

TECHNICAL PAPERS

**ELEMENTI
METODOLOGICI
PER L'ANALISI
DELLA MOBILITÀ
DELLE PERSONE
ATTRAVERSO
L'USO DI BIG DATA**



MINISTERO
DELLE INFRASTRUTTURE
E DEI TRASPORTI



SISTAN
SISTEMA STATISTICO
NAZIONALE



Istat
Istituto Nazionale
di Statistica



FERROVIE
DELLO STATO
ITALIANE

I Technical Papers sono pubblicazioni monografiche a contenuto scientifico, che approfondiscono i temi legati al settore dei trasporti e della mobilità, dal punto di vista tecnico, ambientale, economico, di pianificazione e modellazione, tenendo in considerazione l'interazione con le tecnologie di avanguardia.

Gli autori fanno parte dell'FS Research Centre, il centro di alta competenza interno al Gruppo FS per lo sviluppo di studi e ricerche sulla mobilità e sulle tematiche ad essa connesse, focalizzate soprattutto sulla sostenibilità e sui cambiamenti climatici, la trasformazione verde e la decarbonizzazione dei trasporti, la rivoluzione digitale, i big data e l'Internet of Things, i sistemi di trasporto innovativi, la sharing economy, gli impatti socioeconomici ed ambientali dei sistemi di trasporto, la governance, l'etica e l'equità della mobilità. FS Research Centre si avvale di una squadra di esperti di mobilità, trasporti, economia, ambiente, territorio, data intelligence, sistemi informativi geografici e tecnologie innovative, che sviluppano e utilizzano modelli matematici e sistemi di analisi dei dati all'avanguardia.

Le pubblicazioni sono spesso redatte in collaborazione con università ed enti di ricerca, istituzioni e attori pubblici e privati, sia nazionali che internazionali.

ISBN: 978-88-940638-7-5

Ferrovie dello Stato Italiane SpA
Piazza della Croce Rossa, 1 – 00161 Roma

www.fsitaliane.it

Tutti i diritti sono riservati. Nessuna parte di questa pubblicazione può essere riprodotta, archiviata, tradotta, memorizzata o trasmessa in qualsiasi forma o mezzo – elettronico, meccanico, reprografico, digitale – se non nei termini previsti dalla legge che tutela il Diritto d'Autore.

A cura di:

Gruppo Ferrovie dello Stato Italiane

Istat

Ministero delle Infrastrutture e dei Trasporti

Fotografie

© Archivio FS Italiane

© Adobe Stock (1, 11, 16, 25, 29, 32, 48, 50, 51, 62, 63, 64, 72, 74, 77, 78, 84, 95, 122)

Marzo 2024

Le opinioni espresse sono puramente quelle degli autori e non possono in alcun caso essere considerate come una posizione ufficiale di FS Italiane, Ministero delle Infrastrutture e dei Trasporti e Istat.

TECHNICAL PAPERS

**ELEMENTI
METODOLOGICI
PER L'ANALISI
DELLA MOBILITÀ
DELLE PERSONE
ATTRAVERSO
L'USO DI BIG DATA**

08 1 | INTRODUZIONE

- 1.1 Sintesi dei contenuti
- 1.2 Bibliografia

14 2 | DEFINIZIONI

- 2.1 Tipologie di indicatori ed infrastrutture di riferimento per la raccolta dati
- 2.2 Definizione di Big Data.

22 3 | LINEE GUIDA E NORMATIVE DI RIFERIMENTO PER LA RILEVAZIONE DEI DATI SULLA MOBILITÀ

- 3.1 La normativa per la produzione statistica di indicatori sul trasporto passeggeri per modo di trasporto
- 3.2 Riferimenti normativi per la produzione di indicatori sui comportamenti di mobilità
- 3.3 Quadro sintetico delle normative per le statistiche sulla mobilità delle persone, con riferimento alla produzione statistica.

37 4 | LE FONTI DATI SULLA MOBILITÀ IN ITALIA

- 4.1 Dati da indagine, dati da fonti amministrative e da altre fonti
- 4.2 Fonti complementari a completamento del quadro informativo
- 4.3 Fonti di riferimento per la rappresentazione infrastrutturale e dell'offerta di servizi
- 4.5 Sintesi delle principali indagini statistiche

58 5 | ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

- 5.1 Analisi delle limitazioni delle attuali fonti dati della mobilità delle persone
- 5.2 Rafforzare il potere informativo delle indagini tradizionali attraverso l'uso dei Big Data
- 5.3 Bibliografia





Sommario

- 73** **6 | IL CICLO DI VITA DEI BIG DATA**
- 6.1 Architettura di riferimento
 - 6.2 Ruoli e Attori
 - 6.3 La standardizzazione dell'accesso ai dati
 - 6.4 Bibliografia
- 85** **7 | L'ESTRAZIONE DELL'INFORMAZIONE**
- 7.1 Analisi sperimentale
 - 7.2 Conclusioni
- 103** **8 | ANALISI E MODELLIZZAZIONE**
- 8.1 Formati e strumenti di archiviazione, gestione ed analisi dei Big Data
 - 8.2 Analisi dei Big Data e utilizzo per la previsione di fenomeni di interesse
 - 8.3 Utilizzo e prospettive future dei Big Data e Machine Learning nell'ambito dei trasporti
 - 8.4 Valutazioni conclusive
 - 8.5 Bibliografia
- 115** **9 | CONCLUSIONI E RACCOMANDAZIONI**
- 9.1 Prospettive future
- 120** **10 | ACRONIMI E DEFINIZIONI**

1 | INTRODUZIONE

Mario Tartaglia ¹ [0000-0003-3216-8150]

1. FS Research Centre, Firenze, Italia

08

Il progetto di cui fa parte il presente elaborato nasce in risposta alla necessità conoscitiva da parte dei decisori istituzionali e del settore industriale riguardo gli aspetti della mobilità non ancora coperti in maniera esaustiva dalle fonti statistiche disponibili, ufficiali o meno, e la conseguente richiesta di maggiore tempestività e continuità delle informazioni ad essi relative. Il periodo storico attuale, in cui si inserisce questa analisi, viene definito da alcuni come la "Quarta Rivoluzione industriale": una quantità sempre crescente delle attività quotidiane degli individui (pianificazione degli spostamenti, acquisti, operazioni economiche, etc.) si trasferiscono nel "mare" del mondo digitale, che per contro restituisce delle "onde di ritorno" che sono costituite da grandi moli di dati. Questi dati sono caratterizzati da volumi talmente importanti e da caratteristiche così peculiari che è stato coniato il nuovo termine "Big Data" per indicarli. A questi dati sono inoltre attribuite delle proprietà che li caratterizzano, le principali delle quali sono qui di seguito riportate e verranno poi esplicitate nel capitolo 2 "Definizioni":

- Volume – riferendosi alla quantità dei dati.
- Velocità - velocità di produzione.
- Varietà – differente tipologia dei formati dei dati archiviati (foto, video, json, geo-json,..).
- Veridicità - grossa incertezza perché noi non sappiamo bene qual è il campione da cui questi dati vengono estratti.
- Valore- intendendo il potenziale ma importantissimo valore informativo.
- Variabilità - la variabilità anche della forma dei dati che crea grandissimi problemi informatici.

Dunque, alle indagini tradizionali e strutturate che danno origine ai dati "primari" si affianca-

no sempre più nell'analisi dei fenomeni i cosiddetti dati "secondari" provenienti spesso dai Big Data. È però importante sottolineare che i Big Data esistono a prescindere dai quesiti scientifici cui si cerca di rispondere attraverso il loro utilizzo e spesso sono il risultato secondario di altre attività. Ad esempio, nell'ambito della mobilità, le informazioni che si possono collezionare dai dispositivi mobili derivano dall'interazione con le antenne ed i sistemi GPS per la finalità ultima di effettuare chiamate o scambiare messaggi e/o dati, fine ben diverso da quello di tracciare gli spostamenti per indagare sulle abitudini di mobilità delle persone. Con una metafora relativa all'evoluzione umana, dove si è passati dal nomadismo per necessità di caccia e ricerca del cibo alla sedentarietà con l'introduzione dell'allevamento e dell'agricoltura, nell'ambito dei dati si sta verificando il processo inverso. Infatti, mentre il "coltivatore" di informazioni semina il dato che poi vuole raccogliere e quindi progetta e struttura l'indagine in funzione del tipo di informazione voluta, con l'utilizzo dei Big Data si vaga alla ricerca delle informazioni laddove si pensa possano essere presenti e dove si pensa sia possibile la loro estrazione. In qualche modo, l'impiego dei Big Data a fini statistici è simile a un processo di riciclo/riuso. L'altro elemento di differenza sostanziale fra i Big Data e le indagini tradizionali è che queste ultime sono focalizzate sulle relazioni di causa-effetto esistenti nel fenomeno che devono indagare mentre i primi analizzano gli aspetti fenomenologici (ovvero quelli cinematici per quanto riguarda la mobilità) dei fenomeni, ossia il modo in cui accadono e non le motivazioni. In sintesi, come mostrato in (fig. 1), i Big Data hanno un grande potere in termini di fornitura di volumi di informazione, ma in realtà hanno un minor potere in termini di profondità della

conoscenza. I dati che provengono dalle indagini tradizionali, detti anche “small data” oppure “thick data” se di tipo prevalentemente qualitativo, hanno la possibilità di dare rispo-

ste più pregnanti proprio perché sono rilevati su un progetto che è definito da una specifica necessità di conoscenza.

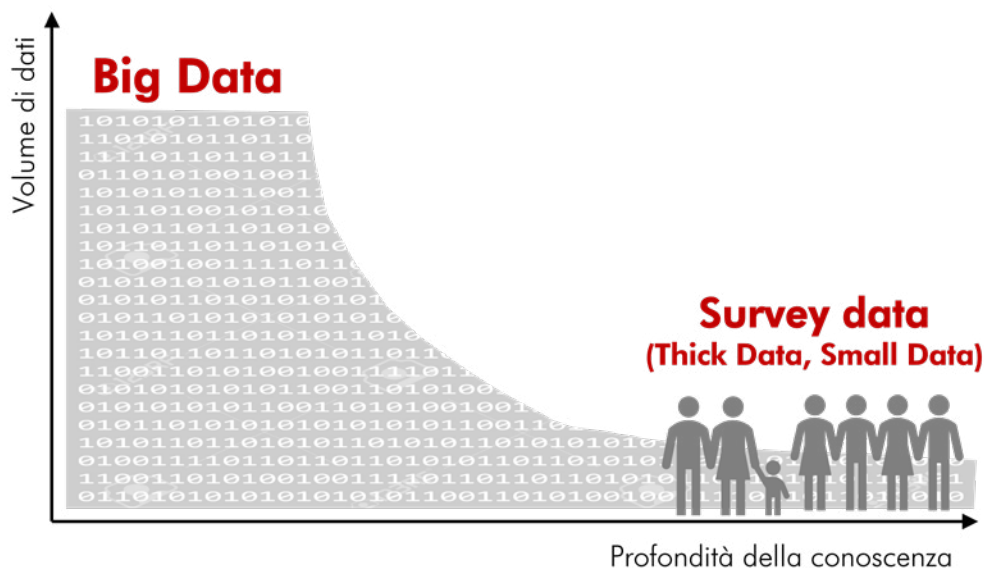


Figura 1 - La relazione fra Big Data ed indagini tradizionali (liberamente adattato da Tricia Wang, 2016)

Ai fini dell'analisi della mobilità, risulta quindi necessaria una riflessione sulle modalità di utilizzo dei Big Data e di integrazione con le campagne di raccolta dati esistenti ed in questo contesto si inserisce il presente studio. Il progetto ha l'obiettivo di definire e validare una strategia che utilizzi i dati IOT e i Big Data per colmare la lacuna informativa in merito agli spostamenti delle persone effettuati in diversi ambiti territoriali, provvedendo a descrivere per ciascuno di essi:

- il limite di applicabilità del dato;
- il livello di qualità e affidabilità della stima;
- le definizioni da adottare per identificare le caratteristiche di ciascun dominio;
- gli indicatori di rappresentatività di una matrice O/D degli spostamenti per dominio.

Lo studio si focalizza con particolare atten-

zione sulla mobilità delle persone, individuando ed ampliando con un taglio territoriale disaggregato, le principali istanze riportate nei documenti internazionali di riferimento. L'obiettivo generale, in un'ideale prosecuzione del lavoro qui presentato, è la definizione e validazione di un documento metodologico anche tramite lo sviluppo di elementi prototipali finalizzati alla costruzione di indicatori sulla mobilità delle persone, ottenuti da fonti IOT/Big Data, con particolare riguardo alla fonte Mobile Network Data (MND).

La prima sezione del presente elaborato è dedicata alla sistematizzazione delle definizioni di mobilità come elemento cardine dei fenomeni da indagare per soddisfare le esigenze informative sul tema (popolazione di riferimento, mobilità ricorrente, per modo di trasporto, per motivazione, domini territoriali di interesse, classificazione per classi di distanza,

etc), con riferimento sia alle Eurostat Guidelines on Passenger Mobility Statistics¹, sia al complesso della produzione statistica finalizzata a fornire una rappresentazione della mobilità ai principali utilizzatori. Dopo l'analisi normativa, si entra quindi nella descrizione delle indagini con un focus sugli indicatori di output riguardanti la mobilità; si analizzano le differenze informative delle varie indagini e se ne evidenziano i limiti a cui si cerca soluzione nell'utilizzo dei Big Data. Lo step successivo riguarda la definizione del processo produttivo

e del controllo di qualità dei Big Data a partire dalle best practice e dalle linee guida in materia. Si illustrano pertanto le potenzialità e i limiti dei Big Data, in particolare dei dati di telefonia mobile, con la narrazione di esempi applicativi in cui sono stati coinvolti i membri del gruppo di lavoro. In questa fase, vengono esplicitati i limiti di applicabilità delle singole tipologie di dati. È presente, infine, una breve rassegna delle tecniche di trattamento dei Big Data in relazione anche al tema di crescente interesse del Machine Learning.

1.1 | SINTESI DEI CONTENUTI

10

Si riportano di seguito le sintesi dei principali argomenti trattati nei capitoli che costituiscono l'elaborato.

- **Capitolo 2: Definizioni**

Le variabili e gli indicatori utilizzati per rappresentare la mobilità sono di vario tipo, a seconda dell'ottica con cui si osserva il fenomeno. Nell'ambito delle fonti tradizionali, si possono distinguere gli indicatori di attività per modo di trasporto (di consistenza, di flusso, di traffico e di performance), quelli relativi alla domanda/offerta dei servizi (dotazione, livelli di domanda/offerta, qualità e soddisfazione), e infine i dati relativi ai comportamenti e alle scelte individuali (durata e lunghezza degli spostamenti, origine/destinazione, motivazioni, split modale). A questi si possono affiancare le 'nuove fonti', i cosiddetti Big Data, che nascono per scopi differenti dalla rappresentazione statistica dei fenomeni, e le cui caratteristiche specifiche rendono necessaria l'introduzione di tecniche, tecnologie e professionalità ad hoc per il loro trattamento e utilizzo a fini statistici. Inoltre, per entrambe le tipologie di fonti, è necessario tenere conto delle infrastrutture di

riferimento che li generano e in base alle quali i dati stessi sono codificati.

Si riportano quindi nel capitolo 2 i principali indicatori utilizzati per rappresentare il fenomeno della mobilità e le infrastrutture di riferimento; si introducono alcuni concetti relativi al mondo dei Big Data e si descrivono sinteticamente le tipologie di Big Data utilizzabili nell'ambito delle analisi di mobilità.

- **Capitolo 3: Linee guida e normative di riferimento per la rilevazione dei dati sulla mobilità**

In Italia, il panorama delle norme di riferimento per la produzione statistica di dati sulla mobilità dei passeggeri è caratterizzato da un'impostazione per "modo di trasporto", con un gap importante dato dall'assenza di norme per le statistiche del trasporto di passeggeri su strada; inoltre non esiste un riferimento normativo a livello nazionale per la rappresentazione statistica dei comportamenti di mobilità. La legislazione che determina la produzione di statistiche per modo di trasporto è principalmente definita in ambito UE, in forma di Regolamenti o Direttive, poi recepite

1. https://ec.europa.eu/eurostat/documents/29567/3217334/Guidelines_on_Passenger_Mobility_Statistics+%282018_edition%29.pdf/f15955e3-d7b4-353b-7530-34c6c94d2ec1?i=1611654879518

dal sistema legislativo nazionale. Sia in ambito Europeo che nazionale, sono state messe a punto alcune Linee Guida connesse direttamente o indirettamente con la produzione di indicatori sulla mobilità.

Si illustrano nel capitolo 3 le principali linee guida, regolamenti e normative distinguendo per ambito di applicazione, europeo o nazionale. Si riportano inoltre i riferimenti normativi per attività che prevedono la raccolta di dati di mobilità definiti da linea guida o legge.

- **Capitolo 4: Le fonti dati sulla mobilità in Italia**

La produzione statistica sulla mobilità delle persone è articolata in numerose fonti che possono essere classificate sia rispetto alle loro caratteristiche metodologiche e al loro potere informativo, e quindi all'essere o meno 'statistiche ufficiali', sia rispetto alle tipologie di indicatori che le costituiscono, e quindi al taglio prospettico con cui rappresentano il fenomeno. Abbiamo pertanto un insieme variegato di processi, definizioni e visioni,

che contribuiscono a una descrizione estesa ma non organica né esaustiva della mobilità. Spesso, inoltre, le fonti disponibili rispondono ad esigenze informative e normative (si veda il Capitolo 3 "Linee guida e normative di riferimento per la rilevazione dei dati sulla mobilità") di differente natura e non sovrapponibili. Peraltro, le rilevazioni richiedono forti investimenti in termini di risorse da poter utilizzare in un'ottica sostenibile nel tempo, per poter produrre output complessi e significativi. Per questi motivi, avere una rappresentazione organica ed esaustiva del fenomeno considerato è un importante obiettivo a tendere, in cui l'uso dei Big Data costituirebbe un indubbio valore aggiunto.

Si descrivono nel capitolo 4 le fonti dati sulla mobilità raggruppandole per tipologia di rilevazione (comportamenti di mobilità, offerta e domanda, modo di trasporto). Vengono inoltre individuate alcune fonti complementari utili per la definizione dei modelli di trasporto e/o per la valutazione di strategie nazionali ed internazionali.



- **Capitolo 5: Analisi di alcuni casi studio relativi alla mobilità delle persone**

L'insieme delle fonti dati disponibili relative alla mobilità in Italia compone un quadro articolato e frastagliato; da una parte si hanno grandi quantità di dati legate però a settori specifici, dall'altra indagini complete sulle abitudini di mobilità ma su campioni di dimensioni ridotte. Nel primo gruppo rientrano i dati provenienti dai gestori dei servizi e delle infrastrutture di trasporto, come i dati di bigliettazione ferroviaria, i dati autostradali e aeroportuali; molte informazioni, difficilmente sovrapponibili con quelle provenienti dagli altri settori, che forniscono una visione della mobilità rispetto all'organizzazione dell'infrastruttura di trasporto specifica. Prendendo per esempio in considerazione i primi elementi nella definizione della domanda di trasporto, la generazione e la distribuzione degli spostamenti sul territorio, le origini e le destinazioni sono infatti riferite a punti notevoli dell'infrastruttura, quali stazioni, barriere autostradali ed aeroporti e forniscono quindi una visione parziale delle origini/destinazioni degli spostamenti. Dall'altra parte, indagini come il censimento ISTAT, l'indagine Audimob di Isfort, forniscono informazioni molto precise su origine e destinazione degli spostamenti ma su un campione parziale di popolazione o perché limitato rispetto al motivo dello spostamento, lavoro o studio per il censimento, o per il numero stesso delle persone intervistate.

Nel capitolo 5 si raccolgono alcune delle limitazioni dei dati a disposizione e si individuano, nell'ottica della sostenibilità tecnico-economica, alcune possibili soluzioni con uno sguardo alle nuove prospettive introdotte dall'utilizzo dei Big Data.

- **Capitolo 6: Il ciclo di vita dei Big Data**

Il ciclo di vita dei Big Data, per le caratteristiche dei dati e dei processi che li generano, richiede la definizione di una pipeline dedicata che preveda fasi di processo completamente nuove rispetto ai classici processi di

produzione statistica di stime e indicatori come quelli realizzati su dati da indagini o da fonti amministrative. Infatti, ad esempio, le indagini vengono disegnate appositamente per rispondere a specifiche esigenze conoscitive, mentre le fonti amministrative vengono scelte in funzione di quanto sono in grado di descrivere il fenomeno di interesse.

