

Open Data e trasporto ferroviario

Proposta di tirocinio interno

Marco Aceti

2 marzo 2023

Sommario

In Italia non esistono Open Data sulle performance del trasporto pubblico ferroviario: le metriche definite nei contratti di servizio tra gli enti locali committenti e le imprese ferroviarie sono insufficienti e spesso inaccessibili. La proposta di tirocinio si articola sull'idea di preservare i dati istantanei della circolazione ferroviaria dalla piattaforma ViaggiaTreno per produrre Open Data storici, *machine-readable* e di qualità. Infine, si propone un'analisi dei dati raccolti a fini statistici e di verifica.

1 Stato dell'arte

In Italia, il servizio di trasporto pubblico è operato da aziende¹ private o partecipate. Sul territorio nazionale sono autorizzate² una ventina di *Imprese Ferroviarie* (IF) adibite al trasporto passeggeri aventi in essere numerosi *Contratti di Servizio* (CdS) con gli enti locali (tipicamente le Regioni). La qualità del servizio è misurata da **metriche di performance** stabilite nei CdS e comunicate agli enti dalle IF.

1.1 Esempio: il servizio ferroviario lombardo

In Lombardia, Trenord S.r.l. definisce³ un *indice di puntualità entro i 5 minuti* che considera il “numero di corse circolanti giunte puntuali o con ritardo fino a 5 minuti”, ma esclude i “ritardi maturati per cause esterne” o “per lavori”. La Regione pubblica mensilmente un rapporto sulla puntualità dei treni⁴ in formato PDF, ma con diverse criticità:

- vengono considerati solo i ritardi in arrivo alla destinazione finale, escludendo quindi le stazioni intermedie;
- i dati forniti non sono granulari ma *brutalmente* aggregati per mese;
- sono escluse le *cause esterne* e le *circostanze occasionali*: gli indici di puntualità effettivi non sono pubblicati;
- i rapporti non rispettano neanche una *stella* dei livelli definiti da Tim Berners-Lee per valutare gli Open Data: non è nemmeno presente una licenza d'uso.

¹https://it.wikipedia.org/wiki/Aziende_di_trasporto_pubblico_italiane

²<https://www.mit.gov.it/documentazione/elenco-imprese-ferroviarie-titolari-di-licenza-1>

³<https://www.regione.lombardia.it/wps/wcm/connect/7144d5b9-7e3c-4e44-82ad-30a1652e2642/Contratto+Trenord+con+firme.pdf> – Allegato 11

⁴<https://www.regione.lombardia.it/wps/wcm/connect/4eae62eb-dfcf-4446-82ea-72dbdfdb2c4a/Puntualit%C3%A0.pdf>

C'è da considerare inoltre che Trenord (società tra l'altro partecipata al 50% da Regione Lombardia stessa) comunica al committente gli indici già calcolati, senza che quest'ultimo abbia modo di verificarli.

Infine, non tutti gli enti committenti pubblicano rapporti sulla qualità del servizio: per esempio, la Regione Campania prevede nel suo CdS⁵ con Trenitalia S.p.A. la fornitura di indici simili per il calcolo di penali e mitigazioni, ma non è reperibile nessun documento che li attesti.

1.2 Open Data storici

In conclusione, non esistono attualmente Open Data storici, completi, strutturati e *machine-readable* sul servizio di trasporto ferroviario in Italia. Gli indici di puntualità (e affidabilità) definiti nei CdS possono essere utili agli enti committenti per calcolare penali o comparare offerte di mercato, ma i Cittadini Digitali meritano una **maggiore trasparenza** per poter verificare autonomamente lo stato reale del *Sistema Ferrovia*.

