questa opzione non è prevista di default in quanto introduce una latenza aggiuntiva specialmente nel caso in cui sia necessario accedere a un feed GTFS di grandi dimensioni.



I dati NeTeX vengono forniti in fase di configurazione e utilizzati per costruire il grafo del trasporto pubblico all'interno di OTP, rappresentando un passaggio fondamentale per l'abilitazione del motore di calcolo dei percorsi. I flussi di dati SIRI, invece, sono configurati per essere consumati in modo continuo: il motore di calcolo interroga regolarmente le API per ottenere informazioni in tempo reale.

2.2 Report sviluppo dei mapping

L'**attività di mapping** svolta come parte dell'analisi preliminare è stata ulteriormente approfondita nel corso del progetto per l'implementazione dei mapping tra i formati GTFS/GTFS-RT e i profili italiani di NeTEx e SIRI nei converter. In questa sezione riportiamo i **macro-risultati ottenuti**, evidenziando le **potenziali aree per successivi sviluppi**. Le specifiche di regole di mapping sono disponibili in formato **Mapping Template Language (MTL)**⁹ nella repository del converter¹⁰ e possono essere facilmente modificate senza dover intervenire sul codice.

Per la generazione di NeTEx secondo le linee guida del profilo italiano:

- **Livello 1:** informazioni su nodi di accesso delle fermate, calendario di operatività, topologia della rete e itinerari/linee, orari di operatività, operatori dei trasporti. Il mapping implementato permette di ottenere la maggior parte dell'informazione richiesta considerando i file principali del feed GTFS (feed_info.txt, agency.txt, stops.txt, routes.txt, trips.txt, stop_times.txt, calendar.txt, calendar_dates.txt, shapes.txt). Il convertitore implementato supporta anche i files levels.txt, transfers.txt e pathways.txt per una migliore descrizione delle stazioni e fermate. Il contenuto informativo presente e/o esprimibile in GTFS risulta carente riguardo la descrizione degli operatori dei trasporti (e.g., ResourceFrame - organisations - Authority e ResourceFrame - organisations - Operator) e delle reti di trasporto (ServiceFrame - Network e ServiceFrame - AdditionalNetworks).
- **Livello 2:** informazioni su contratti di servizio, tipologie di veicoli e servizi a bordo. Il mapping implementato è limitato ad alcuni campi delle entità del Livello 1. Le entità principali aggiunte dal Livello 2 non sono coperte da GTFS (JourneyAccounting, ResponsibilitySet, GroupOfOperators, Parking, VehicleTypes, VehicleModels,

⁹ [https://github.com/cefriel/mapping-template/wiki/Mapping-Template-Language-\(MTL\)](https://github.com/cefriel/mapping-template/wiki/Mapping-Template-Language-(MTL))

¹⁰ <https://github.com/cefriel/convergo/tree/main/src/main/resources/templates>

Vehicles, StopAreas) a meno di non considerare estensioni custom della specifica (per esempio GTFS-PLUS¹¹ permette la descrizione dei veicoli) e non sono ottenibili tramite il converter.

- **Livello 3:** informazioni su tariffe dei parcheggi, titoli di viaggio, tariffe e zone tariffarie, punti e piattaforme di vendita, classi di passeggeri. Mapping implementato considerando le informazioni presenti in GTFS secondo la specifica Fares v1. La specifica Fares v2 risulta infatti ancora poco utilizzata in quanto non è ancora stabile e molti campi sono segnalati come sperimentali. Le informazioni presenti in GTFS permettono di modellare solo alcune tipologie di tariffe, mentre NeTEx offre maggiore espressività. Inoltre, molte delle entità definite nel NeTEx profilo italiano risultano ridondanti come risultato del mapping in quanto GTFS non introduce distinzioni specifiche per alcuni di questi concetti (e.g., FareStructureElement, ValidableElement e FareTable sono mappati genericamente su un singolo fare_id).
- **Livello 4:** informazioni su servizi di mobilità alternativa riguardo alla dotazione e alle caratteristiche della flotta dei veicoli. Il livello 4 non risulta mappato dal converter in quanto non esprimibile in GTFS ma tramite GBFS¹².
- **Livello 5:** informazioni di accessibilità su fermate, stazioni, veicoli, modalità di acquisto e prenotazione. Queste informazioni sono estremamente limitate in GTFS e riguardano la possibilità di accesso per sedie a rotelle a fermate e veicoli. Le informazioni disponibili sono mappate dal converter in NeTEx come definito dal livello 5.

I mapping implementati hanno reso necessario utilizzare alcune entità NeTEx non considerate attualmente dal profilo italiano per migliorare l'esperienza utente in Open Trip Planner. Queste entità sono discusse nella sezione 2.4 come raccomandazioni per il profilo NeTEx italiano.

¹¹ <https://github.com/osplanning-data-standards/GTFS-PLUS>

¹² <https://gbfs.org/>

Per la generazione di SIRI secondo le linee guida del profilo italiano:

- **SIRI-VM (Vehicle Monitoring):** informazioni sulla posizione dei mezzi in tempo reale, gli orari di passaggio effettivi o previsti dei mezzi a specifiche fermate, eventuale ritardo e stato occupazionale del mezzo. Il mapping implementato permette di ottenere la maggior parte dell'informazione richiesta dal messaggio GTFS-RT VehiclePosition. Tuttavia, per alcuni campi è stato necessario reperire l'informazione da messaggi GTFS, per esempio, il nome della fermata a cui si sta facendo riferimento e l'orario programmato di partenza del mezzo dalla fermata.
- **SIRI-ET (Estimated TimeTable):** informazioni sugli orari di partenza/arrivo da/a una fermata registrati durante lo svolgimento di una corsa o frutto di previsione, informazioni sui passeggeri saliti, discesi, a bordo e sui sedili estraibili/posti per carrozzine/posti riservati usati per ciascuno dei transiti in una fermata, e percentuale di occupazione del mezzo. Il mapping implementato considera il messaggio GTFS-RT TripUpdate. Anche in questo caso, tuttavia, è stato necessario estrarre alcuni campi da GTFS, mentre non è stato possibile reperire informazioni per altri campi (come le informazioni sui passeggeri e sui posti riservati). Per quanto riguarda lo stato occupazionale del mezzo, in SIRI viene espresso come percentuale, mentre in GTFS-RT viene fornita una lista di valori discreti (per esempio, *empty*, *few_seats_available*, etc.). Essendo un campo sperimentale della specifica GTFS-RT, ed essendo difficile e soggettivo mappare valori discreti su una percentuale, non è stato implementato il mapping per questa informazione.
- **SIRI-SX (Situation Exchange):** informazioni sugli eventi che hanno un impatto sulla regolarità dei servizi, quali lo stato dell'evento, la causa, linee/corse/fermate impattate, conseguenze (data e ora, gravità, se è impedita o meno la discesa/salita dei passeggeri, ritardo). Il mapping implementato reperisce solo una parte di queste informazioni dal messaggio GTFS-RT Alert. È stato possibile estrarre alcune delle informazioni non rappresentate da GTFS, e sono state definite due mappe per i