L'approccio all'uso dei Big Data richiede l'introduzione di fasi di analisi preliminari non solo per la loro selezione, ma soprattutto per definire nuove metodologie di elaborazione. L'uso di questi dati richiede il coinvolgimento personale con specifiche competenze nell'analizzare grandi volumi di dati, caratterizzati da elevata variabilità e da un valore informativo intrinseco che richiede conoscenze tecniche di vario tipo per essere gestito. Queste caratteristiche, unite ad aspetti come accesso e privacy dei dati, rendono necessaria l'implementazione di una specifica organizzazione non solo dell'intero processo del ciclo di vita del dato ma anche delle risorse e delle relazioni con i fornitori che spesso sono esterni all'utilizzatore dei dati stessi. In particolare, l'analisi di questo tipo di dati richiede l'introduzione di figure professionali innovative come i data scientist, nonché di architetture Software e Hardware dedicate. In generale si deve essere consapevoli che l'uso dei Big Data richiede grandi investimenti non solo in termini tecnici ma soprattutto di acquisizione di conoscenze e di risorse umane qualificate.

Nel capitolo 6, a partire dall'esperienza europea del progetto ESSnet Big Data si riportano le proposte di standardizzazione del processo di trattamento dei Big Data; si forniscono infine gli approcci proposti per la protezione della privacy.

- **Capitolo 7: L'estrazione dell'informazione**

Si descrive in questo capitolo l'esperienza maturata nell'ambito dell'interazione con un operatore telefonico e rispettiva società di analisi, al fine di acquisire e utilizzare i

Big Data di tipo telefonico per analisi della mobilità sul territorio nazionale italiano. Le elaborazioni riportate sono state condotte dal FS Research Centre del Gruppo Ferrovie dello Stato Italiane e dalla Direzione Strategie di Trenitalia sempre del Gruppo FS, con il supporto di Vodafone Business e Motion Analytica, attraverso l'utilizzo dei Vodafone Analytics (Big Data telefonici non personali, anonimizzati e aggregati in conformità assoluta ai requisiti di privacy del GDPR) generati dalla rete Vodafone. Nella prima parte del documento si descrivono le caratteristiche tecniche dei dati utilizzati per poi passare alla descrizione di alcune elaborazioni effettuate con i Big Data telefonici, evidenziando quali sono state le criticità incontrate nel trattare questi nuovi tipi di dati, quali quelle risolte e quali invece le questioni ancora aperte.

- **Capitolo 8: Analisi e modellizzazione**

Grandi quantità di dati, con numerosità completamente differente rispetto ai campioni delle indagini tradizionali, richiedono strumenti e tecniche nuovi per la rappresentazione e l'estrazione delle informazioni e per il loro utilizzo per analisi e previsioni nell'ambito dei trasporti.

Al crescente interesse nei confronti dei Big Data, negli ultimi anni si affianca il grande fermento rispetto a studi e strumenti relativi all'intelligenza artificiale. Il primo utilizzo del Machine Learning nel campo dei Big Data è relativo all'interpretazione ed analisi delle

grandi quantità di informazioni contenute negli stessi. L'altro filone di sviluppo della relazione fra Big Data e Machine Learning è l'applicazione nell'ambito della modellistica dei trasporti per scopi previsionali; gli algoritmi del Machine Learning una volta "addestrati" permettono di prevedere le variabili di interesse al variare degli input iniziali. L'implementazione e l'addestramento degli algoritmi richiedono numerosi dati, esigenza che sembra trovare risposta proprio nella caratteristica "volume" dei Big Data. Nonostante l'accordo fra le necessità per lo sviluppo di "modelli" basati sul Machine Learning ed il potenziale dei Big Data, l'applicazione nel campo della modellistica dei trasporti pone alcuni dubbi rispetto alla sostituzione degli approcci tradizionali, legati alla capacità delle nuove tecniche di riprodurre la discrezionalità della scelta dell'utente ma soprattutto legati all'opacità del funzionamento del modello; si passa infatti dal tradizionale approccio teorico-sperimentale in cui si conoscono le leggi che governano i modelli ad un sistema in cui il modello è una scatola nera che risponde agli input secondo logiche non note all'utilizzatore.

Si illustrano nel capitolo 8 le principali tecniche di intelligenza artificiale affermatesi negli anni 2000, le possibili applicazioni in combinazione con i Big Data nell'ambito dei trasporti, i benefici e i principali limiti delle nuove tecniche. Per la velocità di evoluzione dell'argomento trattato, gli esempi portati non possono considerarsi esaustivi della materia.

1.2 | BIBLIOGRAFIA

[1.1] Schwab K., 2016, "La quarta rivoluzione industriale" Franco Angeli

2 | DEFINIZIONI

Giovanna Astori ¹

1. ISTAT, Roma, Italia

2.1 | TIPOLOGIE DI INDICATORI ED INFRASTRUTTURE DI RIFERIMENTO PER LA RACCOLTA DATI

2.1.1 | TIPOLOGIE DI INDICATORI PER LA RAPPRESENTAZIONE DELLA MOBILITÀ DELLE PERSONE

Il fenomeno della mobilità delle persone è complesso e articolato, in quanto realizza il tessuto connettivo dell'organizzazione socio-economica di un territorio e allo stesso tempo dipende da esso. Per questo motivo, anche la sua rappresentazione quali-quantitativa nel campo della ricerca e della produzione statistica è tutt'altro che lineare, ma presenta numerose prospettive e sfaccettature che di seguito si proverà a schematizzare, mettendone in luce le caratteristiche anche con riferimento alle definizioni e alle variabili ad esse pertinenti. A grandi linee, si possono identificare, nell'ambito della produzione statistica nazionale ed internazionale, quattro prospettive principali, all'interno delle quali la mobilità ed alcuni altri fenomeni ad essa connessi devono essere rappresentati a supporto delle attività del settore pubblico e privato e a beneficio della collettività.

- a. Rappresentazione dell'attività e della performance con riferimento alle imprese e alle infrastrutture sul territorio, per modo di trasporto. In questa categoria si possono individuare alcune fattispecie:
- dati di consistenza/puntuali e di flusso: numero di passeggeri (ad esempio: il numero di passeggeri imbarcati/sbarcati per porto o aeroporto; il numero di passeggeri trasportati per rotta aerea; il numero di passeggeri trasportati per modo e tipologia di flusso: nazionale/internazionale UE/Extra UE/Totale);
 - dati di traffico: Veicoli-chilometro (Vkm), che rappresentano il totale dei km per-

corsi dalle diverse categorie di veicoli in un arco temporale di riferimento, e in un ambito territoriale definito (ad esempio: Totale dei km percorsi dalle autovetture in Italia nell'arco di un anno; Totale dei km percorsi dai veicoli leggeri sulle strade extraurbane nell'arco di un mese, totale dei km percorsi per tipo, età, alimentazione dei veicoli). Numero di veicoli transitati su un territorio in un arco temporale (es. numero di veicoli leggeri transitati sull'autostrada A1 in un mese, numero medio di veicoli/giorno sulla strada X nella fascia oraria y_1 - y_2 , etc);

- dati di performance: Passeggeri-chilometro (Pkm), che rappresentano il totale dei km percorsi da tutti i passeggeri che si sono spostati, per modo di trasporto e in relazione ad un riferimento territoriale e ad un arco temporale definito. Coincidono con i Vkm nel caso di corrispondenza "uno a uno" fra veicolo e passeggero. Nel caso generale, i Pkm sono pari alla somma, per tutti gli occupanti il veicolo, dei km (individuali) percorsi. (ad esempio: Passeggeri-chilometro realizzati su ferrovia in una regione, nel trimestre; Passeggeri-chilometro effettuati in Italia, con autovetture private, in ambito urbano, nell'anno di riferimento).
- I dati di questa categoria sono rilevati generalmente in forma censuaria (con eventuale applicazione di soglie o limitazioni del campo di osservazione, per esempio con riferimento alla 'dimensione'

delle imprese coinvolte, o ad altri parametri) o attraverso dei punti di rilevazione del fenomeno dislocati sul territorio, oppure derivano da elaborazioni di dati da registri amministrativi pubblici o privati. Nella maggior parte dei casi, non sono dati derivanti da indagini che prevedano campioni casuali delle unità statistiche di riferimento.

- b. Rappresentazione del livello/stato dell'infrastruttura, dell'offerta del servizio, del loro uso e soddisfazione. È una prospettiva che definisce il fenomeno dal punto di vista delle potenzialità sia dell'infrastruttura che della domanda/offerta e della programmazione dei servizi, nonché da quello della valutazione dei cittadini rispetto alla qualità degli stessi. Sono dati quantitativi e qualitativi, anche di tipo amministrativo, e connessi alla definizione dei PUT/PUM/PUMS a livello di amministrazioni locali. In questo gruppo abbiamo:
- Dotazione infrastrutturale: numero di infrastrutture presenti su un territorio (es. porti, aeroporti, stazioni, impianti fissi, funicolari, piste ciclabili, punti ricarica veicoli elettrici, ZTL, parcheggi pubblici e di scambio, rastrelliere installate, inventario delle strade, etc). Caratteristiche quali-quantitative delle infrastrutture presenti (es. lunghezza della rete ferroviaria, dimensioni delle banchine portuali, dotazione di collegamenti intermodali, dotazioni dei terminal portuali, dotazioni e caratteristiche delle aree di servizio autostradali, etc);
 - Offerta: livello dell'offerta di mobilità per tipologia/mo di trasporto (es. numero di veicoli disponibili in sharing, numero di treni, posti offerti per il TPL, numero di corse previste per il TPL, numero di licenze taxi, etc);
 - Domanda potenziale ed effettiva: (potenziale) dati di bigliettazione e abbonamenti (es. numero di abbonamenti TPL per tipologia, biglietti ferroviari venduti, etc): esprimono l'impatto potenziale della domanda di mobilità sul territorio, ma non danno una misura del reale utilizzo del sistema dei trasporti. (effettiva) dati di utilizzo del sistema trasporti (es. numero e riferimento territoriale dei biglietti/abbonamenti vidimati, fatturazione dei pedaggi autostradali, percorrenze effettuate con veicoli in sharing, etc);
- Qualità e livello di soddisfazione: indici e indicatori di gradimento per i vari aspetti dei servizi di mobilità offerti (es. per l'uso del TPL, dei servizi di sharing, dell'alta velocità, etc).
- c. Rappresentazione dei comportamenti di mobilità, dei viaggi e degli spostamenti effettuati dalla popolazione. È una prospettiva 'soggettiva', che descrive e quantifica i comportamenti di mobilità della popolazione di riferimento, le scelte modali, le motivazioni degli spostamenti, la struttura degli spostamenti stessi, anche con riferimento ai domini temporali – come i giorni lavorativi o festivi – e agli ambiti territoriali – urbano/non urbano, nazionale/internazionale. Uno dei tratti caratteristici di questa rappresentazione sta nella possibilità di riferire i comportamenti di mobilità alle caratteristiche sociodemografiche dei gruppi di individui, che definiscono i domini di riferimento dei set di indicatori. Le principali variabili/indicatori di riferimento possono essere individuate in:
- Temporali: la durata degli spostamenti (es. durata media degli spostamenti in ambito urbano e giorno feriale, per persona/giorno);
 - Spaziali: la lunghezza degli spostamenti (es. distanza media percorsa per persona/giorno, lunghezza media dello spostamento per motivazione, etc);
 - Di frequenza/consistenza: numerosità (es. numero medio di spostamenti per persona/giorno, numero medio di viaggi per persona/anno, numero di viaggi per

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

classe di distanza con e senza pernottamento, distribuzione degli spostamenti per fascia oraria, tasso di riempimento dei veicoli). Passeggeri-km: a differenza dell'indicatore presentato nella fattispecie a), qui lo stesso è rappresentato secondo domini di diverso tipo (genere, motivazione, età, etc), come indicato in premessa;

- Di flusso: matrici Origine/Destinazione (numero di persone che si spostano abitualmente dalle origini O1, O2, ..., On alle destinazioni D1, D2, ..., Dm; matrici di pendolarismo, etc.).
- d. Rappresentazione dei fenomeni correlati alla mobilità. In questa fattispecie rientrano gli indicatori e le variabili che contribuiscono ad ampliare e caratterizzare ulteriormente il quadro rappresentativo degli spostamenti delle persone o ne sono influenzati. Tra i principali, sono stati identificati alcuni temi della ricerca e della produzione statistica:
 - Turismo;
 - Incidentalità stradale;
 - Accessibilità: studi sperimentali e modelli per misurare l'accessibilità dei luoghi di interesse, di particolare importanza in ambito urbano, correlati anche all'Area 11 degli SDG. (es. quota di popolazione residente entro una soglia di distanza da luoghi di pubblico interesse – es. scuole, fermate TPL, stazioni. La distanza è generalmente definita in termini di durata dello spostamento a piedi, oppure con uno specifico modo di trasporto; in questo caso si parla di accessibilità per quel modo, es. bicicletta);
 - Quantificazione della popolazione incidente, ossia della popolazione che, per qualsiasi motivo, 'incide' ossia si sposta da, verso ed entro un determinato territorio;
 - Quantificazione delle emissioni atmosferiche dei trasporti;
 - Indici di fatturato, risultati economici, investimenti delle imprese del settore del trasporto di persone.



2.1.2 | INFRASTRUTTURE E RIFERIMENTI PER L'ACQUISIZIONE DEI DATI E LA COSTRUZIONE DI INDICATORI DELLA MOBILITÀ

Per avere un quadro teorico ma anche operativo per la sistematizzazione delle metodologie per la produzione degli indicatori di mobilità, è utile mettere in luce le caratteristiche del sistema infrastrutturale e definitorio entro cui vengono rilevati i dati. Gli indicatori prodotti o di interesse, che siano di fonte amministrativa o da indagine statistica, fanno riferimento a entità di tipo territoriale, infrastrutturale o economico, sia dal punto di vista dell'osservazione che della successiva rappresentazione sintetica. Se ne individuano alcune di seguito:

- Porti; aeroporti; stazioni; autostazioni;
- Unità amministrative/territoriali (Paesi; Regioni; Ripartizioni; Province/Città metropolitane; Comuni; etc);
- Tronchi stradali/autostradali; Autostrade; infrastrutture ferroviarie per tipologia;

- Imprese/gestori del servizio;
- Popolazione residente.

È inoltre utile tenere presente che i domini, entro i quali devono essere calcolati gli indicatori, tengono conto di ulteriori classificazioni ed entità di riferimento, tra cui:

- Territoriali: Functional Urban Areas, classi di distanza, ambito urbano/non urbano, etc;
- Modali: modo di trasporto, modo di trasporto prevalente;
- Temporal: giorno feriale/festivo, fasce orarie, mobilità ricorrente, mobilità abituale, etc;
- Socioeconomici: genere, età, residenza, etc;
- Altri attributi: età del veicolo, alimentazione del veicolo, etc.

2.2 | DEFINIZIONE DI BIG DATA

Il concetto di "Big Data" è strettamente legato all'evoluzione tecnologica, all'uso e alla diffusione massiccia dei dispositivi elettronici che ha creato la cosiddetta "Internet of Things" (IoT), alla smaterializzazione delle transazioni economiche e finanziarie e di molte attività quotidiane (acquisti online, socializzazione, *smart working*, pianificazione dei viaggi, etc). Sono il frutto delle 'tracce digitali' (*digital footprint*) che lasciamo continuamente nel sistema in cui viviamo e operiamo. In letteratura si trova più di un paradigma definitorio con cui si è cercato di dare un'identità al concetto di Big Data (BD), i quali sostanzialmente costituiscono dei flussi di dati generati per finalità diverse, massicci e persistenti nel tempo. La definizione più completa è quella del "Modello delle 6V". Secondo questa definizione, i BD sono informazioni, strutturate o non strutturate, generate da dispositivi, social media, piattaforme digitali, sistemi informatici,

e caratterizzate da:

- **Volume:** la mole di dati generati è tale da non poter essere trattata, manipolata, elaborata con i comuni sistemi IT;
- **Varietà:** si tratta di dati di varie tipologie (strutturati, ad esempio secondo un tracciato; semi strutturati, con parti codificate e parti non strutturate; non strutturati, ad esempio segnali generati da sistemi, server etc);
- **Velocità:** sono generati con velocità elevata, o assimilabile ad un flusso continuo;
- **Veridicità:** non sempre garantiscono affidabilità in termini di interpretazione e traduzione in informazione utile a rappresentare altri fenomeni;
- **Valore:** se ne può ottenere un potere informativo e quindi un valore economico;
- **Variabilità:** non è garantita la coerenza e consistenza dei flussi nel tempo, in termini di rappresentatività.

Ai fini della ricerca statistica, e in particolare nell'area della mobilità, è utile e interessante trovare dei punti di contatto fra questi flussi massicci di dati e la rete fisica dei movimenti effettuati dalla popolazione, per valutare se sia possibile utilizzarli per rappresentar-

la attraverso indicatori e modelli. Inoltre, è indispensabile sviluppare strumenti, procedure e metodologie finalizzati a manipolare e trasformare la mole di informazione grezza, rendendola idonea all'utilizzo a fini statistici e predittivi.

2.2.1 | TIPOLOGIE DI BIG DATA POTENZIALMENTE IDONEE ALLA RAPPRESENTAZIONE DELLA MOBILITÀ DELLE PERSONE

Dalla definizione di Big Data riportata alla sezione precedente, si intuisce che vi sono numerose tipologie di fonti che li generano, ed ognuna di esse dà luogo a tipologie di dati completamente diversi fra loro, che si possono prestare o meno a rappresentare la mobilità delle persone, con inconvenienti e vantaggi diversificati. Metteremo in evidenza le principali tipologie di BD che possono essere ulteriormente analizzati con questa finalità, cercando di focalizzarne la connessione con alcuni degli indicatori di interesse, sia in termini di affinità che di divergenza. Nella seconda parte del presente documento, verranno trattati casi di studio che coinvolgono in particolare i dati di telefonia mobile.

a. **Mobile Network Data (MND) o Mobile Phone Data (MPD)**, sono i dati di telefonia mobile. Sono di due tipi: il primo consiste nelle sequenze scambiate con frequenza molto ravvicinata fra le SIM telefoniche inserite nei dispositivi e le antenne per la diffusione del segnale e sono dati generati indipendentemente dal verificarsi di eventi attivi come chiamate, messaggi o connessioni a Internet. Il secondo tipo è invece l'insieme dei dati, sempre riferiti alle SIM, generati invece dal verificarsi di tali eventi. I dati da segnale hanno il vantaggio di essere molto frequenti, e permettono quindi di seguire i movimenti degli individui in maniera più continuativa, ma risentono di una minore precisione per quanto riguarda la localizzazione (dovuta alle specifiche tecniche e dal posizionamento delle an-

tenne e delle celle che fanno riferimento ad esse); i dati generati da eventi invece, pur essendo meno frequenti, consentono una localizzazione più precisa. Nei casi di studio vengono generalmente utilizzati i dati da segnale. In generale, data la diffusione capillare della telefonia mobile, gli MND garantiscono un'ampia copertura della popolazione, ma è necessario analizzarne approfonditamente la natura e il potere informativo per evitare che il loro utilizzo porti a una rappresentazione del fenomeno non sempre rispondente alla realtà. In genere, da questo tipo di dati si estrae un set di informazioni del tipo: identificativo dell'utilizzatore (anonimizzato); riferimento orario del segnale; identificativo dell'antenna/cella dell'aggancio. È importante tenere presente che alcuni protocolli ridefiniscono l'anonimizzazione dell'utilizzatore con alta frequenza (anche ogni due ore), il che rende impossibile identificare comportamenti di mobilità in un orizzonte temporale più esteso.

b. **Global Positioning System (GPS) data**, sono generati da sistemi che registrano il posizionamento dei dispositivi attraverso reti di satelliti dislocati nell'orbita terrestre. Contengono informazioni sul posizionamento dei dispositivi mobili (Smartphone, Smartwatch, dispositivi installati sui veicoli per monitoraggio o navigazione, etc) che vengono raccolti attraverso le App installate sugli stessi. A differenza degli MND, contengono an-

che le coordinate geografiche dei luoghi visitati. A seconda delle opzioni scelte dall'utente, possono tracciare i movimenti sempre, o solo quando la relativa App è in uso, oppure essere disattivati (quest'ultima opzione generalmente non consente il corretto uso dell'App o lo inibisce del tutto, pertanto è una scelta residuale). La raccolta ed elaborazione di dati da più App o dispositivi per uno stesso utente consente inoltre di avere una rappresentazione degli spostamenti effettuati più puntuale, colmando eventuali informazioni mancanti da una o più fonti. Alcune di queste fonti, che registrano la cronologia e la struttura degli spostamenti effettuati dalla persona che ha con sé il dispositivo (Google Live Traffic, Apple maps, Waze, etc) sono particolarmente interessanti per rappresentare la mobilità da un punto di vista oggettivo (ad es. la congestione del traffico) e soggettivo (ad es. i percorsi più utilizzati dagli utenti in un ambito territoriale definito; inoltre gli algoritmi riescono a identificare in maniera abbastanza precisa il mezzo utilizzato e i luoghi visitati). Tuttavia, risentono di alcune criticità, tra cui la difficoltà di identificare il reale utilizzatore del dispositivo (più account sullo stesso dispositivo, o lo stesso account su più dispositivi, generano sovrapposizioni ed errori), la possibilità da parte degli utenti di disattivarle o modificare i percorsi effettuati, e non ultime le difficoltà di accesso essendo dati di proprietà privata.