2 Rilevazioni istantanee

Nella sezione precedente si è discusso di **dati storici**; la situazione è molto più rosea per i **dati in tempo reale**. Esistono innumerevoli siti web e applicazioni, ufficiali e non, che mostrano lo stato attuale di un treno in viaggio. L'app *Orario Treni*⁶ di Paolo Conte, per esempio, presenta con un'interfaccia molto semplice e intuitiva la possibilità di cercare treni per itinerario e numero, visualizzare arrivi e partenze di una stazione e consultare l'*andamento istantaneo* di un treno. Quest'ultimo è composto da informazioni come gli orari programmati ed *effettivi* di partenza e arrivo ad ogni fermata intermedia, ritardo cumulato fino a quel momento e luogo di ultimo rilevamento (non necessariamente corrispondente ad una fermata).

L'idea fondante della proposta in oggetto è sfruttare la ghiotta quantità di dati offerta dalle rilevazioni istantanee nel corso del tempo per produrre Open Data storici.

2.1 ViaggiaTreno

Il Gruppo Ferrovie dello Stato Italiane (*holding* di diverse società⁷ come Trenitalia, RFI, ANAS, ...) permette ai viaggiatori di trovare soluzioni di viaggio e visualizzare l'andamento di una corsa tramite la piattaforma web ViaggiaTreno⁸, similmente all'app *Orario Treni*. Si può infatti speculare che quest'ultima utilizzi proprio ViaggiaTreno come fonte dei dati.

2.1.1 API

Il *motore* dell'interfaccia web di ViaggiaTreno è un insieme di API "REST" non ufficialmente documentate e di scarsa qualità⁹. In rete sono presenti diversi tentativi di documentazione, mantenuti dalla community open source^{10, 11, 12}.

⁵<https://www.regione.campania.it/assets/documents/contratto-di-servizio-tpl-ferro.pdf>
sez. "Penali e forme di mitigazione delle stesse" – Allegato 7

⁶<https://www.orariotreniapp.it/>

⁷https://it.wikipedia.org/wiki/Ferrovie_dello_Stato_Italiane

⁸<http://www.viaggiatreno.it/infomobilita/index.jsp>

⁹<https://medium.com/@albigiu/trenitalia-shock-non-crederete-mai-a-queste-api-painful-14433096502c>

¹⁰<https://github.com/sabas/trenitalia>

¹¹<https://github.com/roughconsensusandruncode/TrainMonitor/wiki/API-del-sistema-Viaggiatreno>

¹²<https://github.com/Razorphyn/Informazioni-Treni-Italiani>

2.1.2 Copyright e licenza d'uso

Le note legali riportate sul portale ViaggiaTreno sono abbastanza aggressive.

I contenuti, la grafica e le immagini sono soggetti a Copyright. Ogni diritto sui contenuti (a titolo esemplificativo e non esaustivo: l'architettura del servizio, i testi, le immagini grafiche e fotografiche, ecc.) è riservato ai sensi della normativa vigente. I contenuti di ViaggiaTreno non possono, neppure in parte, essere copiati, riprodotti, trasferiti, caricati, pubblicati o distribuiti in qualsiasi modo senza il preventivo consenso scritto della società Trenitalia S.p.A.. È possibile scaricare i contenuti nel proprio computer e/o stampare estratti unicamente per utilizzo personale di carattere informativo. Qualsiasi forma di link al sito www.ViaggiaTreno.it deve essere preventivamente autorizzata¹³ e non deve recare danno all'immagine e alle attività di Trenitalia S.p.A.. è vietato il c.d. deep linking ossia l'utilizzo, su siti di soggetti terzi, di parti del Servizio Internet o, comunque, il collegamento diretto alle pagine senza passare per la home page del Servizio Internet. L'eventuale inosservanza delle presenti disposizioni, salvo esplicita autorizzazione scritta, sarà perseguita nelle competenti sedi giudiziarie civili e penali.

Il Gruppo Ferrovie dello Stato Italiane vieta formalmente ai soggetti non autorizzati l'utilizzo di ViaggiaTreno per fini diversi dal mero uso personale, riservando tutti i diritti sui contenuti. Nel 2019, l'applicazione Trenit! è stata costretta¹⁴ a interrompere il servizio in seguito a un processo giudiziario iniziato da Trenitalia, che contestava il riutilizzo dei dati sulla circolazione ferroviaria presenti su ViaggiaTreno. Il giudice nella sua sentenza¹⁵ ha invece stabilito che "la banca dati degli orari dei treni e i prezzi di questi, non è protetta da diritto d'autore" e quindi Trenit! li può utilizzare.