valori previsti nei due standard relativamente alla causa dell'evento (AlertCause) e alla gravità (Severity). Le informazioni che non sono state invece valorizzabili tramite GTFS/GTFS-RT, e che rimangono quindi escluse dal mapping, sono le seguenti: il tipo di sorgente (es, phone, fax, web) e l'azienda che ha fornito i dati, lo stato dell'evento, la data/ora di inizio/fine delle conseguenze dell'evento, il tipo di ritardo e la relativa durata (nel caso in cui l'evento causi un ritardo), se l'evento impedisce o meno la discesa/salita di passeggeri dal mezzo).

- **SIRI-FM (Facility Monitoring):** informazioni sui parcheggi, sugli stalli di ricarica dei veicoli elettrici, sulle biciclette di bike sharing e sugli stalli per biciclette, nello specifico lo stato occupazionale e la disponibilità con relativo conteggio/percentuale in base allo stato. Questo messaggio non ha nessuna corrispondenza in GTFS-RT, per cui il mapping non è stato implementato. Alcune delle informazioni relative ai servizi di sharing mobility sono modellate in GBFS, che non è stato però oggetto di mapping.

2.3 Confronto con soluzioni esistenti

Di seguito riportiamo un **confronto con il convertitore rilasciato da Liguria Digitale**¹³ sottolineando le principali differenze con i convertitori rilasciati dal progetto OTP-Italy. Il convertitore di Liguria Digitale rappresenta, a nostra conoscenza, l'unica soluzione open-source esistente per il mapping verso i profili italiani di NeTEx e SIRI.

Dal punto di vista tecnico, l'approccio alla conversione implementato dai convertitori sviluppati in OTP-Italy si differenzia in quanto definisce le regole di mapping tramite un approccio a template configurabile [2] senza dover necessariamente intervenire nella codebase.

La **principale funzionalità aggiuntiva risiede nel supporto per GTFS-RT** e nella sua conversione verso il profilo italiano di SIRI. Considerando invece il mapping tra GTFS e NeTEx profilo italiano emergono le seguenti differenze:

¹³ <https://github.com/Liguria-Digitale/GTFS2NeTEx-converter>

- per il file stops.txt vengono gestite tutte le tipologie di location_type (mappate in NeTex con Quay, StopPlace, StopPlaceEntrance, BoardingPosition, PathJunction);
- vengono gestiti ulteriori file per la descrizione delle stazioni: levels.txt, pathways.txt, transfers.txt;
- è supportato il mapping delle GTFS Fares v1 verso il livello 3 del NeTex tramite i file fare_attributes.txt e fare_rules.txt;
- è possibile mappare file GTFS con più records nel file agency.txt (e.g., relativi a più operatori). Eventuali configurazioni aggiuntive possono essere fornite per impostare identificativi (e.g., VAT);
- è possibile mappare feed GTFS senza il file feed_info.txt ma è necessario fornire una configurazione minima del converter;
- il campo shape_dist_traveled sia del file stop_times.txt che del file shapes.txt viene valorizzato applicando la formula degli haversines alle coordinate dei punti consecutivi.

2.4 Raccomandazioni e sviluppi futuri

In questa sezione riportiamo una **serie di raccomandazioni** collezionate all'interno del progetto OTP Italy per **migliorare ed evolvere i profili italiani di NeTex e SIRI**. Le raccomandazioni sono il risultato dell'analisi delle specifiche dei profili per effettuare l'attività di mapping, dell'applicazione del converter a feed GTFS e GTFS-RT di operatori italiani e del confronto con dataset NeTex disponibili tramite il National Access Point italiano (dataset SIRI non disponibili attualmente tramite il NAP). In particolare, vengono evidenziate **raccomandazioni considerando l'obiettivo di utilizzo dei dati convertiti in OTP**.

Corrispondenza identificativi

- Problema 1: il profilo SIRI richiede esplicitamente l'utilizzo della partita IVA come identificativo operatore. Il profilo NeTex italiano non lo indica esplicitamente e

questo può creare una mancata corrispondenza quando si integrano i dati NeTEx e SIRI in OTP.

- Problema 2: i dati SIRI possono fare riferimento a identificativi definiti all'interno dei dati NeTEx (e.g., stopPointRef in SIRI si riferisce a uno ScheduledStopPoint in NeTEx). L'applicazione di diverse strategie per determinare gli identificativi (e.g., technical identifier) può creare una mancata corrispondenza quando si integrano i dati NeTEx e SIRI in OTP.
- Problema 3: operatori diversi possono utilizzare diverse strategie per definire identificativi. Per questo motivo, diversi identificativi possono riferirsi alla stessa entità (e.g., diversi operatori che utilizzano le stesse fermate) non permettendo una riconciliazione in OTP.
- **Soluzione a breve termine:** aggiornare e allineare le indicazioni per gli identificativi nei profili NeTEx e SIRI italiani. Prevedere in SIRI un'indicazione esplicita degli identificativi che fanno riferimento a entità NeTEx.
- **Raccomandazione (R1):** definire una linea guida per la gestione degli identificativi separatamente dai due profili NeTEx e SIRI per evitare problemi di disallineamento tra le linee guida dei due profili.
- **Raccomandazione (R2):** creare una strategia nazionale (o su base regionale) per l'assegnazione degli identificativi (e.g., delle fermate) in modo da poter integrare in OTP anche dati di diversi operatori.