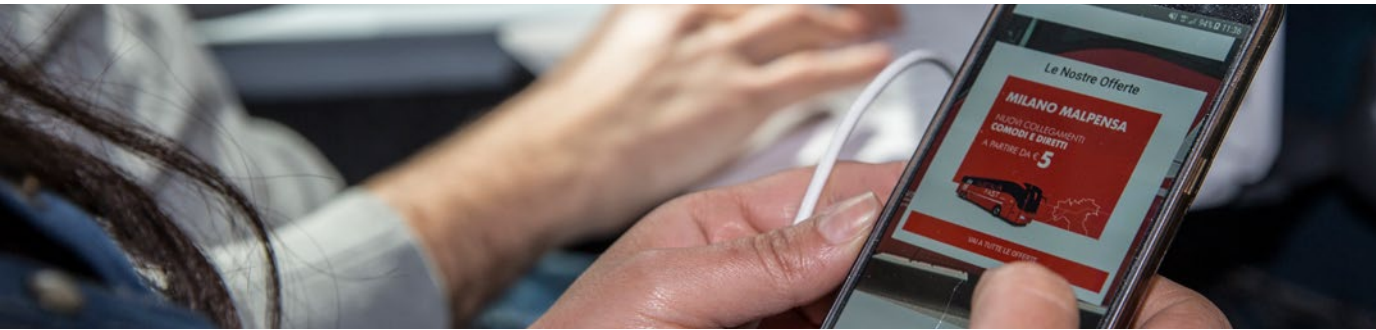
- c. **Social network data**, sono le informazioni che gli utenti inseriscono volontariamente sul proprio profilo dei canali social (Instagram, Twitter, Facebook, WhatsApp, Google etc), dal momento in cui descrivono o raccontano fatti attinenti alle attività svolte nel corso della giornata, registrandosi nei luoghi virtuali corrispondenti ai luoghi fisici in cui svolgono l'attività stessa. Sono dati di tipo discontinuo e soggetti a un margine di errore o

approssimazione, ma possono essere utili per verificare eventi di mobilità derivati da altre fonti. Per questo tipo di dati è necessario porre particolare attenzione alla protezione della riservatezza nell'accedere alle informazioni.

- d. **Floating car data (FCD)**, sono le informazioni raccolte dalle 'scatole nere' (OBU – On Board Unit) installate sui veicoli di ultima generazione, o per fini assicurativi, generati da sistemi satellitari per la localizzazione. Sono utilizzati soprattutto per quantificare il traffico, ma possono essere utilizzati per la rappresentazione di dati di tragitto. Le informazioni contenute sono relative alla georeferenziazione dello spostamento, ai riferimenti temporali, alla velocità di marcia, etc. Questo tipo di dati è utilizzabile in sinergia con i dati di traffico e transito acquisiti con l'implementazione e la sperimentazione di percorsi "Smart road" (V2I – comunicazione fra infrastruttura e veicolo – e V2V – comunicazione fra veicoli), che prevedono l'installazione lungo tratti stradali di sensori di varie tipologie (spire per la rilevazione del passaggio dei veicoli, telecamere per il riconoscimento automatico delle targhe, sensori comunicanti con i veicoli, etc), con l'obiettivo, attraverso la comunicazione digitale con e tra i veicoli di passaggio, di ottimizzare il traffico e i percorsi, aumentare i livelli di sicurezza e promuovere l'uso di veicoli a guida autonoma. Le informazioni contenute nei dati di tipo "smart road" prevedono la rilevazione di numerose variabili, tra cui: georeferenziazione dei punti di osservazione, tipologia del veicolo, distanza fra veicoli, velocità, riferimento temporale puntuale, riferimento alle corsie/carreggiate di rilievo e della direzione di marcia, con rilevazione continua e trasmissione all'hub dati in intervalli di pochi minuti. I dati di tragitto dei veicoli (registrazione sulla strada di un evento 'ingresso' e di

un evento 'uscita' per lo stesso veicolo, identificato e anonimizzato) sono di tipo campionario, e possono essere messi in relazione con l'universo di riferimento, costituito dalla totalità dei veicoli 'registrati' dai sensori sul tratto stradale considerato.

- e. **Sharing mobility data**, dati acquisiti dalle flotte dei veicoli in sharing (automobili, motocicli, monopattini, ebike, etc). Sono informazioni relative al posizionamento dei veicoli in origine e destinazione degli spostamenti (anche in relazione a nodi di scambio come stazioni ferroviarie e TPL), al chilometraggio effettuato, ai tempi di percorrenza. I dati possono essere messi in relazione con le caratteristiche socioeconomiche dell'utilizzatore. Possono essere impiegati per una rappresentazione dell'uso dei diversi modi di trasporto in ambito urbano e le scelte intermodali.
- f. **Smart cards data**, acquisiti attraverso l'utilizzo di carte elettroniche (che incorporano microchip con tecnologie RFID, NFC, etc) che comunicano attraverso onde radio con dispositivi fissi pertinenti all'accesso a luoghi di diverso tipo. Sono dispositivi di questo tipo gli abbonamenti elettronici per il trasporto pubblico locale, o le cards che includono, oltre all'accesso al TPL, anche l'ingresso a luoghi di interesse (musei etc) e sono generalmente utilizzate da turisti e visitatori. Attraverso tali dispositivi si registrano pertanto una serie di informazioni molto dettagliate sia con riferimento all'individuo, sia ai luoghi visitati che ai mezzi di trasporto utilizzati. Un limite nel loro uso ai fini della rappresentazione dei comportamenti di mobilità individuale sta nel fatto che in molti casi permettono di registrare l'accesso o l'utilizzo di un mezzo, ma non la traiettoria (accesso e uscita).
- g. **Points of interest data (POI)**, derivano da App e database (es. Google places), che identificano, in corrispondenza delle coordinate geografiche, dei luoghi classificati secondo i loro attributi (se per esempio si tratta di scuole, musei, uffici, ministeri, luoghi di intrattenimento, centri commerciali etc). Queste informazioni consentono pertanto di identificare, se associate ad altre fonti di dati come quelli da telefonia mobile, gli obiettivi degli spostamenti e quindi, con un certo grado di approssimazione, le motivazioni della mobilità.
- h. **Dati sull'uso del territorio**, sono in genere immagini satellitari da cui si possono trarre informazioni di varia natura che possono essere di ausilio per alimentare modelli di descrizione e predizione della mobilità delle persone, o rappresentare fenomeni ad essa connessi. Un'area in cui vengono impiegati è quella dello studio dell'accessibilità. Un altro tipo di dati che rientra in questa fattispecie, sono le rappresentazioni del territorio realizzate con sistemi collaborativi, in cui gli utenti su base volontaria contribuiscono a definire la struttura del territorio e delle infrastrutture in esso collocate, utilizzando delle applicazioni, come nel progetto Open Street Map.
- i. **Automatic Identification System data (AIS)**. AIS è un Sistema basato su tecnologia GPS e canali radio, che permette di identificare il posizionamento, e quindi le traiettorie, delle navi. I dati AIS sono utilizzati in molte sperimentazioni per la produzione statistica, in particolare del traffico e degli incidenti marittimi, e per stimare le emissioni. Sono delle sequenze di caratteri che necessitano di un pre-trattamento di filtraggio e decodifica per poter essere utilizzati a fini statistici. Una delle maggiori criticità sta nella disponibilità dei dati e quindi nella loro accessibilità, che sono detenuti da più soggetti pubblici e privati.



2.2.2 | INFRASTRUTTURE E RIFERIMENTI DEI BIG DATA PER LA MOBILITÀ

Per individuare i punti di contatto e divergenza fra i riferimenti spazio-temporali e concettuali con cui sono osservati i dati 'tradizionali' e prodotti gli indicatori statistici sulla mobilità delle persone, e quelli relativi ai Big Data che potrebbero essere utilizzati allo stesso scopo, è utile una breve ricognizione di alcuni di questi ultimi, come fatto nella sezione 2.1.2 per quelli tradizionali e gli indicatori.

- I dati acquisiti attraverso dispositivi mobili personali (SIM, smartphone, smartwatch, cards, etc) fanno riferimento ad account o intestatari, senza alcuna certezza che l'utilizzatore del dispositivo corrisponda all'intestatario. Se questa corrispondenza è più probabile in alcuni casi (Smart cards), è invece certa la non corrispondenza in casi come quello delle Sim aziendali. Questo rende problematico, ad esempio, riferire i dati rilevati alle caratteristiche sociodemografiche della persona che si sposta;
- I dati acquisiti attraverso sensori fissi sulla sede stradale, ad esempio contatori del traffico, semafori intelligenti, smart roads, hanno come riferimento delle entità spaziali che potrebbero essere riconducibili ad altre entità geografiche con difficoltà (ad esempio: il tronco stradale riferito alla provincia o alla regione, se vicino ai confini);
- I dati da telefonia mobile, acquisiti con riferimento alle antenne e alle loro celle, necessitano di complesse elaborazioni probabilistiche per essere ricondotti ai riferimenti geografici o a quelli spaziali comunemente utilizzati, soprattutto in aree ad alta densità, dove le celle spesso si sovrappongono o hanno forma irregolare.
- I dati acquisiti attraverso sensori installati sui veicoli, come illustrato nella sezione precedente, sono di diverse tipologie. I floating car data danno informazioni sul passaggio o sulle traiettorie dei veicoli, ma non dicono nulla circa gli occupanti degli stessi. Le coordinate GPS, peraltro, quando disponibili, sono direttamente riferibili ad ambiti territoriali (es. regioni, comuni) ben definiti, almeno fino a una certa soglia. I dati relativi al conteggio (contatori di salita/discesa, non i dati di tipo smart card che prevedono un'azione da parte del passeggero) dei passeggeri del TPL che accedono a un veicolo danno una misura del riempimento del veicolo stesso, e dei passeggeri-km realizzati ma non delle traiettorie origine/destinazione in quanto non è identificato il singolo passeggero.

3 | LINEE GUIDA E NORMATIVE DI RIFERIMENTO PER LA RILEVAZIONE DEI DATI SULLA MOBILITÀ

Giovanna Astori ¹, Giovanni Zacchi ², Francesca Sieli ²

1. ISTAT, Roma, Italia

2. MIT, Roma, Italia. Direzione generale per la digitalizzazione, i sistemi informativi e statistici..

3.1 | LA NORMATIVA PER LA PRODUZIONE STATISTICA DI INDICATORI SUL TRASPORTO PASSEGGERI PER MODO DI TRASPORTO

Le fonti normative per la produzione statistica per modo di trasporto sono:

a. Regolamenti (CE) 437/2003 e 1358/2003 (Trasporto aereo).

Il regolamento (CE) n. 437/2003 ha l'obiettivo di stabilire una solida base statistica al fine di definire le politiche dei trasporti aerei dell'Unione Europea. Il fabbisogno di dati statistici comparabili, coerenti, sincronizzati e regolari circa il livello e le dinamiche dei trasporti aerei di passeggeri, merci e posta ha portato alla definizione di tutti gli aspetti necessari per produrre indicatori standardizzati. Il succitato Regolamento è integrato dal regolamento attuativo (CE) n. 1358/2003 che contiene importanti informazioni metodologiche, quali ad esempio le categorie di aeroporti soggette agli obblighi di rendicontazione, ulteriori definizioni e linee guida per la codifica. La raccolta comune di dati su base comparabile o armonizzata permette di disporre di un sistema integrato con informazioni affidabili, coerenti e tempestive.

Ogni paese dell'UE deve raccogliere dati statistici su: passeggeri, merci e posta, voli, posti a sedere disponibili per i passeggeri, movimenti di aeromobili, di tutti gli aeroporti nel suo territorio che abbiano un traffico superiore a 150.000 unità di passeggeri all'anno. Aeroporti con un numero di passeggeri tra 15.000 e 150.000 devono raccogliere statistiche con minore dettaglio, mentre quelli con meno di 15.000 passeggeri non sono tenuti a

raccoglierle. Per ridurre al minimo l'onere per i dichiaranti, la raccolta dei dati si basa, nei limiti del possibile, su fonti già disponibili.

b. Direttiva n. 2009/42/CE del Parlamento europeo e del Consiglio; REGOLAMENTO UE N. 1090/2010 del Parlamento europeo e del Consiglio; REGOLAMENTO UE n. 1239/2019 (Trasporto marittimo).

Per garantire una rappresentazione armonizzata del fenomeno del trasporto marittimo attraverso indicatori statistici, la normativa europea (Direttiva 2009/42/CE e Regolamento UE 1090/2010) prevede l'implementazione di una rilevazione continua a carattere censuario degli arrivi e delle partenze delle navi nei porti italiani, delle merci caricate e scaricate e dei passeggeri trasportati.

L'universo di riferimento sono i porti italiani, per ciascuno dei quali vengono osservati e registrati gli eventi di arrivo e partenza delle navi e di imbarco e sbarco dei passeggeri (e delle merci). La raccolta dati avviene presso Imprese del settore (agenti marittimi, armatori, spedizionieri, raccomandatari).

Il Regolamento UE n. 1239/2019 istituisce un sistema di interfaccia unica marittima europea (EMSWe), introducendo un singolo ambiente interoperabile marittimo europeo con interfacce armonizzate al fine di semplificare l'ottemperanza agli obblighi di segnalazione per le navi che attraccano, stazionano e partono dai porti dell'Unione. Mira a miglio-

rare la competitività e l'efficienza del settore del trasporto marittimo europeo riducendo gli oneri amministrativi e introducendo un sistema di informazione digitale semplificato.

L'EMSWe è il quadro giuridico e tecnico per la trasmissione elettronica di informazioni sugli obblighi di segnalazione per le navi che fanno scalo nei porti dell'UE. Si tratta di una rete di ambienti nazionali marittimi di natura univoca con interfacce di segnalazione armonizzate e include scambi di dati tramite SafeSeaNet e altri sistemi, nonché servizi per: registro utenti e gestione degli accessi; servizio di indirizzamento comune; banca dati EMSWe delle navi; banca dati sulle posizioni comuni; banca dati comune sui materiali pericolosi («hazmat»); banche dati sul risanamento delle navi. Il regolamento gestisce l'attuale sistema nazionale marittimo esistente in ciascun paese dell'Unione come base per un EMSWe tecnologicamente neutro e interoperabile.

Alla Commissione europea è conferito il potere di adottare atti delegati per stabilire un nuovo insieme di dati EMSWe comuni, che incorporano e adattano i requisiti più pertinenti nella legislazione nazionale o dell'Unione, per armonizzare i sistemi nazionali esistenti e ridurre la necessità di supporti cartacei.

c. Regolamento CE n. 91/2003; Regolamento UE n. 2032/2016; Regolamento UE n. 643/2018 (Trasporto Ferroviario)

Il Regolamento UE n.643/2018 costituisce la rifusione e la sostituzione del Regolamento CE n. 91/2003 (e successive modifiche). Si applica a tutte le ferrovie dell'UE. Definisce norme comuni per la produzione di statistiche dei trasporti ferroviari a livello dell'UE. I Paesi membri devono fornire statistiche per tutti i trasporti ferroviari sul loro territorio e ripartirle per paese se il servizio è internazionale. Possono essere escluse dalle statistiche le società che:

- operano all'interno di zone industriali o porti;

- forniscono servizi turistici locali, come i treni storici a vapore.

Gli indicatori hanno cadenza trimestrale, annuale o quinquennale (si veda il successivo Capitolo 4 "Le fonti dati sulla mobilità in Italia"). Le informazioni nazionali raccolte da un ente pubblico o privato possono provenire da: indagini obbligatorie; dati amministrativi o regolamentari; procedure di stima a fini statistici; organizzazioni professionali del settore ferroviario; studi specifici.

Eurostat sviluppa e aggiorna la metodologia armonizzata per garantire dati di qualità. La Commissione necessita di statistiche in ambito ferroviario per monitorare e sviluppare la politica comune dei trasporti, comprese le reti trans-europee, e intervenire per migliorare la sicurezza dei trasporti ferroviari. Le norme e i concetti comuni garantiscono che le statistiche nazionali siano confrontabili e che si evitino duplicazioni.

d. Altre fonti

Infine, la Legge 1085 del 1967 all'art. 3 recita:

"Il Ministero dei trasporti e dell'aviazione civile curerà l'elaborazione di un conto nazionale dei trasporti nel quale siano considerate le spese che lo Stato, altri enti pubblici ed i privati sostengono per l'esercizio e per gli investimenti nei settori di competenza del Ministero dei trasporti e dell'aviazione civile, sia globalmente sia per i singoli mezzi, ai fini della determinazione delle direttive di politica dei trasporti e secondo le indicazioni del piano economico nazionale."

Pertanto, annualmente il MIT pubblica il Conto Nazionale delle Infrastrutture e dei Trasporti (CNIT) che, oltre a soddisfare quanto richiesto dalla fonte normativa citata, contiene in aggiunta e sin dalle prime edizioni, una raccolta sempre più vasta e il più possibile esaustiva di ulteriori indicatori prodotti da più fonti per illustrare il complesso ed articolato settore dei trasporti e della mobilità.

3.2 | RIFERIMENTI NORMATIVI PER LA PRODUZIONE DI INDICATORI SUI COMPORTEMENTI DI MOBILITÀ

Con riguardo alla produzione di indicatori armonizzati sui comportamenti di mobilità, l'unico riferimento metodologico/procedurale, peraltro di recente genesi, sono le Linee guida Eurostat ("Eurostat guidelines on Passenger Mobility Statistics") che, in quanto tali, non determinano un impegno da parte dei paesi membri essendo, per definizione, soltanto un sistema di indicazioni volto a garantire un quadro informativo armonizzato dal punto di vista della qualità e della comparabilità degli indicatori prodotti.

In ambito nazionale, un altro possibile riferimento sono le linee guida per la predisposizione dei PUMS (Piani urbani per la mobilità sostenibile). In questo documento sono definiti sia gli obiettivi strategici che devono essere soddisfatti dai PUMS, sia le batterie di indicatori da utilizzare per la loro valutazione, che comprendono indicatori infrastrutturali e di accessibilità, dell'organizzazione dei servizi

e del loro grado di soddisfazione, della domanda e offerta di mobilità, dell'incidentalità. A differenza delle Linee guida Eurostat, l'obiettivo qui non è strettamente legato alla produzione di informazione statistica, pertanto non sono definiti gli elementi metodologici necessari per realizzarla.

Inoltre, con il Decreto Interministeriale n. 179 del 12/5/2021, recante "Modalità attuative delle disposizioni relative alla figura del mobility manager" è stato definito il perimetro informativo per l'attività di gestione e ottimizzazione da parte dei Mobility Manager aziendali, scolastici e di area, della mobilità per motivi di lavoro (e della mobilità scolastica, solo su base volontaria) ai fini di ridurre l'uso del mezzo privato, con la definizione dei Piani di Spostamento Casa-Lavoro (PSCL) e le linee guida per la loro predisposizione e monitoraggio.

3.2.1 | LINEE GUIDA EUROSTAT PER LE STATISTICHE SULLA MOBILITÀ DELLE PERSONE (EUROSTAT GUIDELINES ON PASSENGER MOBILITY STATISTICS)

Il manuale è un supporto tematico e metodologico, che esamina tutti gli aspetti utili a circoscrivere il campo di osservazione e a impostare la produzione armonizzata di informazione sui comportamenti di mobilità delle persone. In esso sono riportati: le definizioni concordate dai Paesi membri per descrivere i suddetti fenomeni (concetti, variabili, classificazioni, metodologie); i set di indicatori di riferimento ('minimi', ossia uno schema di base delle disaggregazioni da adottare al fine di dare una rappresentazione complessiva del fenomeno e 'ottimali', con un dettaglio più disaggregato delle modalità per alcuni domini di stima, di riferimento per la fornitura delle stime a Eurostat anche in relazione ai progetti

finanziabili con grant agreement); i parametri di qualità; i metadati e le procedure impiegate dai Paesi che già conducono la rilevazione, inclusi fac-simile dei questionari utilizzati per la raccolta dei dati.

Le Linee Guida sono strutturate seguendo un'impostazione basata sulle categorie di mobilità per distanza. In particolare, si distinguono gli spostamenti a breve percorrenza (mobilità 'locale'), dai viaggi su distanze medio-lunghe. Questo approccio individua due fenomeni nei comportamenti di mobilità complementari ma differenti, che necessitano di un contesto definitorio e metodologico differenziato per una corretta misurazione e lettura.