Ritengo quindi che non ci siano reali limiti legali nell'utilizzo della piattaforma ViaggiaTreno e in particolare delle sue API per i fini della proposta in oggetto.

2.2 Avvisi Trenord sulla circolazione

Trenord, oltre alla tracciabilità dei suoi treni in ViaggiaTreno, offre anche un servizio di avviso delle criticità di tutte le linee (simile all'InfoMobilità di Trenitalia). Gli avvisi sono rilasciati da esseri umani, ma hanno un formato simile. Di seguito ne sono riportati alcuni della linea Verona-Brescia-Milano¹⁶.

Criticità — 01/03/2023 06:24

Aggiornamento: Il treno 10913 (MILANO GRECO PIRELLI 05:52 - BRESCIA 07:12) sta viaggiando con un ritardo di 30 minuti perché è stato necessario prolungare i controlli tecnici che precedono la partenza del treno.

Criticità — 01/03/2023 10:07

Il treno 2624 (VERONA PORTA NUOVA 09:43 - MILANO CENTRALE 11:35) viaggia con 12 minuti di ritardo in seguito alla sosta prolungata di un altro treno della linea.

¹³L'autore di questo documento si dichiara reo del reato di *linking non autorizzato*

¹⁴<https://www.startmag.it/smartcity/perche-trenitalia-ha-tamponato-lapp-trenit-per-il-momento/>

¹⁵<https://www.startmag.it/innovazione/trenit-trenitalia/>

¹⁶<https://www.trenord.it/linee-e-orari/circolazione/le-nostre-linee/brescia-treviglio-milano/?code=R4>

3 Proposta operativa

La proposta si articola in tre fasi.

3.1 Indagine esplorativa

Come concordato a voce nello scorso colloquio, in questa fase potrei indagare più a fondo sullo stato degli Open Data nel trasporto ferroviario in Italia e negli altri Paesi europei. Progetti simili potrebbero influenzare positivamente scelte come la granularità e il formato dei dati.

3.2 Raccolta e produzione degli Open Data

Lo scopo di questa fase è progettare e implementare uno strumento che raccoglie i dati in tempo reale dal portale ViaggiaTreno di **tutti i treni** in circolazione e li salva in un database. Quindi, creare successivamente un altro strumento che li esporta in un formato concordato.

Per quanto riguarda gli avvisi di Trenord, un semplice script che li scarica dal sito web dovrebbe essere sufficiente.

Al fine di avere dati significati nella fase successiva, è importante iniziare il prima possibile l'attività di raccolta dati.

3.3 Analisi dei dati raccolti

Innanzitutto, è necessario definire opportunamente i concetti di *tratta* e *corsa*: considerando che i numeri identificativi dei treni mutano da un giorno all'altro e non sono univoci, non è un compito banale. Quindi, si possono ricalcolare gli indici di puntualità e affidabilità *effettivi* per ogni tratta individuata e trovare correlazioni tra le performance del servizio e giorno della settimana, orario, condizioni meteo, ecc...

Si potrebbe anche verificare la regolarità e correttezza degli avvisi ai passeggeri nelle tratte affidate a Trenord e analizzare le cause dichiarate più comuni.

4 Sviluppi futuri

Nonostante ritenga l'attività di analisi dei dati estremamente interessante, l'obiettivo principale della proposta in oggetto è fornire strumenti liberi e Open Data di qualità per permettere a chiunque dotato delle capacità necessarie di continuare il lavoro. Con il supporto del Dipartimento di Informatica si potrebbe rendere l'attività di raccolta dati permanente e costante fornendo *dump* regolari accessibili da un portale web, anche di semplice costruzione.