Ordinamento elementi per validazione

- Problema: la validazione tramite XSD dei dati NeTEx/SIRI segnala errori nel caso i campi non siano nell'ordine specificato dall'XSD. I profili italiani di NeTEx e SIRI non sempre riportano i campi nell'ordine definito dagli XSD (es. tariffZones in ScheduledStopPoint) e per questo motivo l'utente può ottenere NeTEx non validi.
- **Raccomandazione (R3):** Se si mantiene l'XSD con ordine specifico degli elementi, rivedere le specifiche dei profili per garantire allineamento. Valutare altrimenti la

rimozione di alcuni vincoli di ordinamento nell'XSD e documentare esplicitamente nei profili dove l'ordine ha significato semantico (es. FareStructureElementInSequence).

CompositeFrame NeTex e rappresentazione multi-operatore

- Problema 1: in caso di dati multi-operatore non è specificato nel profilo italiano NeTex se produrre un CompositeFrame per ogni operatore o un CompositeFrame contenente i dati di tutti gli operatori.
- Problema 2: il profilo italiano NeTex menziona tre tipologie di CompositeFrame (EU_PI_LINE_OFFER, EU_PI_NETWORK_OFFER, EU_PI_STOP_OFFER) tuttavia mancano indicazioni specifiche su quali Frame devono appartenere ad un CompositeFrame di una data tipologia e non è chiaro se le tipologie siano in alternativa o complementari (i.e., se necessario avere più CompositeFrame di diverse tipologie).
- **Raccomandazione (R4):** specificare ulteriormente questi aspetti nel profilo italiano NeTex.

Integrazione di informazioni aggiuntive nel profilo NeTex

- Problema 1: I codici visivi per rappresentare le linee sono supportati da GTFS (file routes.txt) e NeTex (Presentation - Colour / TextColour) ma non indicati nel profilo italiano NeTex. Questi campi sono di grande valore per lo sviluppo di interfacce grafiche di trip planning per l'utente finale e devono per questo motivo essere forniti in input a OTP.
- Problema 2: il profilo italiano NeTex non copre elementi GTFS che permettono di descrivere le stazioni e che risultano utili per OTP. La specifica NeTex permette di descrivere queste entità come StopPlaceEntrance (file stops.txt con location_type=2), BoardingPosition (location_type=4), PathJunctions (location_type=3) e PathLink (file pathways.txt).

- Problema 3: GTFS (file stop_times.txt) e NeTEx (StopPointInJourneyPattern - changeOfDestinationDisplay) prevedono la possibilità di rappresentare un cambio nel display di destinazione durante un servizio, ma questa casistica non è coperta dal profilo NeTEx italiano. Questo campo può essere utile per coprire casistiche del tipo capolinea intermedio, linea circolare, linea con diramazione, linea multi-tratta.
- Problema 4: il profilo italiano NeTEx prevede per un ServiceJourney l'indicazione del DepartureTime ma non ArrivalTime. Per alcune applicazioni può essere utile avere accesso a entrambe le informazioni in modo da dover evitare di accedere alla lista dei passingTimes per ottenere il valore.
- Problema 5: il profilo italiano NeTEx non permette di specificare informazioni di dettaglio sugli interscambi (ServiceJourneyInterchange) come la durata del trasferimento (MinimumTransferTime) o la non necessità di cambiare veicolo (StaySeated).
- Problema 6: il profilo italiano NeTEx non permette di specificare i ContactDetails per DistributionChannel (es. per indicare URL per acquisto biglietti online come reso possibile nel file agency.txt di GTFS).
- Problema 7: il profilo italiano NeTEx non permette la rappresentazione di tariffe con un numero massimo di trasferimenti ammessi che sono supportate da GTFS. La specifica NeTEx supporta queste casistiche tramite le limitazioni (FareStructureElement - limitations) di Interchanging - MaximumNumberOfInterchanges e FrequencyOfUse.
- **Raccomandazione (R5):** Valutare la possibilità di estendere il profilo italiano NeTEx per supportare le casistiche riportate.

Dimensione dei file NeTEx

- Problema 1: I file NeTEx che seguono le indicazioni del profilo italiano possono raggiungere dimensioni elevate, diventando difficili da ispezionare da parte degli

sviluppatori in caso di errori di validazione e/o da parte di terzi come esempi di utilizzo della specifica.

- Problema 2: i dataset NeTEx caricati sul NAP italiano contengono file NeTEx di grandi dimensioni e non è possibile scaricare informazioni limitatamente a specifici operatori.
- **Raccomandazione (R6):** definire delle linee guida per la suddivisione dei dati NeTEx in file di dimensioni ridotte (es. approccio del Nordic NeTEx Profile¹⁴ che raccomanda un file per le entità comuni e un file per ciascuna linea).

Definizione Livello 3 NeTEx (Fares Frame)

- Problema 1: Complessità delle entità NeTEx per modellare le tariffe è molto diversa dall'approccio semplificato adottato in altri standard come GTFS. Il mapping delle informazioni di GTFS a NeTEx richiede l'istanziamento di molteplici entità che sono necessarie per una corretta modellazione in NeTEx. Abbiamo riscontrato un utilizzo eterogeneo delle entità indicate nel profilo NeTEx italiano nei dataset di Livello 3 attualmente disponibili nel NAP.
- **Raccomandazione (R7):** introdurre nel profilo NeTEx italiano una sezione di overview dei vari concetti del Fares Frame NeTEx e delle loro relazioni facendo riferimento a esempi concreti di tariffe e loro modellazione (parzialmente disponibili ma come files XML su GitHub¹⁵).
- Problema 2: L'utilizzo dell'entità FareStructureElementInSequence è raccomandata in NeTEx solo se la tariffa preveda dei vincoli relativi all'ordine di validazione dei titoli associati (es. Park&Ride richiede prima ingresso al parcheggio e poi al trasporto pubblico) mentre è utilizzata indistintamente nel profilo NeTEx italiano

¹⁴ <https://enturas.atlassian.net/wiki/spaces/PUBLIC/pages/728891481/Nordic+NeTEx+Profile>

¹⁵ <https://github.com/5Tsr/netex-italian-profile/tree/main/Examples>

- **Raccomandazione (R8):** modificare il profilo NeTEx italiano per indicare l'utilizzo di FareStructureElementInSequence solo per casi specifici in cui la sequenza deve essere posta come vincolo.

Modellazione Servizi Frequency-Based e On-Demand in NeTEx

- Problema: necessità di rappresentare servizi on-demand (area-based), analoghi a quelli descritti da GTFS-Flex¹⁶, e frequency-based, modellabili in GTFS tramite il file frequencies.txt.
- **Raccomandazione (R9):** valutare l'aggiornamento del profilo NeTEx italiano per fornire linee guida su come modellare queste tipologie di servizi.