I viaggi a breve percorrenza (fino a 300 km), che includono il sotto-dominio specifico della mobilità urbana, sono in genere spostamenti quotidiani, effettuati con più modi e per diversi motivi, di durata contenuta (minuti, ore) e come strumento di rilevazione elettivo hanno il 'diario'.

I viaggi a percorrenza medio-lunga sono invece meno frequenti, prevedono l'utilizzo di modi in parte differenti rispetto ai trasporti locali (es. aereo), e possono comportare o meno il pernottamento. L'indagine retrospettiva è la tecnica di rilevazione da preferire per questa categoria della mobilità.

I set di indicatori individuati nelle Linee Guida sono differenziati per le due categorie di mobilità.

Gli indicatori per la mobilità locale hanno come criterio di categorizzazione il paradigma della mobilità "urbana" e "non urbana" (secondo la definizione convenuta, che fa

riferimento all'Area Urbana Funzionale o FUA - Functional Urban Area. È possibile utilizzare una definizione alternativa, dove la "mobilità locale" è definita dagli spostamenti entro i 300 km, di cui si definiscono "mobilità urbana" quelli entro i 100 km). Gli indicatori per i viaggi a medio-lungo raggio hanno come criterio la mobilità per classi di distanza (media - da 301 a 999 km e lunga - oltre i 1000 km), e i viaggi sono ulteriormente distinti se vi è presenza o assenza di pernottamento.

La popolazione di riferimento è quella di età compresa fra i 15 e gli 84 anni.

Le dimensioni fondamentali sono: il viaggio, lo spostamento, la tappa, la distanza, la durata, il modo (o modalità) e il mezzo, la motivazione (secondo una classificazione concordata). Gli indicatori fanno riferimento a tutti i giorni della settimana, distinti nei domini: giorni lavorativi, non lavorativi e totale.

I set di indicatori non prevedono l'elabora-

**ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE
ATTRAVERSO L'USO DI BIG DATA**

zione di matrici O/D, tuttavia l'acquisizione di informazioni puntuali georeferenziate (indirizzi e/o coordinate di origine e destinazione degli spostamenti, CAP) può essere utile per individuare con maggiore precisione sia i

domini di stima (mobilità urbana/non urbana) che le distanze percorse.

I set di indicatori individuati nelle linee guida sono riportati nel prospetto di riepilogo:

Gruppo di indicatori	Modalità di riferimento	
	Mobilità locale	Distanze medio-lunghe
1. QUALITÀ DELL' INDAGINE	<ul style="list-style-type: none"> • Dimensione del campione • Popolazione di riferimento • Tasso di risposta • Campione netto (numero di rispondenti) • Quota di viaggiatori • Numero totale di spostamenti 	<ul style="list-style-type: none"> • Dimensione del campione • Popolazione di riferimento • Tasso di risposta • Campione netto (numero di rispondenti) • Numero totale di viaggi
2. INDICATORI DEL NUMERO DI SPOSTAMENTI/ VIAGGI PER PERSONA	Numero di spostamenti per persona/giorno distinti per: <ul style="list-style-type: none"> • Mobilità urbana/totale • Modalità prevalente • Motivazione 	Numero di viaggi per persona/anno distinti per: <ul style="list-style-type: none"> • Classi di distanza • Modalità prevalente • Motivazione
3. INDICATORI DI DISTANZA PER PERSONA	Distanza media per persona/giorno distinta per: <ul style="list-style-type: none"> • Mobilità urbana/totale • Giorno lavorativo/non lavorativo • Modalità • Alimentazione del veicolo (autovettura) • Motivazione 	Distanza media per persona/anno distinta per: <ul style="list-style-type: none"> • Modalità • Alimentazione del veicolo (autovettura) • Motivazione
4. INDICATORI DI DURATA DELLO SPOSTAMENTO/ VIAGGIO	Tempo medio di viaggio per persona/giorno distinto per: <ul style="list-style-type: none"> • Mobilità urbana/totale • Giorno lavorativo/non lavorativo • Modalità • Motivazione 	Numero totale di pernottamenti
5. PASSEGGERI-KM ANNUI PER LA POPOLAZIONE DI RIFERIMENTO	Km totali percorsi (riportati alla popolazione di riferimento e all'anno solare) distinti per: <ul style="list-style-type: none"> • Mobilità urbana/totale • Giorno lavorativo/non lavorativo • Modalità • Alimentazione del veicolo (autovettura) • Motivazione 	Km totali percorsi (riportati alla popolazione di riferimento e all'anno solare) distinti per: <ul style="list-style-type: none"> • Modalità • Alimentazione del veicolo (autovettura) • Motivazione
6. TASSO DI RIEMPIMENTO DEL VEICOLO (PER AUTOVETTURE E TAXI)	Tasso di riempimento del veicolo distinto per: <ul style="list-style-type: none"> • Mobilità urbana/totale • Giorno lavorativo/non lavorativo 	Tasso di riempimento del veicolo per autovetture e taxi

Tabella 1-Indicatori individuati nelle Linee Guida Eurostat (Eurostat guidelines on Passenger Mobility Statistics)

3.2.2 | NORMATIVA E LINEE GUIDA IN AMBITO NAZIONALE

- a. **Normativa e linee guida per i Piani Urbani di Mobilità Sostenibile (PUMS):**
 - **DM MIT 397/2017** "Individuazione delle linee guida per i piani urbani di mobilità sostenibile, ai sensi dell'articolo 3, comma 7, del decreto legislativo 16 dicembre 2016, n. 257" e
 - **DM MIT 396/2019** "Modifica delle linee guida per la redazione dei piani urbani della mobilità sostenibile (PUMS), di cui al decreto ministeriale 397/2017".
 - **Vademecum** per la redazione del piano urbano di mobilità sostenibile (PUMS) redatto dal Gruppo di lavoro fra esperti della Struttura tecnica di missione e della DG TPL del MIT e del Politecnico di Milano, pubblicato a settembre 2022.

Queste tre fonti normativo-procedurali individuano una serie di azioni finalizzate alla predisposizione e al monitoraggio dei Piani Urbani per la Mobilità Sostenibile (PUMS). Dal Vademecum:

"[Definizione, dall'Allegato I al DM 397/2017] Il PUMS è uno strumento di pianificazione strategica che, in un orizzonte temporale di medio-lungo periodo (10 anni), sviluppa una visione di sistema della mobilità urbana (preferibilmente riferita all'area della Città metropolitana, laddove definita), proponendo il raggiungimento di obiettivi di sostenibilità ambientale, sociale ed economica attraverso la definizione di azioni orientate a migliorare l'efficacia e l'efficienza del sistema della mobilità e la sua integrazione con l'assetto e gli sviluppi urbanistici e territoriali."

[...] Per le città metropolitane, i comuni e le associazioni di comuni con più di 100.000 abitanti, le Linee guida italiane stabiliscono l'obbligo di adozione del PUMS (art.3, comma 1 del D.M. 397/2017 e s.m.i.)

[...] Per le città metropolitane e i comuni con popolazione superiore a 100.000 abitanti, non ricompresi nel territorio di città metro-

politane, l'adozione del PUMS è anche una condizione indispensabile per accedere ai finanziamenti statali destinati a nuovi interventi per il trasporto rapido di massa e per la mobilità ciclistica.

[...] Le Linee guida italiane costituiscono il principale riferimento normativo da seguire per la redazione del PUMS e sono costituite da:

- A. procedura uniforme per la redazione e approvazione dei PUMS, articolata in 8 passi procedurali e definita nell'Allegato 1 al D.M. 397/2017 e s.m.i.;
- B. individuazione delle strategie di riferimento, degli obiettivi macro e specifici, delle azioni e degli indicatori da utilizzare per il monitoraggio, secondo l'Allegato 1 al D.M. 397/2017 e s.m.i.

[...] processo di pianificazione che può essere suddiviso in 2 blocchi principali: attività preliminari e vera e propria redazione del PUMS.

Le attività preliminari sono definite nei passi procedurali dedicati a:

- Definizione del gruppo interdisciplinare/interistituzionale di lavoro
- Progettazione del processo partecipato
- [...] La redazione del PUMS si articola poi in quattro passi procedurali ben definiti:
- Predisposizione del quadro conoscitivo
- Definizione degli obiettivi
- Costruzione dello scenario di piano
- Definizione del piano di monitoraggio
- [...] Le Linee guida italiane prevedono altri due passi procedurali:
- Valutazione ambientale strategica (VAS)
- Adozione ed approvazione del PUMS

Sempre nei riferimenti citati, troviamo che:

[...] Il PUMS ha come perimetro geografico minimo quello dei confini amministrativi dell'ente obbligato alla redazione del PUMS. Tuttavia, tale perimetro può essere ampliato e a tal fine si può far riferimento al concetto di Aree Urbane Funzionali adottato dalle Linee guida europee.

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

Pertanto, vi è una convergenza fra la definizione territoriale adottata dalla normativa per i PUMS e quella delle Linee Guida Eurostat per le statistiche sulla mobilità delle persone (vedi sopra p. 3.2.1, alla definizione della mobilità urbana).

La predisposizione del quadro conoscitivo prevede l'analisi dei flussi di mobilità attraverso vari strumenti, dalla raccolta di informazioni presso gli attori del TPL e dei servizi di vehicle sharing, alla realizzazione di indagini dirette sui cittadini; esso prevede inoltre l'utilizzo di modelli trasportistici per la stima di una o più

matrici O/D (generale, per modo di trasporto, per orario di punta/morbida etc).

È prevista, nel piano procedurale, la definizione di macro-obiettivi (17 obiettivi obbligatori definiti nelle Linee Guida italiane, più eventuali facoltativi in aggiunta), obiettivi specifici funzionali al raggiungimento dei macro-obiettivi, e specifici della realtà locale.

Per ogni macro-obiettivo sono previste delle batterie di indicatori di risultato, per ciascuno dei quali è necessario definire le fonti e i target previsti a breve (2-3 anni), medio (5 anni) e lungo termine (10 anni).

28

Area di interesse	Macro-obiettivi minimi obbligatori
A) Efficacia ed efficienza del sistema di mobilità	a.1. Miglioramento del TPL
	a.2. Riequilibrio modale della mobilità
	a.3. Riduzione della congestione
	a.4. Miglioramento della accessibilità di persone e merci
	a.5. Miglioramento dell'integrazione tra lo sviluppo del sistema della mobilità e l'assetto e lo sviluppo del territorio (insediamenti residenziali e previsioni urbanistiche di poli attrattori commerciali, culturali, turistici)
	a.6. Miglioramento della qualità dello spazio stradale e urbano
B) Sostenibilità energetica e ambientale	b.1. Riduzione del consumo di carburanti tradizionali diversi dai combustibili alternativi
	b.2. Miglioramento della qualità dell'aria
	b.3. Riduzione dell'inquinamento acustico
C) Sicurezza della mobilità stradale	c.1. Riduzione dell'incidentalità stradale
	c.2. Diminuzione sensibile del numero generale degli incidenti con morti e feriti
	c.3. Diminuzione sensibile dei costi sociali derivanti dagli incidenti
	c.4. Diminuzione sensibile del numero degli incidenti con morti e feriti tra gli utenti deboli (pedoni, ciclisti, bambini e over 65)
D) Sostenibilità socio economica	d.1. Miglioramento della inclusione sociale
	d.2. Aumento della soddisfazione della cittadinanza
	d.3. Aumento del tasso di occupazione
	d.4. Riduzione dei costi della mobilità (connessi alla necessità di usare il veicolo privato)

Figura 2 – Macro obiettivi del PUMS (Fonte del prospetto: Vademecum per la redazione del Piano Urbano di Mobilità Sostenibile)



Ai macro-obiettivi obbligatori si possono affiancare altri macro-obiettivi aggiuntivi, ed eventualmente degli obiettivi specifici per il sistema territoriale interessato.

Per ognuno di questi, è necessario definire delle batterie di indicatori per monitorare e valutare le azioni intraprese con il Piano. Le Linee Guida PUMS europee propongono dei set di indicatori comuni per i macro-obiettivi obbligatori, mentre per i macro-obiettivi aggiuntivi e per gli obiettivi specifici, sono pro-

posti gli indicatori riportati nel DM 397/2017 (Tabella 2, Allegato 2) e gli indicatori per la Mobilità Urbana Sostenibile sviluppati dalla Commissione Europea.

Il set di indicatori per i macro-obiettivi obbligatori è riportato nel Vademecum, e fra questi sono di particolare interesse in relazione alla misurazione dei flussi di persone, quelli relativi ai macro-obiettivi a.1, a.2 e a.3 di seguito riportati con le possibili fonti:

**ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE
ATTRAVERSO L'USO DI BIG DATA**

Macro obiettivo	Definizione indicatore	Indicatori da DM 396/2019 (ed eventuale metodologia di calcolo)
a.1. - Miglioramento del TPL	<i>Passengeri trasportati</i>	Aumento dei passeggeri trasportati
a.2.- Riequilibrio modale della mobilità		% di spostamenti in autovettura
		% di spostamenti sulla rete integrata del TPL
		% di spostamenti in ciclomotore/motoveicolo (mezzo privato)
		% di spostamenti in bicicletta (mezzo privato)
		% di spostamenti a piedi
a.3.- Riduzione della congestione	<i>Congestion - RETE PRIMARIA</i>	a.3. - Riduzione della congestione - RETE PRIMARIA
a.4.- Miglioramento della accessibilità di persone e merci	<i>Indice dell'accessibilità delle persone - TPL</i>	a.4.a.-Miglioramento della accessibilità di persone - TPL
	<i>Indice dell'accessibilità delle persone - Sharing</i>	a.4.b.-Miglioramento della accessibilità di persone - Sharing
	<i>Indice dell'accessibilità delle persone - Taxi NCC</i>	a.4.c.- Miglioramento accessibilità persone servizi mobilità taxi e NCC
		a.4.d.-Accessibilità - pooling
	<i>Indice dell'accessibilità - Merci</i>	a.4.e-Miglioramento della accessibilità sostenibile delle merci



30

Figura 3 – Indicatori per la misurazione del raggiungimento dei macro-obiettivi PUMS (Fonte del prospetto: Vademecum per la redazione del piano urbano di mobilità sostenibile)

LINEE GUIDA E NORMATIVE DI RIFERIMENTO PER LA RILEVAZIONE DEI DATI SULLA MOBILITÀ

		Unità di misura	Possibili fonti dati (o metodologie di raccolta dati)
		n. passeggeri/anno/1000 abitanti	Numeratore: Operatori locali TPL; Denominatore: Anagrafe comunale
		adimensionale	Indagine a campione
		adimensionale	
		adimensionale	
		adimensionale	
	<p>Rapporto tra il tempo complessivo impiegato su rete stradale congestionata ed il tempo complessivo "virtuale" impiegato in assenza di congestione:</p> <ul style="list-style-type: none"> Tempo di percorrenza su rete congestionata: calcolata in una fascia oraria di punta concordata tra le 7.30 e le 9.30 da prendersi per un minimo di 10 giorni nei giorni infrasettimanali (martedì/mercoledì/giovedì) feriali dei mesi di ottobre e novembre su un campione rappresentativo di almeno il 10% della rete. Tempo di percorrenza sulla rete in assenza di congestione: sarà la stessa rete in orario notturno alla velocità massima consentita dal codice della strada. 	adimensionale	Elaborazioni modellistiche (possono essere utilizzate, nelle realtà più piccole dati provenienti da provider open source di navigazione elaborate dalle strutture).
	Sommatoria numero popolazione residente che vive a 250 metri da fermata autobus e filobus, 400 da fermata tram e metro e 800 metri da stazione ferroviaria.	numero assoluto	Elaborazioni GIS
	Numero di veicoli condivisi (auto, bici, e moto)/popolazione residente	n. veicoli/popolazione residente	Numeratore: Operatori locali sharing; Denominatore: Anagrafe comunale
		n. licenze/popolazione residente	Numeratore: Comune. Motorizzazione civile; Denominatore: Anagrafe comunale
	Forme di incentivi al pooling censiti	n° di incentivi al pooling	Comune
	n. veicoli commerciali "sostenibili" (cargobike, elettrico, metano, idrogeno) attivi in ZTL/kmq tot. di ZTL - ora	n. veicoli commerciali attivi in ZTL rispetto alla sua estensione (kmq) nell'unità di tempo (ora)	Comune

La normativa prevede che il monitoraggio debba essere eseguito con cadenza almeno biennale. A questo scopo le amministrazioni dovrebbero dotarsi di un *Cruscotto di Monitoraggio*, ossia di un sistema di raccolta ed elaborazione dati in modo da poter avere il quadro completo degli indicatori necessari.

b. Normativa e linee guida per i Piani di Spostamento Casa Lavoro (PSCL):

- **Legge 221/2015** "Disposizioni in materia ambientale per promuovere misure di green economy e per il contenimento dell'uso eccessivo di risorse naturali." (art. 5 c. 6, istituzione della figura del Mobility Manager);
- **Decreto Interministeriale MITE/MIT n. 179 del 12/5/2021** recante "Modalità attuative delle disposizioni relative alla figura del Mobility Manager" e successive modifiche;
- **DL 16 giugno 2022 - convertito in L. 108/2022; Linee guida** per la redazione e l'implementazione dei piani degli spostamenti casa-lavoro (PSCL).

Il **DM 179/2021** stabilisce l'istituzione, obbligatoria nei Comuni capoluogo di Regione, nelle Città Metropolitane, nei Comuni capoluogo di Provincia e nei Comuni con po-

polazione superiore a 50.000 abitanti, delle figure del **Mobility Manager (MM)**, con funzioni di supporto professionale continuativo alle attività di decisione, pianificazione, programmazione, gestione e promozione di soluzioni ottimali di mobilità sostenibile:

- **MM Aziendale**, (nelle imprese e nelle Pubbliche Amministrazioni di cui all'art. 1, comma 2, del Decreto Legislativo 30 marzo 2001, n. 165, con singole unità locali con più di 100 dipendenti) figura specializzata nel governo della domanda di mobilità e nella promozione della mobilità sostenibile nell'ambito degli spostamenti casa-lavoro del personale dipendente;
- **MM di Area**, figura specializzata nel supporto al Comune territorialmente competente, presso il quale è nominato, nella definizione e implementazione di politiche di mobilità sostenibile, nonché nello svolgimento di attività di raccordo tra i MM Aziendali.

Le imprese e le pubbliche amministrazioni che non rientrano tra quelle richiamate nella definizione possono comunque procedere facoltativamente alla nomina del Mobility Manager aziendale.



- **MM Scolastico** (DECRETO LEGGE 16 giugno 2022, n. 68 art. 8 c. 12bis - convertito in L. 108/2022):

“Le istituzioni scolastiche, singolarmente o in rete, individuano il Mobility Manager scolastico tra il personale docente, senza esonero dall’insegnamento, ovvero ricorrendo a figure professionali esterne, in coerenza con il piano dell’offerta formativa. Il MM scolastico ha il compito di:

- *diffondere la cultura della mobilità sostenibile;*
- *promuovere l’uso della mobilità ciclo-pedonale e dei servizi di noleggio e condivisione di veicoli elettrici o a basso impatto ambientale;*
- *supportare il MM di Area, ove nominato, e le competenti amministrazioni locali ai fini dell’adozione delle misure di mobilità sostenibile, fornendo elementi per favorire la sostenibilità degli spostamenti del personale scolastico e degli studenti dell’istituto scolastico;*
- *segnalare al competente ente locale eventuali esigenze legate al trasporto scolastico e delle persone con disabilità.”*

Il MM di Area, insieme ai MM Aziendali, ha il compito di predisporre il piano degli spostamenti casa-lavoro (PSCL), ossia uno strumento di pianificazione degli spostamenti sistematici casa-lavoro del personale dipendente, a partire da quelli relativi ad ogni singola unità locale lavorativa. Per facilitare questo compito, sono state predisposte delle Linee Guida, in cui sono definiti anche gli indicatori per il monitoraggio.

Dalle Linee Guida:

“È necessario raccogliere tutte le informazioni ed i dati relativi alle esigenze di mobilità del personale e conoscere le condizioni strutturali aziendali, l’offerta di trasporto sul territorio, nonché le risorse disponibili per l’attuazione delle possibili misure utili a migliorare la mobilità del personale. La parte informativa e di analisi del PSCL deve contenere:

- *Analisi delle condizioni strutturali azien-*

dali e dell’offerta di trasporto;

- *Analisi degli spostamenti casa-lavoro”*

[...]