Il **presente report è già stato condiviso con 5T¹⁷**, coordinatore italiano dei tavoli di lavoro per i profili italiani di NeTEx e SIRI, e con il **NOI Techpark¹⁸**, esperto di dominio coinvolto nel progetto. Le modifiche minori verranno implementate all'interno delle prossime release delle specifiche. Altre raccomandazioni (e.g., relative alla modellazione delle Fares) possono beneficiare invece di un coordinamento a livello di profilo NeTEx europeo e verranno valutati da 5T e Cefriel nell'ambito del **progetto NAPCORE-X** in cui sono coinvolti insieme al Ministero delle Infrastrutture e dei Trasporti (MIT). Inoltre, il NOI Techpark sta effettuando una sperimentazione con dati reali per supportare il **trasporto pubblico on-demand** sulla base del profilo NeTEx tedesco, e i risultati potranno essere considerati come input per lo sviluppo del profilo italiano di NeTEx. Infine, **ulteriori sviluppi dei converter per supportare dati** non presenti in GTFS (e.g., sharing data da GBFS) sono da considerarsi come input per successive attività progettuali.

Dopo la fine del progetto, **i converter sviluppati in OTP-Italy resteranno disponibili in maniera open-source** tramite il repository su GitHub. Eventuali contributi alla repository verranno valutati e possibilmente integrati nella repository principale secondo le condizioni

¹⁶ <https://github.com/google/transit/pull/433>

¹⁷ <https://www.5t.torino.it/>

¹⁸ <https://noi.bz.it/>

del Contributor License Agreement (CLA). Cefriel si impegna ad identificare nuove forme di finanziamento per continuare a fornire supporto e manutenzione dei converter.

3. Miglioramento ed ottimizzazione dei dati in formato GTFS e GTFS-RT

Oltre alle attività di sviluppo dei due **converter** e alla creazione del **fork di OpenTripPlanner (OTP)**, una parte centrale del progetto ha riguardato l'**analisi approfondita dei dataset GTFS** attualmente in uso e l'**implementazione di interventi mirati al miglioramento della loro qualità e completezza**. L'obiettivo di tali attività è stato quello di **potenziare l'affidabilità, la coerenza e la ricchezza informativa dei dati di trasporto pubblico**, al fine di offrire all'utente finale informazioni più accurate, dettagliate e coerenti con la realtà operativa dei servizi.

Le azioni intraprese hanno riguardato diversi ambiti del modello GTFS, tra cui:

- la **generazione automatica dei file *shapes.txt***, per garantire una rappresentazione geometrica realistica dei percorsi svolti dai mezzi di trasporto;
- l'**utilizzo del campo *parent_station*** per il corretto raggruppamento delle fermate fisicamente contigue in nodi o stazioni comuni;
- l'**adozione dell'estensione *route_type_extension*** per una classificazione più precisa dei servizi e dei mezzi di trasporto;
- la **transizione dal modello Fares V1 al modello Fares V2**, finalizzata a una gestione più articolata e interoperabile delle tariffe;
- l'**integrazione dei file *pathways.txt* e *levels.txt***, per descrivere in modo dettagliato le connessioni pedonali e la struttura verticale delle stazioni e degli hub di interscambio.

Queste attività, sviluppate in modo coordinato, hanno contribuito a **innalzare significativamente la qualità semantica, spaziale e funzionale dei feed GTFS**, rendendoli più adatti all'utilizzo in contesti avanzati di pianificazione multimodale,

conversione verso standard europei (NeTeX, Transmodel) e integrazione all'interno di ecosistemi MaaS.

3.1 Creazione di file *shapes.txt*

Durante la pianificazione di un viaggio, il motore **OpenTripPlanner (OTP)** restituisce, per ciascun itinerario calcolato, una **polyline** del percorso, ossia una catena poligonale, costituita da una sequenza ordinata di punti geografici. Collegati tra loro, questi punti delineano graficamente il tragitto effettivamente seguito dal mezzo lungo la rete stradale o ferroviaria. Affinché la rappresentazione sia precisa e fedele alla realtà, è indispensabile che il feed GTFS di input contenga un file ***shapes.txt*** completo, coerente e allineato con la topologia della rete. Il file *shapes.txt* descrive, per ogni *trip*, l'insieme ordinato delle coordinate (latitudine e longitudine) che costituiscono il **tracciato effettivo del servizio**. Tale informazione è fondamentale per assicurare che gli strumenti di routing possano mostrare il percorso reale seguito dal mezzo e non una semplice interpolazione geometrica tra le fermate.

Tuttavia, **non tutti i feed GTFS pubblicati dagli operatori includono il file *shapes.txt***, o lo includono in forma incompleta o incoerente. In assenza di queste informazioni, **OpenTripPlanner** è costretto a visualizzare i percorsi come **linee rette tra le fermate**, con una conseguente perdita di realismo nella rappresentazione e un potenziale impatto negativo sull'esperienza utente e sulla leggibilità dei risultati. Questo problema si accentua nei contesti urbani complessi o nei percorsi che seguono infrastrutture con andamento curvilineo (ad esempio linee ferroviarie, tratte collinari o costiere).

Per ovviare a tale criticità, all'interno della **pipeline di elaborazione e conversione GTFS-NeTeX** è stato implementato un **modulo di generazione automatica delle geometrie (shapes)** nei casi in cui il file originale risulti mancante o incompleto. A questo scopo, è stato impiegato **Pfaedle**, un software **open source** progettato per ricostruire il file *shapes.txt* a partire da:

- il feed GTFS disponibile (fermate, sequenze dei viaggi, route identifiers);
- i dati **OpenStreetMap (OSM)**, che forniscono la rete stradale o ferroviaria di riferimento.

Il processo si basa su un **algoritmo di map matching**, che confronta la sequenza delle fermate del *trip* con la rete OSM, individuando il percorso più plausibile che le collega. Il risultato è un tracciato continuo e coerente con la topologia reale, che viene salvato come file *shapes.txt* associato ai corrispondenti *trip_id*. Questo approccio consente di ottenere una **rappresentazione geometrica ad alta fedeltà**, migliorando sensibilmente la qualità visiva e informativa dei percorsi, nonché la consistenza del feed GTFS nei successivi processi di conversione in formato NeTEx. Ad esempio, L'immagine seguente, tratta dalla documentazione ufficiale del progetto **Pfaedle** su GitHub¹⁹, illustra graficamente il funzionamento dell'algoritmo di generazione delle geometrie a partire dalle sequenze di fermate e dai dati OSM.



Figura 3: Caso pratico di applicazione di geometrie (*shapes*)

In conclusione, l'integrazione di questa fase nella pipeline di elaborazione ha permesso di **colmare una delle principali lacune dei feed GTFS forniti dagli operatori**, garantendo una **maggiore qualità cartografica**, una **migliore leggibilità dei percorsi** e una **coerenza geometrica ottimale** per l'utilizzo dei dati nei sistemi MaaS e negli strumenti di pianificazione multimodale.

¹⁹ <https://github.com/ad-freiburg/pfaedle>

3.2 Raggruppamento di fermate tramite *parent_station*

Nel modello dati **GTFS (General Transit Feed Specification)**, il campo *parent_station* del file *stops.txt* consente di **stabilire relazioni gerarchiche tra fermate (stops) e stazioni (stations)**, permettendo di raggruppare più punti di arresto fisicamente vicini sotto un nodo comune. Questa struttura gerarchica è essenziale per rappresentare correttamente **aree di interscambio multimodale** come stazioni ferroviarie, terminal autobus, fermate con banchine su entrambi i lati della carreggiata o hub condivisi tra diversi operatori.

Dal punto di vista funzionale, la corretta valorizzazione del campo *parent_station* migliora notevolmente la **coerenza semantica e visiva** del dataset GTFS, consentendo di:

- rappresentare con maggiore realismo la **topologia del nodo di trasporto**, distinguendo tra ingressi, banchine e aree comuni;
- **semplificare la visualizzazione cartografica**, mostrando ad esempio soltanto le stazioni “parent” nei livelli di zoom più bassi e i relativi “children stops” (banchine, fermate, accessi) nei livelli più dettagliati;
- facilitare i processi di **conversione verso formati di livello superiore**, come **NeTEx** o **Transmodel**, che si basano su relazioni esplicite tra “stop places” e “quays”;
- migliorare la **qualità della pianificazione multimodale** nei motori di routing (es. OpenTripPlanner), che possono così considerare correttamente le connessioni tra stop fisicamente contigui ma logicamente appartenenti alla stessa stazione.

All'interno della **pipeline di trasformazione GTFS-NeTEx** è stato pertanto introdotto un **passaggio specifico di aggregazione topologica delle fermate**, finalizzato a identificare automaticamente gruppi di stop geograficamente vicini e ad assegnare loro una *parent_station* comune. Questo processo si basa su una logica di **clustering spaziale**, che analizza le coordinate geografiche degli stop e applica soglie di prossimità (distanza massima) calibrate in funzione del contesto urbano o extraurbano. Una volta individuato

un gruppo coerente di fermate, il sistema genera una nuova entità *parent_station*, associandovi come “*children*” tutti gli stop che ricadono entro il raggio definito.

Per garantire una **rappresentazione topograficamente realistica e coerente**, la stazione principale (*parent_station*) viene posizionata **nel punto medio del gruppo di fermate**. A tale scopo, si utilizza il **centroide geometrico** del cluster, calcolato come media delle coordinate latitudinali e longitudinali delle fermate appartenenti al gruppo. Questa scelta consente di collocare la stazione virtuale nel baricentro del nodo di trasporto, migliorando sia la resa grafica sulle mappe, sia la coerenza con le successive trasformazioni in formati interoperabili.

Nel complesso, l'introduzione di questa logica di aggregazione e valorizzazione del campo *parent_station* ha portato a un **miglioramento tangibile della qualità strutturale dei feed GTFS**, garantendo una rappresentazione più fedele delle infrastrutture reali e un incremento della **precisione nei processi di visualizzazione, routing e interoperabilità** con altri standard del dominio della mobilità.

3.3 Implementazione di *route-type extension*

Un ulteriore ambito di intervento ha riguardato il **miglioramento della classificazione semantica delle rotte** all'interno dei feed GTFS, attraverso l'adozione, ove possibile, dell'estensione denominata **Route Type Extension**²⁰.

Il formato GTFS prevede, all'interno del file *routes.txt*, un campo denominato *route_type* che consente di specificare la modalità di trasporto associata a ciascuna rotta (ad esempio bus, treno, metropolitana, traghetto, funivia, ecc.). Tuttavia, la versione base dello standard risulta limitata a un set ristretto di categorie generiche, che non permettono di rappresentare in modo accurato la crescente varietà di servizi di trasporto oggi presenti nei contesti multimodali.

²⁰ <https://developers.google.com/transit/gtfs/reference/extended-route-types>

Per superare tali limitazioni, Google e il gruppo di lavoro internazionale sul GTFS hanno introdotto la **Route Type Extension**, un'estensione ufficiale che consente l'uso di un insieme ampliato di codici denominati **HVT (Hierarchy of Vehicle Types)**, derivati dallo standard europeo **TPEG (Transport Protocol Experts Group)**. Grazie a questa estensione, è possibile distinguere con maggiore precisione tra varianti e sottotipi di veicoli, garantendo una **maggiore precisione semantica** e una rappresentazione più aderente alla realtà del servizio. Alcuni esempi di applicazione includono:

- differenziazione tra **treni regionali, intercity e ad alta velocità**;
- distinzione tra **autobus urbani, suburbani o a lunga percorrenza**;
- classificazione specifica di mezzi alternativi come **traghetti, funivie, impianti di risalita o servizi a chiamata**.

L'introduzione di questa classificazione estesa comporta **benefici diretti sia in termini tecnici che funzionali**, tra cui:

- **Maggiore precisione semantica**: la possibilità di distinguere varianti di un medesimo mezzo consente una descrizione più accurata dell'offerta di trasporto e facilita l'analisi dei servizi per modalità o segmento operativo;
- **Migliore esperienza utente e visualizzazione**: applicazioni e strumenti di routing possono mostrare icone, colori o diciture coerenti con il tipo effettivo di mezzo, migliorando la leggibilità e la comprensione del viaggio da parte dell'utente finale;
- **Allineamento con standard europei (TPEG / HVT)**: l'adozione dei codici HVT consente di mantenere la compatibilità con altri standard di rappresentazione e scambio dati nel dominio del traffico e del trasporto, facilitando l'interoperabilità e l'integrazione con sistemi ITS e MaaS a livello europeo;
- **Supporto a logiche di filtraggio e pianificazione avanzata**: l'uso di route type estesi consente ai motori di pianificazione come **OpenTripPlanner (OTP)** di applicare regole specifiche per diverse categorie di servizio, ad esempio dando

priorità ai treni ad alta velocità nei risultati di ricerca o escludendo determinati servizi speciali dalle pianificazioni standard.