“Al fine di ricostruire il quadro degli spostamenti casa-lavoro delle sedi aziendali è necessario, preliminarmente, inquadrare il personale dipendente in relazione a residenza e tipologia di turnazione aziendale. Ai fini dell’analisi della distribuzione territoriale delle residenze dei dipendenti, la forza lavoro può essere disaggregata per “Codice di Avviamento Postale”, oppure per appartenenza alle “zone di traffico”, in coerenza con la zonizzazione del territorio adottata dai modelli di simulazione della mobilità disponibili presso l’Amministrazione comunale.

Relativamente alla turnazione, il personale può essere classificato in funzione dei tipi di turni lavorativi tipici: i parametri di riferimento sono i giorni lavorativi e gli orari di inizio e fine dei turni di lavoro.

[...] *Per indagare gli elementi utili a comprendere le abitudini e le esigenze di spostamento dei dipendenti, nonché la loro propensione al cambiamento, è necessario che il mobility manager aziendale realizzi anche una specifica campagna di raccolta dati, tramite un questionario informativo da somministrare a ciascun dipendente.”*

Pertanto, la redazione del PSCL prevede una vera e propria fase di rilevazione statistica puntuale della mobilità abituale. Le indicazioni sulle **informazioni minime necessarie da raccogliere**, sono riportate nell’Allegato 3 delle Linee Guida, e prevedono un dettaglio fine sia sulla localizzazione territoriale degli spostamenti, sia sui mezzi utilizzati (anche in combinazione multimodale) e sulla motivazione di tali scelte, sia sulla struttura dei comportamenti di mobilità (orari, frequenza settimanale, distanza e durata degli spostamenti, etc).

Allegato 3 - Scheda informativa sugli spostamenti casa-lavoro dei dipendenti

Per ciascuna sede di lavoro, con più di 100 dipendenti, è necessario rilevare, attraverso una specifica campagna di indagine, gli elementi utili a comprendere le abitudini e le esigenze di spostamento dei dipendenti, nonché la loro propensione al cambiamento verso forme di mobilità sostenibili.

Si riportano di seguito le *informazioni minime necessarie da rilevare*.

1. ANAGRAFICA

- 1.1. COMUNE DI RESIDENZA/DOMICILIO
- 1.2. CAP
- 1.3. SESSO
- 1.4. ETÀ

2. ATTIVITÀ LAVORATIVA

- 2.1. TIPOLOGIA DI CONTRATTO (tempo pieno, part time)
- 2.2. PRESENZA IN SEDE (per ciascun giorno della settimana)
 - 2.2.1. Orario di ingresso
 - 2.2.2. Orario di uscita

3. MEZZI DI TRASPORTO E ABBONAMENTI AI SERVIZI DI TRASPORTO NELLE DISPONIBILITÀ DEL DIPENDENTE

- 3.1. DISPONIBILITÀ DI MEZZI DI TRASPORTO (specificare)
- 3.2. DISPONIBILITÀ DI ABBONAMENTI AI SERVIZI DI TRASPORTO (specificare)

4. SPOSTAMENTI CASA-LAVORO

- 4.1. MODALITÀ ABITUALE DI SPOSTAMENTO
 - 4.1.1. Utilizzo di un solo mezzo/modo di trasporto (specificare)
 - 4.1.2. Combinazione di più mezzi/modalità di trasporto (specificare)
- 4.2. DISTANZA PERCORSA
- 4.3. TEMPO IMPIEGATO
- 4.4. MOTIVAZIONI SULLA MODALITÀ ABITUALE DI SPOSTAMENTO
- 4.5. GRADO DI SODDISFAZIONE SULLA MODALITÀ ABITUALE DI SPOSTAMENTO

5. PROPENSIONE AL CAMBIAMENTO NEGLI SPOSTAMENTI CASA-LAVORO: MOTIVI E CONDIZIONI VERSO FORME DI MOBILITÀ SOSTENIBILI

- 5.1. CAMBIAMENTO A FAVORE DEL TRASPORTO PUBBLICO
- 5.2. CAMBIAMENTO A FAVORE DELL'USO DELLA BICICLETTA
- 5.3. CAMBIAMENTO A FAVORE DEL CAR POOLING
- 5.4. CAMBIAMENTO A FAVORE DEL CAR SHARING
- 5.5. CAMBIAMENTO A FAVORE DELLO SCOOTER SHARING
- 5.6. CAMBIAMENTO A FAVORE DEL BIKE SHARING
- 5.7. CAMBIAMENTO A FAVORE DELL'USO DI MONOPATTINI IN SHARING

3.3 | QUADRO SINTETICO DELLE NORMATIVE PER LE STATISTICHE SULLA MOBILITÀ DELLE PERSONE, CON RIFERIMENTO ALLA PRODUZIONE STATISTICA

L'analisi delle fonti normative di cui ai paragrafi 0 e 3.2 può essere sintetizzata nel prospetto che segue, per una visione di insie-

me, in raccordo con la produzione statistica esistente o potenziale.

Norme o linee guida specifiche per le statistiche sulla mobilità delle persone	PSN	Descrizione	Produzione statistica sulla mobilità delle persone	Ente produttore	Attuazione della produzione
Legge 1085 del 31/10/1967	SI	Inserisce fra i compiti istituzionali del Ministero dei Trasporti la pubblicazione di una raccolta di informazioni statistiche su alcuni fenomeni relativi ai trasporti e alla mobilità	1) Trasporto pubblico locale (MIT-00018) 2) Collegamenti marittimi con le Isole (PSN MIT-00024)	MIT	in essere
Regolamenti (CE) 437/2003 e 1358/2003	SI	Stabiliscono una solida base statistica al fine di definire le politiche dei trasporti aerei dell'Unione europea (UE) con dati comparabili, coerenti, sincronizzati e regolari, attraverso elementi metodologici e definitori comuni.	Indagine sul trasporto aereo (PSN IST-00145)	ISTAT	in essere
1) Direttiva n. 2009/42/CE del Parlamento europeo e del Consiglio 2) REGOLAMENTO UE N. 1090/2010 del Parlamento europeo e del Consiglio 3) REGOLAMENTO UE n. 1239/2019	SI	Produzione e diffusione di statistiche relative al trasporto marittimo	Trasporto Marittimo (PSN IST-00818)	ISTAT	in essere

**ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE
ATTRAVERSO L'USO DI BIG DATA**

Norme o linee guida specifiche per le statistiche sulla mobilità delle persone	PSN	Descrizione	Produzione statistica sulla mobilità delle persone	Ente produttore	Attuazione della produzione
1) Regolamento CE n. 91/2003 2) Regolamento UE n. 2032/2016 3) Regolamento UE n. 643/2018	SI	Norme comuni per la produzione di statistiche dei trasporti ferroviari a livello dell'UE.	Trasporto Ferroviario (IST-01646)	ISTAT	in essere
Eurostat Guidelines on Passenger mobility statistics	SI	Metodologia di riferimento definita in ambito UE per la produzione statistica armonizzata di set di indicatori sulla mobilità delle persone.	Audimob - Indagine su stili e comportamenti di mobilità dei residenti in Italia (PSN IFT-0001)	ISFORT	in essere
1) DM MIT 397/2017 2) DM MIT 396/2019 3) Vademecum per la redazione del piano urbano di mobilità sostenibile (pums)	SI	Normativa e linee guida per i Piani Urbani di Mobilità Sostenibile (PUMS): individuano una serie di azioni finalizzate alla predisposizione e al monitoraggio dei Piani Urbani per la Mobilità Sostenibile (PUMS).	1) Monitoraggio a 2-5-10 anni dei PUMS 2) Rilevazione dati ambientali nelle città (PSN IST-00907)	1) Città metropolitane, Comuni e associazioni di Comuni con più 100.000 abitanti (obbligatorio); Facoltativo per gli altri Comuni 2) ISTAT	potenziale (in parte in essere)
1) Legge 221/2015 2) DM MITE/MIT 179/2021 3) Legge 108/2022 4) Linee guida per la redazione e l'implementazione dei piani degli spostamenti casa-lavoro (PSCL).	NO	Istituzione della figura del Mobility Manager e dell'implementazione e monitoraggio dei piani degli spostamenti casa-lavoro (PSCL)	Monitoraggi periodici dei PSCL	Comuni capoluogo di Regione, nelle Città Metropolitane, nei Comuni capoluogo di Provincia e nei Comuni con popolazione superiore a 50.000 abitanti	potenziale (in parte in essere)

Tabella 2 – Sintesi delle fonti normative relative alla produzione statistica

4 | LE FONTI DATI SULLA MOBILITÀ IN ITALIA

Giovanna Astori ¹, Mario Tartaglia ²[0000-0003-3216-8150]

1. ISTAT, Roma, Italia

2. FS Research Centre, Firenze, Italia

4.1 | DATI DA INDAGINE, DATI DA FONTI AMMINISTRATIVE E DA ALTRE FONTI

4.1.1 | LE RILEVAZIONI SUI COMPORAMENTI DI MOBILITÀ

La produzione statistica di indicatori sui comportamenti di mobilità presente in ambito SISTAN è abbastanza ricca e articolata, ed è una buona base di partenza per la costruzione di un panorama informativo ampio per rappresentare il fenomeno. Tuttavia, nessuna delle fonti individuate è in sé esaustiva, e sotto molti aspetti si trovano divergenze rispetto alle definizioni e alle raccomandazioni metodologiche contenute nelle Linee Guida Eurostat. Si illustrano di seguito, per ciascuna fonte, le caratteristiche principali e i possibili impieghi per le finalità considerate.

a. Osservatorio Audimob “Indagine su stili e comportamenti di mobilità dei residenti in Italia” (Isfort)

- È una rilevazione a cadenza annuale, dove si osserva principalmente il fenomeno dalla mobilità locale e urbana. A partire dal 2021 la sua metodologia di riferimento è stata completamente adeguata a quanto previsto nelle Linee Guida Eurostat. Essa prevede:
 - La popolazione di riferimento dei residenti di età 14-84 anni;
 - Un campione per quote di oltre 16.000 individui, ottimizzato con riferimento ai domini di genere, età e regione di residenza (con un sovracampionamento per le città metropolitane);
 - L’osservazione di tutti gli spostamenti effettuati nel giorno precedente l’intervista (con copertura sia dei feriali che dei festivi) tramite la compilazione di un diario

- con il dettaglio di ciascuno spostamento. Sono rilevate anche le caratteristiche socioeconomiche del rispondente, e informazioni relative ai veicoli in disponibilità;
 - Le variabili principali sono: orario, durata, lunghezza, origine, destinazione, motivazione, mezzo utilizzato (dettagliato per tappa modale), alimentazione del veicolo se autovettura privata, altri accompagnatori durante lo spostamento.;
 - La tecnica di indagine è mista: CATI (70%) CAWI (30%);
 - La mancata risposta è azzerata tramite sostituzione delle unità con riferimento alle quote.

Gli indicatori pubblicati nel rapporto Audimob hanno un riferimento temporale annuale e territoriale nazionale. Gli indicatori disponibili sono indicati di seguito per ogni variabile di dominio:

- Classi di lunghezza degli spostamenti (distribuzione %);
- Classi di frequenza degli spostamenti giornalieri (distribuzione %, criterio cluster);
- Motivo dello spostamento (distribuzione %);
- Modo di trasporto utilizzato (distribuzione %, tasso di mobilità sostenibile);
- Fasce orarie (distribuzione %);
- Sistematicità degli spostamenti (distribuzione %);
- Mobilità urbana, extraurbana di medio o lungo raggio e totale (distanza media, tempo medio, distribuzione % sposta-

menti); anche in combinazione: con la segmentazione socioanagrafica (genere, età e condizione professionale); con la segmentazione territoriale (ripartizione geografica, ampiezza demografica del comune); con motivazione, fascia oraria, sistematicità; col mezzo di trasporto.

- (distribuzione % spostamenti);
- Genere (Tasso di mobilità %);
- Fasce di età (Tasso di mobilità %);
- Condizione professionale (Tasso di mobilità %);
- Ripartizione geografica (Tasso di mobilità %);
- Ampiezza demografica del comune di residenza (Tasso di mobilità %);
- Tipologia di Comune (classificazione SNAI);
- Smart working in combinazione: con i comportamenti di mobilità (statici/ di prossimità/mobili); con motivazione, fascia oraria, sistematicità; col mezzo di trasporto; (distribuzione %);
- Smart working (numero medio, lunghezza e durata spostamenti).

A partire dal 2019 (con una post armonizzazione dei dati 2019 e 2020 e con la metodologia armonizzata a partire dal 2021) vengono forniti a Eurostat anche tutti gli indicatori previsti nelle Linee Guida con riferimento alla mobilità locale. Anche per questo set di indicatori (si veda il Capitolo 3 "Linee guida e normative di riferimento per la rilevazione dei dati sulla mobilità") il livello territoriale è quello del Paese nel suo complesso, con dettaglio di dominio per mobilità locale urbana e totale. Il riferimento temporale è annuale.

b. TUS – Indagine sull'uso del tempo - Multiscopo (Istat).

Da SiQual (Sistema informativo della qualità delle indagini – Istat):

"L'Indagine sull'Uso del Tempo si colloca nel sistema di indagini "Multiscopo" sulle famiglie. La principale peculiarità di tale rilevazione sta nel fatto che attraverso la compilazione di un diario è possibile conoscere il modo

in cui ciascun rispondente ripartisce 24 ore (suddivise in intervalli di 10 minuti) tra le varie attività giornaliere, gli spostamenti, i luoghi frequentati e le persone con cui le ha trascorse. Tale indagine è considerata strategica per la conoscenza dell'organizzazione dei tempi di vita della popolazione in un'ottica di genere, poiché consente lo studio della divisione dei ruoli nella società e nelle famiglie. Per tale ragione l'indagine è regolata dalla legge 53 del 2000, art.16 "Statistiche ufficiali sui tempi di vita: L'ISTAT assicura un flusso informativo quinquennale sull'organizzazione dei tempi di vita della popolazione attraverso la rilevazione sull'uso del tempo, disaggregando le informazioni per sesso ed età".

L'indagine ha cadenza quinquennale.

Il registro di base è LAC/ANPR.

È utilizzata la tecnica PAPI (in parte con rilevatore, in parte autocompilato). Sono attivi progetti sperimentali a livello internazionale per lo sviluppo di metodi alternativi di raccolta dati, in particolare con l'uso di personal device (App per Smartphone, etc).

Il campo di osservazione copre le attività svolte durante tutti i giorni della settimana (feriali e festivi) di un intero anno solare (sono osservati 365 giorni di attività, il campione è ripartito sull'intero anno).

Il questionario comprende una prima sezione (A) di schede individuali per ciascuno dei componenti della famiglia. Fra le altre notizie vengono richieste informazioni sulla distanza dal luogo di studio o lavoro e sul tempo impiegato per raggiungerlo, sull'eventuale partecipazione regolare ad attività di vario tipo (volontariato, sport, etc) nel tempo libero; inoltre, fra i quesiti anagrafici è richiesto un ampio dettaglio sulla nazionalità, la cittadinanza, anche con riferimento ai genitori. È prevista anche una scheda complessiva per la famiglia, in cui si richiede la disponibilità di veicoli e il loro numero per tipologia.

La sezione B del questionario è un diario giornaliero, dove il rispondente deve descrivere (compilando i campi dello schema) tutte le

attività svolte (con cadenza di 10 minuti) in un giorno fissato, dalle 4 del mattino e per tutte le 24 ore. Fra le informazioni richieste, è inclusa l'indicazione dei mezzi di trasporto utilizzati per spostarsi (inclusi i cambi di mezzo) e del motivo dello spostamento; si chiede inoltre per tutte le attività (quindi anche per gli spostamenti, che sono considerati come attività a sé stanti) l'eventuale presenza di altre persone. Infine, si richiede l'eventuale presenza di viaggi non abituali in altre località (nazionali o estere) nel periodo di osservazione, e l'indicazione della distanza percorsa. Il diario deve essere compilato per tutti i componenti della famiglia, di qualsiasi età.

Vi è poi una sezione C relativa ai tempi di lavoro, che non è di interesse con riferimento al fenomeno della mobilità.

L'indagine rileva anche la condizione professionale, se la compilazione sia relativa a una giornata 'ordinaria', di ferie, malattia etc. (informazione utile per individuare la mobilità ricorrente) e la condizione che definisce l'eventuale presenza di disabilità.

Di particolare interesse per il tema mobilità è il trattamento dei dati relativi agli spostamenti, con riferimento alle motivazioni. Lo schema logico adottato in TUS coincide con quello definito nelle Linee Guida Eurostat, anche riguardo la codifica degli spostamenti per rientrare all'abitazione principale.

L'indagine TUS può essere considerata un utile strumento di confronto per gli indicatori di mobilità su domini ampi (es. mobilità per motivazione, per modo di trasporto, per età/genere). Ai fini della rappresentazione del fenomeno della mobilità, tuttavia non rileva due fattori: il riferimento territoriale degli spostamenti e le distanze percorse. Non è pertanto possibile produrre alcuna informazione specifica di carattere territoriale - geografica o contestuale - relativa ai luoghi in cui ci si muove, salvo il riferimento al territorio di residenza di chi effettua lo spostamento.

Fra gli indicatori pubblicati con cadenza annuale (ultimo disponibile è l'anno 2013, si

prevede una nuova edizione per il 2023) di particolare interesse sono quelli relativi alle attività classificate come "Spostamenti finalizzati", ulteriormente disaggregati secondo la motivazione dello spostamento. I domini di riferimento sono:

- Il giorno di osservazione (giorno medio settimanale/feriale/sabato/domenica);
- Sesso;
- Classe di età;
- Titolo di studio;
- Ampiezza della famiglia;
- Cittadinanza (Italiano/straniero), solo per la modalità "Spostamenti finalizzati" aggregata;
- Condizione lavorativa, solo per la modalità "Spostamenti finalizzati" aggregata;
- Stato civile, solo per la modalità "Spostamenti finalizzati" aggregata;
- Posizione nel nucleo familiare, solo per la modalità "Spostamenti finalizzati" aggregata;
- Regione, solo per la modalità "Spostamenti finalizzati" aggregata;
- Ampiezza demografica del comune, solo per la modalità "Spostamenti finalizzati" aggregata;
- Fasce orarie, solo per la modalità "Spostamenti finalizzati" aggregata;
- Tipologia di coppia, solo per la modalità "Spostamenti finalizzati" aggregata;
- Mezzi di spostamento (in TUS è considerata una modalità della variabile "luoghi", e disaggregata in: a piedi/bicicletta/motocicletta, ciclomotore, scooter/autovettura privata/altro mezzo privato diverso da auto e moto/mezzi pubblici), in combinazione con sesso, età, condizione lavorativa, stato civile, posizione nel nucleo, regione, tipo di comune.

Gli indicatori disponibili sono riferiti alla variabile "tempo", quindi: durata media generica e specifica (in hh:mm), tempo dedicato all'attività di spostamento in percentuale sulle 24 ore, incidenza percentuale delle persone che hanno svolto l'attività di spostamento.

c. Censimento Permanente della Popolazione e delle abitazioni (Istat)

Dal Piano Generale del Censimento:

“La strategia del CP è basata sull’integrazione di dati amministrativi e dati da indagini statistiche campionarie, con l’obiettivo di produrre informazioni ogni anno e contenere i costi e il disturbo statistico sulle famiglie. L’obiettivo primario dell’impianto metodologico del Censimento permanente è quello di mantenere l’elevato livello di dettaglio classificatorio garantito tradizionalmente dal Censimento decennale per un insieme di variabili fondamentali (di natura demografica, sociale ed economica), aumentando, altresì, la frequenza temporale dell’informazione prodotta e la tempestività della sua diffusione. Il passaggio a un nuovo modello di Censimento è reso possibile dall’acquisizione, dal trattamento e dall’utilizzo a fini statistici di fonti amministrative che generano, mediante processi di validazione dei dati, registri statistici aggiornati con elevata frequenza temporale. Il CP si avvale sia delle informazioni prodotte dai Registri statistici che compongono il Sistema integrato di registri (SIR), sia di quelle rilevate attraverso apposite indagini periodiche [...]. In particolare, si avvarrà di due specifiche indagini campionarie: una Areale (A) e una da Lista (L). Il campo di osservazione del CP è costituito, a livello di ciascun Comune, dalla popolazione dimorante abitualmente [...]. Fanno parte della popolazione dimorante abitualmente le persone di cittadinanza straniera regolarmente soggiornanti in Italia. [...] Il CP, mediante le indagini campionarie effettuate dall’Istat e l’integrazione con il SIR, acquisisce informazioni relative alla struttura delle famiglie, alle caratteristiche anagrafiche, di stato civile, socioeconomiche e di mobilità territoriale delle persone dimoranti abitualmente.”

L’indagine da Lista, ripetuta annualmente, ha lo scopo di rilevare tutte le informazioni di interesse che non possono essere desunte dai registri amministrativi utilizzati per produrre il Registro di Base degli Individui (RBI). Il cam-

pione L è composto da 500.000 famiglie. Il registro di riferimento è RBI.