Nel contesto del progetto, **ove tecnicamente possibile e in presenza di dati sufficienti**, il team ha provveduto a **introdurre autonomamente la classificazione estesa delle rotte**, aggiornando il campo *route_type* con i relativi codici HVT coerenti con la tipologia di servizio erogato da ciascun operatore. Questa attività ha richiesto un'analisi puntuale dei dataset GTFS e una mappatura dei servizi rispetto alle definizioni TPEG, al fine di garantire uniformità di codifica e corretto allineamento tra operatori differenti. Ad esempio, nel caso del Trentino, la classificazione estesa è stata utilizzata per distinguere gli skibus (autobus invernali dedicati a località sciistiche) e le navette estive dagli autobus del trasporto pubblico di linea.

3.4 Transizione al modello Fares V2

Un'ulteriore area di significativa evoluzione nel corso del progetto ha riguardato la **qualità e la struttura dei dati tariffari**. In particolare, ove tecnicamente possibile, è stata condotta una **migrazione dal modello Fares V1²¹ al modello Fares V2²²**, con l'obiettivo di migliorare la granularità, la coerenza e l'interoperabilità delle informazioni tariffarie all'interno dei dataset GTFS (General Transit Feed Specification).

Il modello **Fares V2** rappresenta infatti un'evoluzione sostanziale rispetto alla precedente versione Fares V1, nata per superarne le principali **limitazioni strutturali e semantiche**. Dal punto di vista tecnico, Fares V2 si configura come un'estensione del **GTFS Schedule**, progettata per offrire una maggiore flessibilità nella rappresentazione delle **regole tariffarie**, delle **politiche di prezzo** e delle **condizioni di utilizzo dei titoli di viaggio**.

²¹ <https://gtfs.org/documentation/schedule/examples/fares-v1/>

²² <https://gtfs.org/community/extensions/fares-v2/>

In particolare, il nuovo schema consente di descrivere in modo più articolato elementi quali:

- **Abbonamenti e biglietti multiviaggio**, con la possibilità di definire la durata di validità, il numero di corse incluse e le relative condizioni d'uso;
- **Tariffe variabili in funzione delle fasce orarie**, delle zone percorse o della distanza, consentendo una modellazione più accurata dei sistemi di pricing dinamico;
- **Scontistiche e agevolazioni dedicate a specifiche categorie di utenza** (studenti, lavoratori sotto una certa età, persone con disabilità, ecc.);
- **Trasferimenti e integrazioni tariffarie tra operatori diversi**, favorendo la rappresentazione di schemi multimodali e interoperabili, elemento chiave nei contesti MaaS.

Nonostante i miglioramenti introdotti, il modello **Fares V2** presenta tuttora alcune **lacune funzionali e di copertura informativa**, che ne limitano parzialmente l'adozione come standard completo per la descrizione dell'intero ecosistema tariffario. Tra le principali criticità ancora riscontrate si evidenziano:

- **Assenza di informazioni strutturate sui canali di distribuzione** dei titoli di viaggio (ad esempio: vendite online, app mobile, rivenditori fisici, biglietterie automatiche);
- **Mancanza di dati relativi alle condizioni di vendita e di utilizzo**, quali termini di rimborso, validità minima e compatibilità con altre tariffe o promozioni;
- **Incapacità di rappresentare tariffe accessorie**, come quelle relative ai servizi di **parkeggio o park&ride**, che rappresentano un elemento sempre più rilevante nell'ambito della mobilità integrata.

L'attività di transizione al modello Fares V2 ha comunque consentito di **innalzare la qualità complessiva dei dataset** e di porre le basi per una futura estensione verso

modelli tariffari ancora più completi e interoperabili, in linea con gli standard emergenti del settore.

3.5 Integrazione dati Pathways e Levels

Un ulteriore ambito di intervento del progetto ha riguardato la **qualità e la completezza dei dati relativi alle connessioni pedonali interne alle stazioni e agli hub di interscambio multimodale**, elementi fondamentali per una pianificazione dei percorsi accurata e inclusiva.

Nel formato **GTFS (General Transit Feed Specification)**, i file *pathways.txt* e *levels.txt* consentono di descrivere, in relazione a **fermate create ad hoc e non direttamente associate alle rotte del trasporto pubblico**, i collegamenti fisici e logici tra diversi punti di accesso, banchine e spazi interni delle stazioni. Tali informazioni rivestono un ruolo essenziale nel garantire un'esperienza utente coerente con la realtà fisica, poiché permettono di rappresentare con precisione le modalità di percorrenza di ciascun segmento (ad esempio a piedi, tramite scale, rampe o ascensori) e di modellare la **verticalità degli spazi** attraverso la definizione di livelli, piani e sottopassi.

Questa componente dei dati è cruciale non solo per migliorare l'accuratezza della **pianificazione dei percorsi di accesso e uscita dalle stazioni**, ma anche per potenziare la **valutazione dell'accessibilità universale**, in particolare per utenti con mobilità ridotta o disabilità. Tuttavia, nella pratica, i feed GTFS pubblicati dalle agenzie di trasporto pubblico **raramente includono i file *pathways.txt* e *levels.txt***, e anche quando presenti, la loro efficacia dipende fortemente dalla **qualità e completezza dei dati OpenStreetMap (OSM)**²³ utilizzati per la costruzione del grafo multimodale della rete.

Tuttavia, anche dati provenienti da OSM incompleti o non aggiornati possono compromettere la capacità di motori di routing come **OpenTripPlanner (OTP)** di generare

²³ <https://www.openstreetmap.org/#map=6/42.09/12.56>

percorsi realistici e accessibili, soprattutto in contesti complessi come gli scambi ferro-gomma o le grandi stazioni ferroviarie a più livelli.

Per affrontare questa criticità, nel corso del progetto è stato avviato un **processo di miglioramento dell'attraversabilità pedonale** di diverse stazioni nelle aree del **Trentino e della Sardegna**, articolato in più fasi successive.