La tecnica di indagine prevede un primo contatto per il coinvolgimento dell’unità campionaria, tramite informativa postale. Successivamente, la raccolta dei dati è basata su due canali: CAWI su portale Istat dedicato (con supporto dei centri di rilevazione comunali) e CATI in-bound con un numero verde dedicato. Le unità non rispondenti vengono contattate in una fase successiva con CATI out-bound e CAPI, per recuperare le mancate risposte. Riguardo i contenuti tematici, fra gli altri sono presenti quesiti dettagliati sulla nazionalità, sulla condizione professionale, sul possesso di automobili in famiglia.

Nel questionario è presente una sezione che riguarda gli spostamenti sistematici per studio e lavoro, in cui sono richieste informazioni anche sul dominio territoriale degli spostamenti (con dettaglio a livello comunale o generico estero), in particolare se gli spostamenti avvengano nello stesso comune ovvero altrove, e se il riferimento di partenza e rientro sia l’abitazione nel comune di dimora abituale ovvero altro alloggio (e l’ubicazione di questo a livello comunale); sono richieste inoltre informazioni sui tempi di percorrenza, e sui mezzi (fino a tre indicazioni) utilizzati, in ordine di importanza rispetto alla distanza percorsa per modo. La classificazione dei modi di trasporto è differente da quella proposta nelle LG Eurostat (e mancano alcune voci, per esempio l’aereo) e la modalità ‘a piedi’ è considerata solo come esclusiva, non essendo ricompresa nell’elenco.

I dati sulla mobilità per studio e lavoro acquisiti con il CP sono utilizzati per costruire le matrici di pendolarismo, con disaggregazione delle O/D a livello comunale (attualmente disponibile con aggiornamento all’anno 2011).

Nella matrice di pendolarismo, la variabile Numero di individui è stimata su domini dati dall’intersezione delle seguenti variabili:

- Tipo record S: Provincia di residenza, Comune di residenza, Sesso, Motivo dello

spostamento, Luogo di studio o di lavoro, Provincia abituale di studio o di lavoro, Comune abituale di studio o di lavoro, Stato estero di studio o di lavoro.

- Tipo record L: Provincia di residenza, Comune di residenza, Sesso, Motivo dello spostamento, Luogo di studio o di lavoro, Provincia abituale di studio o di lavoro, Comune abituale di studio o di lavoro, Stato estero di studio o di lavoro, Mezzo, Orario di uscita, Tempo impiegato.

Gli indicatori che riguardano la mobilità, pubblicati con cadenza annuale, sono riferiti ai domini:

- Sesso;
- Motivazione (studio/lavoro);
- Area di pendolarismo (stesso comune di dimora abituale/fuori dal comune di dimora abituale);
- Territorio (tutti i comuni/province/regioni).

Tali indicatori rappresentano la consistenza della popolazione residente che si sposta giornalmente.

Il vantaggio di utilizzare questa fonte è soprattutto dovuto al fatto che si tratta di una rilevazione con un impianto solido e un campione molto consistente, e può senza dubbio essere considerata un punto di riferimento per confrontare gli indicatori di mobilità. Ai fini del fabbisogno conoscitivo, il gap più importante è relativo agli spostamenti per motivi diversi da studio e lavoro.

d. AVQ – Indagine Multiscopo Aspetti della Vita Quotidiana (Istat).

Dalla Nota metodologica:

“L’indagine “Aspetti della vita quotidiana” fa parte del sistema integrato di Indagini Multiscopo sulle famiglie avviato nel 1993 che ha l’obiettivo di produrre informazioni su individui e famiglie. Le informazioni statistiche raccolte, integrate con quelle desumibili da fonte amministrativa e dalle imprese, contribuiscono a determinare la base informativa del quadro sociale del Paese. Attraverso la rilevazione sono indagate diverse aree tematiche,

esplorate da un punto di vista individuale e familiare. I contenuti informativi possono essere raggruppati in quattro grandi aree: famiglia, abitazione e zona in cui si vive; condizioni di salute e stili di vita; cultura, socialità ed attività del tempo libero e interazione tra i cittadini e servizi”.

L’indagine è campionaria, il registro di riferimento è LAC/ANPR. La popolazione di riferimento comprende tutti gli individui (attraverso l’unità campionaria della famiglia), senza limitazioni di età.

L’indagine ha cadenza annuale ed è condotta tra il primo e il secondo trimestre dell’anno di riferimento.

La tecnica è mista CAWI/CAPI-PAPI, dove la compilazione con intervistatore è limitata ai non rispondenti, contattati in un momento successivo alla prima fase della raccolta e alle famiglie sprovviste di accesso a internet. Per la CAWI si accede al Portale Istat dedicato, con credenziali personali.

Il modello di rilevazione è composto da una scheda generale in cui sono raccolte tutte le informazioni sociodemografiche della famiglia e dei suoi componenti, e da sezioni specifiche (Questionario familiare e Questionario individuale) per tema.

Tra gli aspetti della vita quotidiana delle famiglie, sono rilevate in un’apposita sezione le informazioni relative agli spostamenti abituali per studio e lavoro. I quesiti proposti in questa sezione, relativamente agli spostamenti, sono in parte simili a quelli proposti nel modello del Censimento Permanente. In particolare, la classificazione dei modi di trasporto coincide, ma in AVQ c’è la possibilità di indicare più di tre mezzi e, successivamente, quello prevalente con riferimento alla distanza percorsa. Comprende inoltre una serie di quesiti sul grado di utilizzo e soddisfazione dei mezzi pubblici, dei servizi di car e bike sharing e dei veicoli privati. Gli indicatori prodotti si riferiscono agli individui che hanno effettuato spostamenti, e sono di due tipi:

- Numero di individui;

- Percentuale sul totale delle persone con le stesse caratteristiche.
- I domini di output sono dati dall'intersezione delle variabili:
- Motivazione (Studio/lavoro);
- Sesso;
- Classe di età;
- Modo di trasporto;
- Tempo impiegato;
- Ripartizione geografica;
- Regione;
- Tipo di comune (ampiezza demografica);
- Centro/periferia area metropolitana.

e. Spese delle famiglie – focus Viaggi e vacanze (Istat)

A completamento del quadro informativo sulle fonti disponibili, si vuole esaminare il fenomeno della mobilità su distanze medio-lunghe, la cui natura differisce con quello della mobilità locale sia negli indicatori di interesse che nelle tecniche di indagine appropriate per osservarlo e rappresentarlo. Le indagini in questo ambito sono generalmente di tipo retrospettivo. Con riferimento a quest'area della mobilità, generalmente relativa ai viaggi effettuati per lavoro o motivi personali è stato individuato, come fonte statistica attualmente presente nel panorama della produzione nazionale, il focus 'Viaggi e vacanze' inserito nella rilevazione sulle spese delle famiglie.

L'indagine rileva le componenti del turismo nazionale, dato dalle componenti 'domestic' e 'outbound', ossia i viaggi effettuati dai residenti in Italia, sia sul territorio nazionale che all'estero.

Il campo di osservazione comprende i viaggi con pernottamento brevi o lunghi (più di tre notti) e le escursioni (viaggi con andata e ritorno in giornata e permanenza di almeno tre ore senza pernottamento), effettuati per motivi personali e lavoro. Le escursioni verso l'estero sono rilevate con cadenza annuale, quelle domestiche con cadenza triennale. Le informazioni sui viaggi con pernottamento sono raccolte con cadenza annuale o triennale, a seconda delle variabili. Pertanto, il quadro

completo del fenomeno si ha con cadenza triennale.

Ai fini di questa ricognizione tale indagine non è essenziale, in quanto le informazioni raccolte sono in gran parte relative agli aspetti descrittivi del soggiorno o di altri fattori legati al fenomeno turistico e alle spese effettuate; riguardano invece solo in maniera marginale l'aspetto del viaggio inteso come trasporto a destinazione e ritorno, che è pertinente allo studio della mobilità.

Il questionario, nella sezione dedicata ai viaggi e alle escursioni, raccoglie per ogni viaggio informazioni sulla destinazione (comune o Stato estero) e sulla motivazione (lavoro/vacanza/altri motivi). È richiesta anche l'indicazione relativa ad eventuali viaggi abituali. Rispetto alle variabili necessarie per il calcolo degli indicatori stabiliti nelle Linee Guida Eurostat, nel focus non vengono rilevati:

- La distanza percorsa fino alla destinazione (meta iniziale del viaggio) e/o dall'ultima località al rientro a casa;
- La durata dello spostamento per/dalla destinazione principale;
- Il mezzo di trasporto utilizzato per lo spostamento iniziale (e/o il rientro). Il focus richiede di specificare il mezzo di trasporto prevalentemente utilizzato, ma con riferimento all'evento viaggio, che potrebbe non coincidere con quello utilizzato per l'evento trasporto (es. trasporto iniziale in aereo, uso prevalente di auto a noleggio durante il viaggio);
- Eventuale presenza di altri occupanti il veicolo privato durante lo spostamento per/dalla destinazione principale.

La rilevazione esaminata comporta un carico statistico elevato sul rispondente, anche in considerazione del fatto che fa riferimento ad entità di analisi diverse (spese della famiglia – viaggi dell'individuo) e volendo ampliarla per raccogliere anche le informazioni sulla mobilità sarebbe necessario inserire una terza entità, ossia l'evento trasporto.

È stato realizzato uno studio sperimentale

nell'ambito della road map per le Trusted smart statistics in Istat per l'uso di dati telefonici per la stima degli indicatori che riguardano la rilevazione "Viaggi e vacanze".

f. Altre fonti da enti territoriali locali.

A livello territoriale disaggregato (Regioni, Aree metropolitane, Comuni) sono disponibili numerose esperienze di ricerca e produzione di indicatori sulla mobilità, realizzate con carattere episodico o a cadenza irregolare.

Fra queste si può menzionare, come esperienza a carattere prototipale, la smart survey realizzata nell'ambito del progetto HARMONY finanziato con i fondi europei H2020, che ha visto come caso di studio un'applicazione nell'Area Funzionale Urbana di Torino.

La survey si basa su un campione per quote di 584 individui, stratificato su una zonizzazione della FUA torinese in otto aree (55% in quattro aree della municipalità, 33% in tre aree dell'area metropolitana, il 12% negli altri comuni della FUA), e poi secondo le variabili:

- Sesso (equiripartizione M-F)
- Età (40% classe 18-34, 60% classe 35-64)
- Condizione occupazionale (60% occupati, 10% pensionati, 30% studenti)
- Numero di automobili possedute in famiglia (25% nessuna automobile, 50% un'automobile, 25% due o più automobili)

Il campione è stato suddiviso su due periodi di osservazione, la prima e la seconda metà di febbraio 2022.

Per la raccolta è stata utilizzata un'App (MobyApp) sviluppata in ambito europeo e testata anche in altri casi di studio.

La survey è stata organizzata in tre fasi:

- i. Raccolta di variabili preliminari e di contesto sull'individuo e le sue abitudini di mobilità;
- ii. Raccolta delle informazioni sugli spostamenti individuali, qualificati con le variabili: origine, destinazione, durata, modo di trasporto, motivazione, tramite la MobyApp. Tale fase prevede interventi di validazione ed integrazione, da parte del rispondente, dei dati acquisiti in autonomia dall'App;
- iii. Questionario focus sulle scelte individuali, basato sulle informazioni raccolte nella prima fase.

È stato richiesto al rispondente di verificare e validare i dati degli spostamenti per almeno quattro giorni nel periodo di osservazione. L'acquisizione e la conservazione dei dati individuali è effettuata in forma anonimizzata (ID assegnata ad ogni individuo in modo da preservarne l'identità separatamente dai dati raccolti).

Complessivamente l'esperienza ha avuto un



riscontro positivo, soprattutto in termini di gradimento della tecnica utilizzata (acquisizione con App) da parte dei rispondenti, e potrebbe costituire un punto di partenza per ulteriori applicazioni.

Un bacino informativo che sarebbe opportuno monitorare è quello che si potrebbe generare a partire dall'attività dei Mobility Manager. Con il Decreto Interministeriale n. 179 del 12/5/2021 recante "Modalità attuative delle disposizioni relative alla figura del mobility manager" e successive modifiche, i ministeri della Transizione ecologica e delle Infrastrutture e Trasporti hanno disciplinato le attività dei Mobility Manager Aziendale, Scolastico e di Area, che hanno il compito principale di definire i Piani di Spostamento Casa Lavoro (PSCL). Il decreto, all'art 6 c.3c, specifica fra le attività del MM di Area:

"l'acquisizione dei dati relativi all'origine/destinazione ed agli orari di ingresso e uscita dei dipendenti e degli studenti forniti dai mobility manager aziendali e scolastici e trasferimento dei dati in argomento agli enti programmatori

dei servizi pubblici di trasporto comunali e regionali."

Il Decreto prevede inoltre la definizione di Linee Guida per la redazione e l'implementazione dei PSCL, in cui sono indicate le informazioni minime riguardanti gli spostamenti casa-studio/lavoro che il MM deve raccogliere per definire il piano, quali ad esempio:

- Modalità abituale di spostamento;
- Utilizzo di un solo mezzo/modo di trasporto, da specificare;
- Combinazione di più mezzi/modalità di trasporto, da specificare;
- Distanza percorsa;
- Tempo impiegato;
- Disponibilità di mezzi di trasporto, da specificare;
- Disponibilità di abbonamenti ai servizi di trasporto, da specificare.

Il patrimonio informativo alla base della definizione dei PSCL potrebbe essere utilizzato per rappresentare la mobilità abituale a livello urbano e intercomunale.

4.1.2 | FONTI SU OFFERTA E DOMANDA DI MOBILITÀ

Un'area di ricerca e produzione ulteriore, che arricchisce il quadro delle prospettive da cui si osserva il fenomeno della mobilità, è quella della rappresentazione dell'offerta di infrastrutture e servizi per la mobilità e della domanda di tali servizi da parte dell'utenza. Fra i lavori in essere, sono di particolare interesse:

a. Rilevazione Dati ambientali nelle città - Indicatori sulla mobilità urbana (Istat)

La rilevazione ha cadenza annuale e raccoglie informazioni ambientali relative ai 109 comuni capoluogo di provincia/città metropolitana e dal 2020 si è aggiunto su base volontaria il Comune di Cesena. I dati e l'informazione statistica hanno l'obiettivo di fornire un quadro informativo a supporto del monitoraggio dello stato dell'ambiente

urbano e delle attività poste in essere dalle amministrazioni, per assicurare la buona qualità dell'ambiente nelle città. La rilevazione si articola in sette questionari d'indagine - Aria, Eco management, Energia, Mobilità, Rifiuti, Rumore e Verde urbano. La sezione Mobilità raccoglie informazioni su domanda e offerta di trasporto pubblico locale, mobilità sostenibile e infomobilità. Il riferimento per la raccolta dei dati, veicolati a Istat dalla rete dei Comuni Capoluogo, sono i PUT/PUMS. È presente una sezione dedicata alla mobilità condivisa (*sharing mobility*).

Il dettaglio territoriale degli indicatori prodotti è quello dei Capoluoghi di Provincia.

Per l'**offerta** gli indicatori sono:

- I. Densità
- II. Lunghezza della rete

III. Numero di fermate

Distinti per:

- reti tranviarie e della metropolitana;
- corsie preferenziali destinate al TPL;
- fermate di autobus e filobus;
- fermate del tram e stazioni della metropolitana;
- fermate o stazioni della funicolare, della funivia e dei trasporti per vie d'acqua.

IV. Disponibilità

V. Utilizzo (numero)

VI. Posti-km offerti

Distinti per:

- Autobus e filobus;
- Autobus a trazione elettrica e alimentati a metano o GPL;
- Vetture del tram e convogli della metropolitana;
- Vetture della funicolare, cabine della funivia e imbarcazioni dei trasporti per vie d'acqua.

VII. Velocità commerciale dei servizi di trasporto pubblico locale

VIII. Licenze di taxi attive

Per la **domanda**, gli indicatori sono:

- I. Passeggeri annui per abitante
- II. Passeggeri annui del trasporto pubblico locale (valori assoluti)

Per la **sharing mobility** e la **mobilità sostenibile**, gli indicatori pubblicati sono:

- I. Presenza di servizi di car sharing
- II. Disponibilità di veicoli dei servizi di car sharing
- III. Veicoli utilizzati per i servizi di car sharing

Distinti per:

- Car sharing;
 - Bike sharing;
 - Scooter sharing.
- IV. Veicoli a basse emissioni utilizzati per i servizi di car sharing (valori assoluti e per 100 veicoli)
 - V. Densità e lunghezza delle piste ciclabili
 - VI. Superficie aree pedonali
 - VII. Presenza e variazione di ZTL e Zone 30

VIII. Presenza di parcheggi di scambio

IX. Numero stalli di sosta nei parcheggi di scambio.

b. **Osservatorio Nazionale Sharing Mobility**

L'Osservatorio è un network composto da una molteplicità di soggetti pubblici e privati coinvolti nello sviluppo, nella diffusione e nell'offerta dei servizi di *sharing mobility* in Italia. Comprende operatori di servizi di mobilità condivisa, Enti locali, istituti di ricerca e associazioni. È promosso dal Ministero della Transizione Ecologica, dal Ministero delle Infrastrutture e dei Trasporti e dalla Fondazione per lo Sviluppo Sostenibile.

Produce, con cadenza annuale dal 2016, il dettagliato **Rapporto sharing mobility** sull'offerta e la domanda di servizi di *vehicle sharing* e *ride sharing* (carpooling aziendale). Il rapporto offre una rappresentazione del fenomeno a livello aggregato per il territorio italiano, con due focus su Roma e Milano. Comprende inoltre sezioni di approfondimento anche sui comportamenti e gli stili di mobilità degli utilizzatori dei servizi considerati (cluster per mix dei mezzi di trasporto utilizzati, sia propri che condivisi).

Gli indicatori pubblicati, con riferimento temporale annuale, sono per l'**offerta**:

- I. Numero di veicoli in flotta
- II. Qualità ecologica della flotta (tipologia di veicolo e alimentazione ove presente)
- III. Presenza e assenza dei servizi nei capoluoghi di provincia

Gli indicatori per la **domanda** sono:

- I. Numero di iscrizioni al servizio
- II. Numero di servizi attivi
- III. Numero di noleggi effettuati (per Capoluogo di provincia)
- IV. Percorrenze (km) (per Capoluogo di provincia)
- V. Noleggi mensili e Noleggi/giorno (solo su Roma e Milano)

La tipologia dei servizi è distinta per:

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

- Car sharing free floating;
- Car sharing station based;
- Bike sharing free floating;
- Bike sharing station based;
- Scooter sharing;
- Monopattino sharing;
- Ride sharing.

Le modalità Bike, Scooter e Monopattino sono ulteriormente classificate in un insieme unico denominato Micromobilità, messo a confronto con il Car sharing, per alcuni indicatori.

Nell'ambito delle attività dell'Osservatorio, è stata condotta una rilevazione pilota sulla mobilità urbana (**Progetto Pollicino**) sul territorio del Comune di Bologna. Questa indagine ha lo scopo di raccogliere informazioni su tutti gli spostamenti effettuati dagli individui del campione, sia con mezzi propri che con servizi di *sharing mobility*.

Dal sito dell'Osservatorio:

"Il progetto ha avuto il suo primo ambito applicativo nella città di Bologna dove sono stati 2.400 i cittadini che hanno scaricato l'app loPollicino sul proprio telefono e hanno acconsentito alla geolocalizzazione del proprio smartphone e circa un migliaio quelli che sono stati coinvolti nell'indagine e che, per almeno 7 giorni, hanno inserito alcune informazioni essenziali come il mezzo di trasporto utilizzato e la motivazione del proprio spostamento. Di questi, seicento Pollicini, stratificati per età e genere, hanno formato un campione rappresentativo della mobilità in città e una banca dati eccezionale per analizzare i comportamenti, le abitudini e gli stili di mobilità delle persone. La raccolta dei dati è avvenuta in forma volontaria, completamente anonima, senza nessuna forma di profilazione o di scoring, vale a dire quei metodi che attribuiscono un punteggio positivo o negativo a determinati comportamenti dei cittadini. L'obiettivo dell'analisi, infatti, è stato quello di far avanzare la ricerca sui comportamenti di mobilità urbana rispetto alle tradizionali indagini campionarie compiute

con interviste telefoniche o via web e che, di norma, raccolgono dati esclusivamente riferiti a un solo giorno lavorativo della settimana, non consentendo un'analisi della mobilità degli individui di ogni giorno della settimana, dal momento in cui si esce sino a che non si fa rientro a casa."