In una prima fase, l'attività si è concentrata sulla revisione e correzione dei dati OSM, intervenendo su nodi, archi e relazioni per garantire la creazione di **segmenti pedonali coerenti con la reale struttura fisica delle stazioni**. Nel caso del Trentino, tali interventi sono stati supportati da **rilievi diretti e conoscenze di campo** delle infrastrutture, mentre per la Sardegna l'attività si è basata su **analisi cartografiche e fotografie satellitari** (ad esempio tramite Google Maps o Bing Aerial Imagery), con l'obiettivo di correggere incongruenze nella connessione tra ingressi, banchine e sottopassi.

Completata la fase di allineamento topologico, l'attenzione si è spostata sui feed GTFS, con una serie di interventi mirati a **rafforzare la rappresentazione interna delle stazioni**:

- **Riposizionamento manuale degli stop ferroviari**, spesso collocati in origine all'esterno dell'edificio della stazione, con conseguente distorsione nei percorsi di accesso;
- **Creazione di stop intermedi** all'interno delle stazioni, per rappresentare in modo più realistico le connessioni tra ingressi, aree di attesa e banchine;
- **Popolamento del file *levels.txt***, con la definizione puntuale dei diversi livelli verticali (piani, banchine, sottopassi, passaggi sopraelevati, ecc.);
- **Compilazione del file *pathways.txt***, mediante la connessione logica tra gli stop interni e l'associazione di ciascun collegamento al tipo di percorso (a piedi, scale, scale mobili, ascensore), con indicazione delle relative caratteristiche di accessibilità.

Il livello di dettaglio aggiuntivo consente a **OpenTripPlanner** di elaborare **percorsi di accesso, transito e attraversamento delle stazioni** molto più aderenti alla realtà fisica,

migliorando significativamente la qualità della navigazione multimodale. Inoltre, l'informazione strutturata sui collegamenti verticali e sui tipi di pathway permette di **valutare automaticamente l'accessibilità per utenti con mobilità ridotta**, identificando, ad esempio, tragitti che evitano scale o prevedono l'uso esclusivo di ascensori. Tali miglioramenti rappresentano un passo importante verso una **modellazione completa e inclusiva dell'ambiente stazionale**, coerente con gli obiettivi di **universal design** e con i principi di accessibilità stabiliti a livello europeo (Direttiva UE 2019/882 – European Accessibility Act).

3.6 Raccomandazioni sulla gestione dei dati GTFS

Sebbene l'obiettivo primario del progetto fosse focalizzato sullo sviluppo dei **converter** e sull'evoluzione del **motore di pianificazione OpenTripPlanner (OTP)**, l'analisi approfondita condotta sui dataset GTFS attualmente in uso ha consentito di formulare alcune **osservazioni trasversali di carattere metodologico e strategico**, emerse come risultato collaterale ma di grande rilevanza per il futuro della gestione dei dati di trasporto pubblico.

In particolare, sono state individuate **due aree di miglioramento sistemico** che, pur non direttamente affrontate nel perimetro progettuale, rappresentano elementi chiave per garantire la qualità, l'affidabilità e la piena valorizzazione del formato GTFS. In primis, la necessità di uno strumento strutturato di controllo e validazione dei feed GTFS ed in secundis la necessità di un utilizzo più esteso e consapevole dello standard GTFS da parte degli operatori.

Con riferimento al primo punto, attualmente, gli operatori e le autorità di trasporto dispongono di strumenti eterogenei e frammentati per la verifica della correttezza formale e semantica dei propri dataset. Sebbene esistano soluzioni open source (come *GTFS*

*Validator*²⁴ sviluppato da MobilityData), tali strumenti non coprono in modo completo le esigenze operative legate ai feed più complessi o multi-operatore, né garantiscono un'integrazione efficace con pipeline di produzione o sistemi di quality assurance automatizzati. È pertanto emersa la necessità di sviluppare o adottare un **tool di validazione evoluto**, capace di verificare non solo la conformità sintattica ai requisiti dello standard, ma anche la **coerenza topologica, spaziale e tariffaria** (sebbene questi ultimi siano perlopiù assenti) dei dati, supportando processi di auditing continuo e miglioramento progressivo della qualità informativa.

Con riferimento al secondo punto, l'analisi dei feed raccolti ha evidenziato come molti operatori utilizzino solo parzialmente le potenzialità offerte dal formato GTFS, limitandosi ai file obbligatori e tralasciando elementi opzionali (come *pathways.txt*, *levels.txt*, *shapes.txt* o *fare_rules.txt*) che risultano invece essenziali per descrivere in modo completo l'offerta di trasporto. Una maggiore **maturità nell'adozione dello standard** permetterebbe di incrementare significativamente il valore informativo dei feed, migliorando la rappresentazione di aspetti quali accessibilità, tariffe integrate, interscambi e percorsi reali. Ciò consentirebbe di abilitare **nuovi livelli di interoperabilità**, di potenziare i sistemi di pianificazione e di garantire all'utente finale un'informazione più accurata e affidabile.

In conclusione, su questo tema, il lavoro svolto ha dimostrato che il potenziamento qualitativo dei dati GTFS non è solo un'esigenza tecnica, ma un passaggio strategico verso la costruzione di un **ecosistema informativo aperto, interoperabile e di alta qualità**, in grado di favorire la piena valorizzazione dei dati pubblici come infrastruttura digitale per la mobilità del futuro.

²⁴ <https://gtfs-validator.mobilitydata.org/>

4. Sviluppo del fork di OpenTripPlanner

Il secondo focus di progetto riguardava la **realizzazione di un fork di OpenTripPlanner (OTP)**, con l'intento di estendere e consolidare il supporto del motore di routing agli **standard europei NeTex e SIRI** per la pianificazione e la gestione integrata dei dati di trasporto.