È una *smart survey*, in cui la raccolta dei dati, opportunamente trattati per essere anonimizzati, è effettuata tramite un'App per dispositivo mobile (loPollicino). L'App raccoglie le posizioni del dispositivo a intervalli regolari tramite GPS (i cosiddetti *bread crumbs*).

L'utente deve completare il quadro della mobilità realizzata, integrando l'informazione relativa ad altri aspetti, come la motivazione dei singoli spostamenti. Può inoltre correggere l'informazione acquisita col GPS se riscontra inesattezze sulla ricostruzione degli spostamenti (sulla struttura degli stessi, sul mezzo utilizzato etc). La tecnica di raccolta App-based consente di registrare tutti gli spostamenti, compresi quelli di percorrenza inferiori al chilometro, rilevanti in ambito urbano, che in genere vengono esclusi dalle survey tradizionali.

Anche tale iniziativa, come quella illustrata nella sezione 4.1.1 lettera f, nell'ambito del progetto HARMONY, potrebbe costituire un'esperienza a carattere prototipale per una *smart survey* su scala più estesa.

4.1.3 | I FLUSSI PER MODO DI TRASPORTO

Un altro gruppo comprende le rilevazioni ufficiali consolidate e soggette a Regolamento europeo che organizzano l'informazione relativa al fenomeno del trasporto con un'ottica "a silos", ossia distinguendo l'acquisizione dei dati e la produzione di indicatori secondo il modo di trasporto.

Esaminando gli aspetti metodologici più rilevanti, e limitando il riferimento alla mobilità delle persone (pertanto qui non riportiamo gli elementi relativi al trasporto di merci), in sintesi possiamo così rappresentarli:

a. **Trasporto marittimo (Istat)**

La rilevazione sul trasporto marittimo fornisce statistiche relative al trasporto di merci e di passeggeri effettuato a fini commerciali. La rilevazione ha carattere censuario, riferendosi all'insieme degli arrivi e delle partenze registrati nei porti italiani. I dati relativi ai passeggeri sono visualizzati per i porti che hanno movimentato almeno 200 mila passeggeri.

I set di indicatori prodotti da Istat ed Eurostat sono:

- Passeggeri trasportati (esclusi crocieristi), per tipo flusso (internazionale, nazionale, totale) e direzione (in ingresso, in uscita, totale), annuale;
- Passeggeri Imbarcati/Sbarcati, per i primi 20 porti europei, annuale;
- Passeggeri in Arrivo (sbarcati) e in Partenza (imbarcati) in tutti i porti italiani per direzione (Arrivo/partenza), annuale;
- Imbarcati/Sbarcati/Totale per NUTS2;
- Trasportati per totale porti principali (esclusi crocieristi), per tipo di flusso (nazionale, internazionale) e direzione (in arrivo, in partenza, totale), trimestrale;
- Trasportati per porto (italiano), bandiera, direzione (in ingresso, in uscita, totale), trimestrale.

La rilevazione non prevede l'elaborazione di una matrice O/D.

Eurostat stima alcuni indicatori per i passeggeri-km (Pkm), pubblicati nella tabella del

Database "Trasporto marittimo di passeggeri nelle Zone Economiche Esclusive (ZEE o EEZ) dei Paesi europei (nazionale, internazionale intra UE27, internazionale extra UE27, in transito, totale) per singolo Paese, annuale".

È in corso uno studio sperimentale nell'ambito della road map per le Trusted Smart Statistics in Istat, per l'uso dei dati AIS per incrementare la qualità dei dati della rilevazione sul trasporto marittimo.

b. **Trasporto aereo (Istat)**

La nuova rilevazione statistica sul trasporto aereo è a carattere totale e raccoglie informazioni sul trasporto di passeggeri, merci e posta, sulle tappe di volo, sui posti disponibili e sui movimenti degli aeromobili. Le unità di rilevazione sono costituite dagli aeroporti e i dati vengono forniti dalle società di gestione degli stessi. I dati trimestrali si riferiscono agli aeroporti che inviano i dati su base mensile. Gli indicatori prodotti da Istat ed Eurostat (che utilizza dati Istat ed Enac) sono i seguenti:

- Passeggeri totali (mensili, trimestrali, annuali);
- Passeggeri trasportati, per rotta (mensili, trimestrali);
- Passeggeri NUTS2 (annuali);
- Passeggeri-km Totali (annuale);
- Passeggeri-km nazionale + UE (annuale);
- Passeggeri-km Extra UE (annuale);
- Passeggeri-km dei Sorvoli (annuale).

È disponibile la matrice O/D dei passeggeri per coppie di Paesi (passeggeri a bordo e passeggeri trasportati).

c. **Trasporto ferroviario (Istat)**

La rilevazione sul trasporto ferroviario è finalizzata a raccogliere l'informazione statistica sul servizio di trasporto fornito dalle imprese ferroviarie operanti sul territorio nazionale, in ottemperanza a quanto previsto nel Regolamento del Parlamento europeo e del Consiglio n. 643/2018 (che ha rifiuto tutti i precedenti atti legali relativi alle statistiche dei trasporti ferroviari). La rilevazione è censuaria

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

e l'unità di rilevazione è l'impresa ferroviaria attiva, vale a dire qualsiasi impresa di diritto pubblico o privato che presta servizio di trasporto merci e/o passeggeri a mezzo ferrovia. Attraverso tale rilevazione vengono raccolti i dati, variamente dettagliati, sul trasporto di merci e di passeggeri, nonché sugli incidenti occorsi sulla rete ferroviaria nazionale e locale e sulle caratteristiche tecniche dell'infrastruttura della rete nazionale.

I set di indicatori pubblicati da Istat ed Eurostat sono:

- Passeggeri trasportati totali, annuale e trimestrale;
- Passeggeri trasportati per tipo flusso e direzione (nazionale, internazionale in ingresso e in uscita), annuale;
- Trasporto internazionale di passeggeri da confine a Paese di sbarco, annuale;
- Trasporto internazionale di passeggeri, da Paese di imbarco a confine, annuale;
- Passeggeri trasportati per velocità del treno (fino al 2011), annuale;

- Percorso medio in km per passeggero (interno/internazionale), annuale;
- Passeggeri-km totali, annuale e trimestrale;
- Passeggeri-km per tipo flusso e direzione (interno/internazionale in ingresso e in uscita), annuale.

È inoltre disponibile la matrice O/D dei passeggeri trasportati, con dettaglio territoriale NUTS2 e a cadenza quinquennale.

d. **Trasporto passeggeri su strada (Eurostat)**

Eurostat pubblica regolarmente una tabella in cui sono riportate le stime annuali a livello Paese dei seguenti indicatori:

- Passeggeri e Passeggeri-km trasportati con autobus e pullman immatricolati in Italia, distinti in Trasporto nazionale/nazionale urbano/nazionale interurbano.

Non è disponibile per l'Italia il dato relativo al trasporto internazionale.



4.1.4 | IL TRAFFICO VEICOLARE

I progetti per la stima del traffico veicolare hanno avuto un forte sviluppo negli anni più recenti, sia a livello internazionale, dove Eurostat/ITF/UNECE hanno predisposto una

tabella dettagliata per l’acquisizione degli indicatori sui veicoli-km per le diverse categorie di veicoli stradali (Odometer reading questionnaire):

Vehicle categories by fuel type	Age of the vehicle								Total
	0-2	2-4	4-6	6-8	8-10	10-15	15-20	Over 20	
M1: Passenger cars - Total.									
• Petrol									
• Diesel									
• Pure electric									
• Petrol-Hybrid									
• Diesel-Hybrid									
• Other fuel - Total									
of which :									
- Bi-fuel petrol/LPG									
- Bi-fuel petrol/CNG									
- LPG									
- CNG									
- Flex fuel									
- Other									
M2: Mini buses - Total.									
• Petrol									
• Diesel									
• Pure electric									
• Other fuel									
M3: Buses and coaches - Total.									
• Diesel									
• Diesel-Hybrid									
• Pure electric									
• CNG									
• Other fuel									
N1: Goods vehicles up to 3.5t - Total.									
• Petrol									
• Diesel									
• Other fuel - Total									
of which :									
- Bi-fuel Petrol/LPG									
- Bi-fuel Petrol/CNG									
- Pure electric									
N2: Goods vehicles between 3.5t and 12t - Total.									
• Diesel									
• Other fuel									
N3: Goods vehicles over 12t - Total.									
• Diesel									
• Other fuel									
L1/L2/L6: Mopeds and light 3- and 4-wheelers - Total.									
• Petrol									
• Pure electric									
• Other fuel									
L3/L4/L5/L7: Motorcycles (with or without sidecar), tri-cycles and quads - Total.									
• Petrol									
• Pure electric									
• Other fuel									
T5: Tractor on wheels - Total.									
• Diesel									
• Other fuel									

Figura 5 – Tabella predisposta da Eurostat/ITF/UNECE, Odometer reading questionnaire

sia a livello nazionale, con raccolte di dati consolidate, recenti e oggetto di studio.

Si riportano di seguito le fonti più rilevanti.

a. Dati AISCAT sul traffico veicolare autostradale.

Gli indicatori con cadenza mensile sono i veicoli-km distinti per:

- Società autostradale (sia associate che esterne ad AISCAT);
- Tipo veicolo (leggeri "i motocicli e gli autoveicoli a due assi con altezza da terra, in corrispondenza dell'asse anteriore, inferiore a 1,30 m.", pesanti "gli autoveicoli a due assi con altezza da terra, in corrispondenza dell'asse anteriore, superiore a 1,30 m., sia tutti gli autoveicoli a tre o più assi.").

Gli indicatori con cadenza trimestrale sono i veicoli-km distinti per:

- Autostrada o traforo;
- Tipo veicolo (leggeri/pesanti).

I dati sono stimati in base agli ingressi in autostrada e alla lunghezza dell'estesa stradale.

b. Dati ANAS sul traffico veicolare sulle strade extraurbane di competenza.

Dal sito ANAS:

"Il sistema automatico di rilevamento statico del traffico, costituito da circa 1.200 sezioni di conteggio, è distribuito sull'intera rete Anas: tutti i sensori inviano i propri dati ad un sistema di monitoraggio centralizzato denominato PANAMA (Piattaforma Anas per il Monitoraggio e l'Analisi), che provvede alla verifica ed elaborazione dei trend dell'Indice di Mobilità Rilevata. L'affidabilità dei dati acquisiti è assicurata da una serie di processi di controllo; in particolare due step automatici di controllo garantiscono la consistenza e

coerenza del database. Il primo processo ha il fine di evidenziare problemi all'interno del file inviato dalla stazione di rilevamento locale. A valle di tali verifiche eventuali errori di coerenza, impossibili da correggere, rendono il file inutilizzabile e quindi rifiutato dal sistema. Una volta che i dati vengono caricati nel database, PANAMA esegue il secondo step, costituito da varie procedure per la valutazione dell'affidabilità dei dati aggregati, non eliminando i dati inseriti, ma classificandoli attraverso un parametro che ne qualifica la coerenza rispetto ad alcune situazioni reali che possono capitare. Tale parametro permette di escludere i dati incerti dai valori che concorrono ai calcoli delle misure desiderate. La veridicità dei dati acquisiti, ovvero la capacità dei sensori di rilevare la realtà effettivamente in transito nella sezione, viene inoltre valutata dal personale Anas con controlli a campione effettuati attraverso un rilievo a vista. Il personale Anas utilizza una tecnologia che permette l'acquisizione di video dei veicoli effettivamente in transito sulla sezione con la sovrapposizione simultanea della stringa di dati rilevati dalla centralina locale. Il filmato viene successivamente visualizzato in ufficio, fase nella quale si rilevano tutti gli errori sia di conteggio sia di classificazione, e quindi si valuta la bontà del dato restituito in automatico rispetto la realtà su strada.

"[La rete dei sensori] Le tecnologie utilizzate sono sostanzialmente due: spire induttive e sensori a microonde, per quanto il sistema riceva attualmente anche i dati di traffico rilevati attraverso il sistema Vergilius per il



controllo elettronico della velocità media. Sono state inoltre implementate ulteriori tecnologie quali pesa dinamiche, telecamere per il rilevamento delle merci pericolose e tracciatori bluetooth per il rilevamento dei tempi di percorrenza.”

I dati sono raccolti per ‘sezioni di conteggio’, a loro volta classificate in base al territorio (Italia, Nord, Centro, Sud, Sicilia, Sardegna), e in base ai tipi di veicoli, distinti in:

- veicoli leggeri che raggruppano i motocicli, le auto con e senza rimorchio e furgoni o camion (la cui sagoma corrisponde alle tipologie con portata inferiore alle 3,5 t.) anch’essi con e senza rimorchio;
- veicoli pesanti che aggregano tutti gli altri veicoli, ovvero i camion “grandi” (la cui sagoma corrisponde a tipologie con portata al di sopra delle 3,5 t), gli autotreni, gli autoarticolati e i pullman.

Gli indicatori prodotti a cadenza mensile sono gli “Indici di mobilità rilevata” (IMR), corrispondenti al numero medio stimato di veicoli rilevati dai sensori per ogni dominio, definito da:

- tipo veicolo (totali/pesanti);
- ambito territoriale (Italia, Nord, Centro, Sud, Sicilia, Sardegna);
- giorno della settimana (totale/feriale/prefestivo/festivo);
- fascia oraria (giorno 6-22h/notte 22-6h).

Viene inoltre pubblicato il “Traffico medio giornaliero annuo” (TGMA) calcolato come media dei passaggi bidirezionali su ogni postazione, per tutte le “giornate valide”,

ossia i giorni dell’anno in cui i passaggi sono rilevati senza malfunzionamenti e se sono caricati almeno il 98% dei passaggi rilevati nei 288 intervalli da 5 minuti in cui è divisa la giornata. Questo indicatore è pubblicato per i domini:

- Postazione (strada, km);
- Comune;
- Provincia;
- Tipo veicolo (leggeri/pesanti).

Infine, ANAS produce dei report a cadenza trimestrale e annuale, disponibili su richiesta, sul traffico rilevato dalle singole stazioni.

Questa informazione è raccolta puntualmente secondo uno schema che sintetizza i passaggi dei veicoli con due indicatori (volumi medi e velocità medie) sui domini di intersezione delle seguenti dimensioni:

- Direzione del flusso (ascendente/discendente);
- Tipo veicolo (leggeri/pesanti);
- Fascia oraria (6-20h/20-22h/22-6h).

È infine possibile ottenere delle elaborazioni specifiche rispetto all’ambito territoriale e temporale, facendone richiesta ad ANAS.

c. **Progetto Istat “Quantificazioni e descrizione del traffico veicolare”**

Il progetto ha l’obiettivo di stimare il traffico, in termini di veicoli-km, realizzato dai veicoli di tutte le categorie immatricolati in Italia, a partire dai dati presenti nell’Archivio Revisioni e dal Parco veicolare, forniti dal Ministero dei Trasporti e dall’ACI. Le stime avranno cadenza annuale. Allo stato attuale non sono ancora disponibili, essendo il progetto in fase di studio con riferimento ad alcuni aspetti metodologici essenziali.



4.2 | FONTI COMPLEMENTARI A COMPLETAMENTO DEL QUADRO INFORMATIVO

4.2.1 | FONTI COMPLEMENTARI PER LA DEFINIZIONE E LO SVILUPPO DI MODELLI DI TRASPORTO

Nell'ambito della produzione statistica nazionale, oltre alle fonti approfondite nella sezione precedente, si ritiene opportuno mettere in evidenza ulteriori rilevazioni ed elaborazioni che costituiscono un interessante supporto per la definizione di modelli di trasporto. Sebbene non verranno qui esaminate nei dettagli, a completamento del quadro generale può essere utile riportarne un elenco, non esaustivo.

- a. Parco veicolare. Titolare ACI (PSN ACI-00002). Obiettivo: fornire un quadro dettagliato in Italia attraverso l'analisi dei dati relativi ai veicoli iscritti al Pubblico Registro Automobilistico.
- b. Il mercato automobilistico: prime iscrizioni, radiazioni e trasferimenti di proprietà dei veicoli. Titolare ACI (PSN ACI-00014). Obiettivo: Completare l'analisi del mercato automobilistico attraverso la descrizione delle principali caratteristiche del mercato dell'usato, delle prime iscrizioni e delle radiazioni.
- c. Immatricolazioni e passaggi di proprietà di autovetture. Titolare MIT (PSN MIT-00010). Obiettivo: Consolidamento e miglioramento della produzione di informazioni statistiche a sostegno della conoscenza e delle decisioni dei policy makers, operatori economici e dei cittadini.
- d. Patenti in corso di validità e neopatentati. Titolare MIT (PSN MIT-00011). Obiettivo: Consolidamento e miglioramento della produzione di informazioni statistiche a sostegno della conoscenza e delle decisioni dei policy makers, operatori economici e dei cittadini.
- e. Trasporto pubblico locale. Titolare MIT (PSN MIT-00018). Obiettivo: diffusione di statistiche relative al trasporto pubblico locale.
- f. Trasporti per vie d'acqua interne. Titolare MIT (PSN MIT-00021). Obiettivo: diffusione di statistiche sul trasporto per vie di acque interne.
- g. Collegamenti marittimi con le Isole. Titolare MIT (PSN MIT-00024). Obiettivo: diffusione di statistiche relative ai collegamenti marittimi con le isole.
- h. Indagine sulle spese delle famiglie. Titolare ISTAT (PSN IST-02396). Obiettivo: L'indagine sulle spese familiari rileva le spese sostenute dalle famiglie residenti in Italia per acquistare beni e servizi destinati al consumo familiare e rappresenta la fonte informativa per descrivere, analizzare e interpretare i comportamenti di spesa. La sezione delle spese per trasporti è disaggregata (con ulteriori sottoclassi) in: spese per acquisto di mezzi di trasporto, per acquisto di servizi di trasporto, per l'esercizio di mezzi di trasporto.

4.2.2 | FONTI COMPLEMENTARI PER IL MONITORAGGIO E LA VALUTAZIONE DELLE POLICY E DEGLI OBIETTIVI

Come fin qui ampiamente delineato, la mobilità si colloca in una posizione strategica nel tessuto socioeconomico di ogni Paese, essendo sia causa che conseguenza di un funzionamento fluido e resiliente degli ingranaggi che legano i sistemi in cui operano imprese, famiglie e istituzioni.

Pertanto, la misurazione del fenomeno con diverse prospettive e tipologie di indicatori, si incardina in numerose raccolte statistiche che utilizzano informazione proveniente da fonti ufficiali già esistenti.

Si ritiene utile in questa sede citarne alcune di particolare rilievo, in modo da completare il quadro dei fabbisogni e del potenziale informativo che potrebbe essere coadiuvato dall'uso dei Big Data.

a. **Sustainable Development Goals (SDGs).**

Dal sito Istat:

“Il 25 settembre 2015, l’Assemblea Generale delle Nazioni Unite ha adottato l’Agenda 2030 per lo sviluppo sostenibile nella quale si delineano le direttrici delle attività per i successivi 15 anni. I 17 Sustainable Development Goals che compongono l’Agenda 2030 rappresentano il piano di azione globale per radicare la povertà, proteggere il pianeta e garantire la prosperità per tutti.

I Sustainable Development Goals fanno riferimento a diversi domini dello sviluppo relativi a tematiche di ordine ambientale, sociale, economico ed istituzionale, delineando un piano d’azione globale per i prossimi 15 anni. La strada da percorrere a livello internazionale è definita dal Cape Town Global Action Plan, la strategia per mettere in atto tutte quelle azioni necessarie alla modernizzazione e al rafforzamento dei sistemi statistici a livello nazionale e globale.

Per identificare un quadro di informazione statistico condiviso quale strumento di monitoraggio e valutazione dei progressi verso gli

obiettivi dell’Agenda, la Commissione Statistica delle Nazioni Unite ha costituito l’Inter Agency Expert Group on SDG Indicators che ha definito un insieme di oltre 200 indicatori.” Gli indicatori definiti dall’Expert Group sono integrati con altri indicatori di contesto nazionale (ossia specifici per ogni singolo Paese che li adotta). Per la tematica Trasporti e mobilità, le sezioni interessate sono la 3 “Salute e benessere” (con riferimento all’incidentalità stradale), la 9 “Imprese, innovazione, infrastrutture”, la 11 “Città e comunità sostenibili” e la 13 “Lotta contro il cambiamento climatico”. Istat pubblica con cadenza annuale il **Rapporto SDGs**, è inoltre disponibile una **dashboard** aggiornata con dati e grafici, dove consultare gli indicatori specifici.

b. **Rapporto sul benessere equo e sostenibile (BES)**

Dal sito Istat:

“Il progetto per misurare il benessere equo e sostenibile nasce con l’obiettivo di valutare il progresso di una società non soltanto dal punto di vista economico, ma anche sociale e ambientale.