Durante la fase iniziale di analisi del codice sorgente, è emerso che **OTP implementava già una parte significativa delle specifiche NeTex**, in particolare per quanto concerne i **livelli 1 e 2** dello standard, relativi alla **descrizione della rete di trasporto** e alla **programmazione oraria dei servizi**. Questa scoperta ha consentito di riconoscere che una base solida di interoperabilità era già presente, riducendo la necessità di un intervento radicale sul codice sorgente. A partire da questa constatazione, il lavoro si è orientato verso un'attività di **analisi tecnica approfondita** e di **ottimizzazione mirata** delle componenti già presenti, al fine di migliorare la leggibilità e la coerenza dei dati NeTex e SIRI all'interno di OTP. In particolare, sono stati condotti interventi di **revisione e affinamento dei moduli di importazione e parsing**, con l'obiettivo di garantire una corretta interpretazione dei dataset ibridi GTFS-NeTex e di ottimizzare i processi di conversione, validazione e caricamento nel grafo di rete. Questa fase di ottimizzazione ha permesso di incrementare la **robustezza del motore** nei confronti di formati eterogenei, migliorando la **consistenza semantica** e l'**efficienza dei processi di lettura dei dati**, e creando le premesse per un'integrazione più fluida dei feed multimodali.

Successivamente, l'analisi funzionale ha evidenziato che, pur disponendo di un supporto consolidato per i livelli 1 e 2 di NeTex, **OTP non include attualmente il livello 3 (FareFrame)**, dedicato alla gestione strutturata delle tariffe e delle regole di prezzo. Poiché tale componente è strettamente connessa al **cuore del motore di routing**, la sua implementazione richiederebbe un intervento profondo nella logica di calcolo dei costi, con un impatto significativo sulla complessità architetturale e sulla manutenibilità del software.

Un confronto diretto con la **community di sviluppo internazionale di OpenTripPlanner** (tramite riunioni dedicate su base bisettimanale con il team di sviluppo e su base mensile con il team di prodotto) ha confermato che la mancanza del supporto al livello 3 è **nota e documentata**, ma non rientra tra le **priorità di sviluppo a breve termine**. Alcuni contributori della community stanno esplorando soluzioni alternative attraverso **moduli esterni dedicati alla gestione tariffaria (fare engines)**, mantenendo così il nucleo principale del motore focalizzato sugli aspetti di routing e pianificazione.

Pur riconoscendo che l'integrazione nativa del livello tariffario rappresenta un'evoluzione auspicabile nel medio periodo, si è ritenuto **non opportuno destinare risorse progettuali significative** a tale implementazione, in considerazione del limitato interesse attuale della community e della complessità del refactoring necessario.

Alla luce di queste considerazioni, le attività sono state orientate verso **miglioramenti specifici della qualità dei dati GTFS** e verso **l'armonizzazione semantica e strutturale con gli standard NeTEx e SIRI**, garantendo così la piena compatibilità informativa dei dataset e ponendo le basi per future estensioni funzionali del motore di pianificazione.

5. Conclusioni

Il progetto **OTP-Italy** ha rappresentato un passo significativo verso il rafforzamento dell'interoperabilità e della qualità dei dati di trasporto pubblico in Italia, in linea con le direttive europee e italiane. Attraverso la collaborazione tra **OpenMove** e **Cefriel**, il progetto ha consentito di sviluppare un insieme di strumenti e metodologie open-source per: (i) facilitare la transizione dai formati **GTFS e GTFS-RT** agli standard europei **NeTEx e SIRI (profilo italiano)**, garantendo una maggiore coerenza semantica e una migliore integrazione tra sistemi eterogenei, e (ii) migliorare la qualità del dato e il servizio fornito agli utenti tramite servizi di trip planning.

Le principali attività di progetto hanno incluso:

- la progettazione e l'implementazione di **due converter** da GTFS a NeTEx (profilo italiano) e da GTFS-RT a SIRI (profilo italiano), basati su un approccio modulare e configurabile;
- un ampio lavoro di **analisi, arricchimento e ottimizzazione dei dati GTFS e GTFS-RT**, volto a migliorare la qualità, in vista della loro integrazione nei processi di pianificazione dei percorsi tramite OpenTripPlanner;
- lo sviluppo di un **fork dedicato di OpenTripPlanner**, mirato a ottimizzare la compatibilità con i formati europei e a migliorare la robustezza del motore di routing;
- la validazione delle soluzioni proposte con dati reali da tre territori principali, Trentino, Alto Adige e Sardegna.

L'approccio metodologico adottato ha permesso di evidenziare criticità strutturali e opportunità di miglioramento dei profili italiani di NeTEx e SIRI, formalizzate in una serie di **raccomandazioni tecniche** che potranno orientare l'evoluzione futura delle specifiche nazionali. Tali raccomandazioni riguardano, tra gli altri aspetti, l'allineamento dei criteri di

identificazione tra NeTEx e SIRI, la gestione di dataset multi-operatore, l'inclusione di informazioni aggiuntive utili ai motori di routing e la razionalizzazione della modellazione tariffaria (FareFrame).

Il progetto ha inoltre dimostrato come il miglioramento della qualità dei dati GTFS attraverso la generazione automatica degli shapes, una gestione gerarchica dei dati e la migrazione verso Fares V2, costituiscano un prerequisito essenziale per l'affidabilità dei sistemi di routing.

Infine, la decisione di rilasciare i converter come **software open-source** sotto licenza EUPL garantisce la massima trasparenza, riusabilità e sostenibilità dei risultati, favorendo il consolidamento di un ecosistema collaborativo per la gestione e la valorizzazione dei dati di mobilità in Italia.

In prospettiva, tali strumenti costituiscono la base per ulteriori sviluppi quali l'estensione ai dati di sharing mobility (GBFS) e ai servizi on-demand (GTFS-Flex) e per la progressiva integrazione dei diversi livelli informativi nel contesto della **mobilità intelligente, integrata e sostenibile**.

6. Bibliografia

[1] Grassi, M., Scrocca, M., Carenini, A., Comerio, M., Celino, I.: Composable Semantic Data Transformation Pipelines with Chimera. In: Proceedings of the 4th, International Workshop on Knowledge Graph Construction. CEUR Workshop Proceedings (2023).

[2] Scrocca, M., Carenini, A., Grassi, M., Comerio, M., Celino, I.: Not Everybody Speaks RDF: Knowledge Conversion between Different Data Representations. In: Proceedings of the 5th, International Workshop on Knowledge Graph Construction. CEUR Workshop Proceedings (2024).

[3] Scrocca, M., Grassi, M., Comerio, M., Carriero, V., Delgado Dias, T., Vieira Da Silva, A., Celino, I.: Intelligent Urban Traffic Management via Semantic Interoperability across Multiple Heterogeneous Mobility Data Sources. In Proceedings of the 23rd Intern. Semantic Web Conference, (2024). Pre-print version available at: <http://arxiv.org/abs/2407.10539>