L’Istat, insieme ai rappresentanti delle parti sociali e della società civile, ha sviluppato un approccio multidimensionale per misurare il “Benessere equo e sostenibile” (Bes) con l’obiettivo di integrare le informazioni fornite dagli indicatori sulle attività economiche con le fondamentali dimensioni del benessere, corredate da misure relative alle disuguaglianze e alla sostenibilità. Sono stati individuati 12 domini fondamentali per la misura del benessere in Italia.

L’analisi dettagliata degli indicatori, pubblicata annualmente nel rapporto Bes a partire dal 2013, mira a rendere il Paese maggiormente consapevole dei propri punti di forza e delle difficoltà da superare per migliorare la qualità della vita dei cittadini, ponendo tale

concetto alla base delle politiche pubbliche e delle scelte individuali.

Nel 2016 il Bes è entrato a far parte del processo di programmazione economica: per un set ridotto di indicatori è previsto un allegato del Documento di economia e finanza che riporti un'analisi dell'andamento recente e una valutazione dell'impatto delle politiche proposte. Inoltre, a febbraio di ciascun anno vengono presentati al Parlamento il monitoraggio degli indicatori e gli esiti della valutazione di impatto delle policy."

Gli indicatori di interesse per il tema mobilità sono inseriti nel dominio "12 – Qualità dei servizi" e derivano da indagini Istat.

Istat pubblica con cadenza annuale un rapporto descrittivo della situazione del Paese nei vari domini tematici, e mette a disposizione disponibile una dashboard aggiornata con tabelle e grafici.

c. Piano Nazionale di Ripresa e Resilienza (PNRR)

Dal sito della Presidenza del Consiglio dei Ministri del Governo italiano:

"Il Piano Nazionale di Ripresa e Resilienza (PNRR) Italia Domani, approvato dalla Commissione europea il 22 aprile 2021, si inserisce all'interno del programma Next Generation EU (NGEU)

[...] Il Piano, che si sviluppa in sei Missioni, ha tre obiettivi principali. Il primo, con un orizzonte temporale ravvicinato, risiede nel riparare i danni economici e sociali causati dalla crisi pandemica.

Con una prospettiva più di medio-lungo termine, il Piano affronta alcune debolezze che affliggono la nostra economia e la nostra società da decenni: i perduranti divari territoriali, le disparità di genere, la debole crescita della produttività e il basso investimento in capitale umano e fisico. Infine, le risorse del Piano contribuiscono a dare impulso a una compiuta transizione ecologica.

[...] Istituita presso la Presidenza del Consiglio dei Ministri, la Cabina di regia è l'organo di indirizzo politico che coordina e dà impulso

all'attuazione degli interventi del PNRR.

[...] Ogni riunione della Cabina di regia consente di fare il punto sullo stato di avanzamento delle riforme e degli investimenti.

Permette inoltre di individuare per tempo ostacoli e criticità, in modo da poter intervenire con prontezza e rispettare il calendario degli impegni che sono stati concordati con la Commissione, che sono determinanti per l'assegnazione dei fondi."

Le missioni che interessano più direttamente l'area di infrastrutture trasporti e mobilità sono la M1 "Digitalizzazione, innovazione, competitività, cultura e turismo", la M2 "Rivoluzione verde e transizione ecologica" e la M3 "Infrastrutture per una mobilità sostenibile", anche nell'ottica della trasformazione digitale (servizi MaaS, monitoraggio del servizio di trasporto pubblico, etc).

d. Raccolte di indicatori a livello internazionale.

Sono state implementate nel tempo numerose raccolte di dati per il monitoraggio e la rappresentazione della mobilità da parte di istituzioni internazionali con differenti compiti istituzionali.

Si cita, ad esempio, il Questionario Comune (Common Questionnaire UNECE/ITF/Eurostat) per la compilazione del quale è richiesta una vasta collezione di indicatori con cadenza annuale, che spazia dalla performance per modo di trasporto, alle statistiche sugli incidenti stradali, alla dotazione infrastrutturale.

Nel PSN è previsto un lavoro specifico (IST-02653) finalizzato alla fornitura degli indicatori per queste raccolte: "Elaborazione di dati nazionali e regionali su trasporti, ambiente e turismo per istituzioni internazionali (Ocde-ift, Eurostat, Unece, Unwto)".

4.3 | FONTI DI RIFERIMENTO PER LA RAPPRESENTAZIONE INFRASTRUTTURALE E DELL'OFFERTA DI SERVIZI

La produzione statistica e lo sviluppo di modelli trasportistici necessitano di riferimenti geospaziali puntuali e il più possibile dettagliati. Come evidenziato nel capitolo 2 "Definizioni", tali riferimenti possono essere di tipo diverso, e differenziarsi nel caso di dati acquisiti con metodi tradizionali e Big Data.

Per creare quindi una rappresentazione delle infrastrutture e dei servizi potenzialmente disponibili su cui si sviluppa la mobilità, abbiamo a disposizione una serie di fonti di cui riportiamo quelle principalmente utilizzate allo stato attuale.

- Conto Nazionale dei Trasporti, fornisce informazioni di carattere generale relative alle infrastrutture di trasporto suddivise per tipologia in differenti capitoli, come ad esempio nel capitolo V "Trasporti su strada" in cui viene riportata l'estensione della rete stradale suddivisa per tipologia di strada (Autostrade, Strade Regionali e Provinciali, Altre Strade di interesse nazionale) a livello nazionale e regionale (appendice al capitolo V).
- Rete ferroviaria (fonte: RFI - Rete Ferroviaria Italiana), mappa con dettaglio per tipo di linea ferroviaria e stazioni;
- Rete stradale/autostradale (fonti: ANAS, AISCAT), mappe con dettagli per tipo di strada, con indicazioni dei caselli autostradali;
- Anagrafica dei porti italiani (fonte: MIT - Ministero delle infrastrutture e dei trasporti), elenco dettagliato e georeferenziato dei 351 principali porti italiani;
- Aeroporti in Italia (fonte: ENAC - Ente Nazionale Aviazione Civile), mappa con dettaglio dei bacini di traffico degli

- aeroporti italiani;
- Dati relativi all'offerta dei servizi di trasporto. Per completare il quadro delle potenzialità offerte dal territorio per soddisfare la domanda di mobilità, si evidenzia il valore aggiunto delle fonti che forniscono i riferimenti spazio-temporali dell'offerta dei servizi, ossia le tabelle orarie del trasporto ferroviario, aeroportuale, portuale e stradale a lunga percorrenza, e i General Transit Feed Specification (GTFS) per il trasporto pubblico locale e ferroviario. Tali fonti, sebbene strategiche per la costruzione di modelli, non sono sempre accessibili o disponibili sull'intero territorio.

I dati delle infrastrutture sopra citate sono disponibili in formato grafico (shapefile per GIS) o corredati da informazioni che ne permettono la georeferenziazione; alcuni sono scaricabili da siti istituzionali come il sito degli Open Data del Ministero delle Infrastrutture e dei Trasporti². Inoltre, nell'ottica delle Direttive Europee INSPIRE 2007/2/CE, attuata dal Decreto Legislativo del 27 gennaio 2010 e finalizzata alla costruzione di un'infrastruttura dati europea per la condivisione dei dati spaziali connessi alle politiche ambientali fra le organizzazioni del settore pubblico delle varie nazioni, i dati delle reti di trasporto dovrebbero essere resi disponibili e condivisi sulla piattaforma INSPIRE³. Per quanto riguarda i GTFS, invece, alcuni sono fruibili liberamente come nel caso di Regione Toscana che mette a disposizione tutti i GTFS del servizio pubblico regionale⁴, o forniti eventualmente dai gestori dei servizi su richiesta per scopi di studio o ricerca.

2. <https://dati.mit.gov.it/catalog/dataset>

3. <https://inspire.ec.europa.eu/>.

4. <https://dati.toscana.it/it/dataset/rf-oraritb>

4.5 | SINTESI DELLE PRINCIPALI INDAGINI STATISTICHE

Indagine	Osservatorio Audimob	Indagine sull'uso del tempo-Multiscopo (TUS)	
Autore	ISFORT	ISTAT	
Tecnica rilievo	CAWI/CATI	PAPI	
Dimensione campione	16 000 individui	circa 34.000 famiglie distribuite in 650 comuni italiani.	
Cadenza	Annuale	Quinquennale	
Omogenea EUROSTAT	Sì	Sì	
Riferimento spaziale geografico	Comune origine-> comune destinazione	Ai fini della rappresentazione del fenomeno della mobilità, tuttavia non rileva due fattori: il riferimento territoriale degli spostamenti e le distanze percorse	
Riferimento temporale	Fascia oraria e durata	Gli indicatori disponibili sono riferiti alla variabile "tempo", quindi: durata media generica e specifica (in hh:mm), tempo dedicato all'attività di spostamento in percentuale sulle 24 ore, incidenza percentuale delle persone che hanno svolto l'attività di spostamento.	
Modo dello spostamento	Combinazione mezzi utilizzati	Mezzi di spostamento (in TUS è considerata una modalità della variabile "luoghi", e disaggregata in a piedi/bicicletta/motocicletta, ciclomotore, scooter/autovettura privata/altro mezzo privato diverso da auto e moto/mezzi pubblici), in combinazione con sesso, età, condizione lavorativa, stato civile, posizione nel nucleo, regione, tipo di comune.	
Frequenza	Sì		
Motivo	Sì	Sì, per alcuni indicatori	
Caratterizzazione socio-economica	Sì	Sì	
Dimensione di espansione temporale	Giorno feriale(festivo) medio		

Tabella 3 – Sintesi delle caratteristiche delle principali indagini statistiche sulla mobilità passeggeri

LE FONTI DATI SULLA MOBILITÀ IN ITALIA

	Censimento permanente della Popolazione e delle abitazioni	Indagine Multiscopo Aspetti della Vita Quotidiana (AVQ)	Spese delle famiglie - Focus Viaggi e Vacanze
	ISTAT	ISTAT	ISTAT
	CAWI/CATI/CAPI	CAWI/CAPI-PAPI	CAPI
	500 000 famiglie	circa 25.000 famiglie distribuite in circa 800 comuni italiani	circa 32 mila famiglie residenti in circa 540 comuni italiani.
	Annuale	Annuale	Annuale e continua su tutti i mesi dell'anno
	Classificazione modi di trasporto differente da Eurostat		
	-Spostamento avviene all'interno dello stesso comune o altrove, riferimento partenza rientro è il luogo abituale di dimora. -Comune Origine --> Comune Destinazione	Regione, ripartizione geografica	Viene rilevato la Regione e la ripartizione geografica di residenza e la regione e provincia di destinazione principale del viaggio o stato estero dell'UE di destinazione principale o macro-area geografica non UE
	Fascia oraria e durata	Tempo impiegato	Non si rileva la durata dello spostamento da e per la destinazione principale
	Fino a 3 mezzi in ordine di importanza rispetto alla distanza percorsa per modo	Possibilità di indicare più di 3 mezzi e successivamente quello principale rispetto alla distanza percorsa	Mezzo prevalente utilizzato nell'ambito del viaggio, che potrebbe non coincidere con quello utilizzato nello spostamento iniziale/rientro
	Sì		Viene richiesto l'indicazione relativa ad eventuali viaggi abituali
	Lavoro/studio	Lavoro/studio	Lavoro/ Vacanza/ altri motivi
	Sì	Sesso/classe di età	Sì
	Giorno feriale medio		

5 | ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

Lorenzo Vannacci²[0000-0001-9587-7611], Martina Farsi²[0000-0002-3132-8071], Giovanna Astori¹

1. ISTAT, Roma, Italia

2. FS Research Centre, Firenze, Italia

5.1 | ANALISI DELLE LIMITAZIONI DELLE ATTUALI FONTI DATI DELLA MOBILITÀ DELLE PERSONE

"Good decision making requires good data and good models, especially for transport planning. This, in turn, requires good data to establish the model base year, a sound representation of travel choice behaviour by users in its formulations and a good estimation of how the future will evolve. Data collection is therefore essential, but it is often perceived by modellers as an endless source of frustration."
[5.1]

Il modellista dei trasporti si confronta spesso con fotografie del fenomeno della mobilità a cui manca qualche porzione; l'attività principale, insieme alla costruzione dei modelli, è costituita dall'interpretazione congiunta di dati di diversa natura per la ricomposizione del quadro nel suo complesso. Le risorse temporali ed economiche sono generalmente limitate; le interviste di qualsiasi tipo, telefoniche, web o al cordone (effettuate cioè intercettando l'utenza del sistema di trasporto in punti nevralgici della rete, per esempio nelle strade principali di accesso all'area di studio), richiedono tempo per la progettazione e per l'esecuzione, oltre ad essere economicamente onerose. Fonti indispensabili sono i dati provenienti dal censimento ISTAT, che permettono di arrivare alla disaggregazione territoriale delle sezioni di censimento ma sono attualmente disponibili fino al 2011. In questo caso il tassello mancante è costituito dalla mobilità occasionale; questa componente, sempre più importante negli ultimi anni, è lasciata quindi alla valutazione tramite altre

fonti dati. A questo fine molto importante è l'indagine di mobilità Audimob in cui vengono indagati tutti i motivi dello spostamento; altra differenza sostanziale fra i due tipi di indagine è legata al tipo di output: gli indicatori che derivano dal censimento ISTAT rappresentano gli individui che si spostano, mentre l'indagine Audimob dà una rappresentazione della struttura degli spostamenti, definiti dalla motivazione principale che induce l'individuo a spostarsi (ad esempio, il tragitto dall'abitazione al luogo di lavoro, con una sosta per il rifornimento di carburante, è considerato come un singolo spostamento con motivazione 'lavoro', senza interruzioni). D'altra parte, il campione Audimob risulta ridotto rispetto al campione del censimento, essendo composto di 16.000 interviste contro le circa 500.000 famiglie coinvolte nell'indagine censuaria. Altre informazioni che contribuiscono a costruire il quadro complessivo della mobilità sono i dati raccolti dai gestori delle infrastrutture di trasporto e dagli operatori dei servizi di trasporto pubblico. Dal lato delle infrastrutture stradali si hanno a disposizione i dati provenienti dal sistema di rilievo dei flussi stradali delle strade gestite da ANAS che conta numerosi punti di conteggio distribuiti su tutto il territorio nazionale (si veda successivo Capitolo 4 "Le fonti dati sulla mobilità in Italia"), talvolta anche le stesse Regioni (per esempio Toscana ed Emilia Romagna) e le Città Metropolitane possiedono un sistema di punti di misura del traffico sulle strade di competenza, a cui si può attingere, analogamente ai dati

ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

ANAS, previo pagamento o gratuitamente su richiesta motivata. Questo tipo di dati fornisce un'informazione puntuale riferita alla georeferenziazione della postazione di rilievo, quantificando il flusso che transita in quella particolare sezione della rete; di contro, non è possibile ricavare indicazioni relative all'origine/destinazione dei flussi veicolari per cui i dati vengono spesso utilizzati o per la correzione delle matrici ottenute da modello di trasporto o per la validazione delle stesse. Sempre con riferimento al traffico veicolare, si hanno anche i dati di passaggio alle barriere autostradali; questi possono essere aggregati per rappresentare i flussi orari in uscita ed in ingresso ai caselli oppure, sfruttando l'informazione di punto di ingresso e uscita contenuta nel biglietto, generando una matrice OD giornaliera. Si tratta comunque di una matrice che rappresenta il fenomeno di mobilità veicolare in maniera parziale, perché si riferisce ai soli utenti del sistema autostradale e le origini e destinazioni sono costituite esclusi-

vamente dalle barriere autostradali. Altro elemento di attenzione è legato al fatto che non si ha un unico gestore autostradale. Nell'integrazione dei dati autostradali con le altre fonti di rilievo veicolare bisogna poi tenere presente che la definizione della tipologia dei veicoli (leggeri/pesanti) potrebbe non essere omogenea, in quanto il rilievo viene effettuato con dispositivi diversi da cui seguono criteri di classificazione non armonizzati, come già si può comprendere dal capitolo 4 "Le fonti dati sulla mobilità in Italia".

Anche per il trasporto ferroviario, attraverso i dati delle bigliettazioni possono essere ricostruite delle matrici ferroviarie, che come quelle autostradali, forniscono un quadro parziale della domanda di trasporto limitato alla specifica utenza ferroviaria e territorialmente limitata alle stazioni di inizio e fine del viaggio; inoltre viaggi che prevedono l'acquisto di due o più biglietti distinti (alta velocità e regionale) vengono registrati in funzione dell'origine e destinazione del singolo biglietto.



ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

Per quanto riguarda il trasporto pubblico su gomma non si hanno molti dati a disposizione; mancano i passeggeri dei bus che compiono viaggi su lunghe distanze e risulta complesso anche definire l'offerta di trasporto; a livello locale si conoscono eventualmente i numeri complessivi di biglietti venduti, mentre il numero di passeggeri per linea deriva da indagini tradizionali, con rilievo manuale dei presenti a bordo e saliti/discesi alle fermate, spesso effettuate su direttrici precise in funzione di specifiche necessità di studio. Manca ancora una raccolta sistematica delle

informazioni tramite dispositivi automatici di conteggio.

A livello locale altre indagini che potrebbero contribuire ad arricchire l'insieme dei dati sono quelle derivanti dai PUMS e dai piani spostamenti casa lavoro. I contenuti minimi richiesti per l'elaborazione di questi piani permettono un inquadramento del fenomeno della mobilità a livello locale molto interessante che però rimane circoscritto al progetto specifico mancando di un sistema informativo unico che permetta di rendere i dati raccolti disponibili per altri tipi di analisi.

60

Riquadro I – Indice del quadro conoscitivo

Quadro normativo, pianificatorio e programmatico

1. Livello regionale;
2. Livello sovralocale (piani e programmi di livello generale e di settore);
3. Livello locale (piani e programmi di livello generale e di settore);

Inquadramento territoriale e socio-economico dell'area di Piano

1. Struttura territoriale e insediativa;
2. Caratteristiche e dinamiche demografiche;
3. Imprese e dinamiche occupazionali;
4. Localizzazione di servizi e dei poli di attrazione;

Offerta di reti e servizi di trasporto

1. Rete stradale esistente e gerarchizzazione;
2. Reti e servizi di trasporto pubblico e nodi di interscambio;
3. Rete ciclabile, aree pedonali, Zone 30 e ZTL;
4. Sistema della sosta;
5. Servizi integrativi al trasporto pubblico e mobilità condivisa;
6. Logistica urbana;
7. Sistemi ITS e di informazione, regolamentazione e controllo della circolazione;
8. Politiche della mobilità;

Domanda di mobilità

1. Zonizzazione;
2. Indagini e rilievi sui flussi;
3. Matrici O/D degli spostamenti delle persone e delle merci, articolate nelle diverse modalità e suddivise per fasce orarie di punta e di morbida ed eventuali picchi stagionali;

Interazione tra domanda e offerta di trasporto

1. Livelli di Servizio della rete stradale e flussi di traffico;
2. Livelli di servizio sul TPL e flussi trasportati;
3. Flussi di traffico ciclabile e pedonale;
4. Rappresentazione delle dinamiche della logistica urbana;
5. Indice di utilizzo della sosta;

Criticità e impatti

1. Grado di accessibilità;
2. Congestione della rete stradale;
3. Saturazione dei servizi di TPL;
4. Incidentalità;
5. Impatti ambientali (parco veicolare, qualità dell'aria, inquinamento acustico, consumi energetici).

Punti di forza e di debolezza, opportunità e minacce

Tabella 3 – Sintesi delle caratteristiche delle principali indagini statistiche sulla mobilità passeggeri

ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

1. ANAGRAFICA
1.1. COMUNE DI RESIDENZA/DOMICILIO 1.2. CAP 1.3. SESSO 1.4. ETÀ
2. ATTIVITÀ LAVORATIVA
2.1. TIPOLOGIA DI CONTRATTO (tempo pieno, part time) 2.2. PRESENZA IN SEDE (per ciascun giorno della settimana) 2.2.1. Orario di ingresso 2.2.2. Orario di uscita
3. MEZZI DI TRASPORTO E ABBONAMENTI AI SERVIZI DI TRASPORTO NELLE DISPONIBILITÀ DEL DIPENDENTE
3.1. DISPONIBILITÀ DI MEZZI DI TRASPORTO (specificare) 3.2. DISPONIBILITÀ DI ABBONAMENTI AI SERVIZI DI TRASPORTO (specificare)
4. SPOSTAMENTI CASA-LAVORO
4.1. MODALITÀ ABITUALE DI SPOSTAMENTO 4.1.1. Utilizzo di un solo mezzo/modo di trasporto (specificare) 4.1.2. Combinazione di più mezzi/modalità di trasporto (specificare) 4.2. DISTANZA PERCORSA 4.3. TEMPO IMPIEGATO 4.4. MOTIVAZIONI SULLA MODALITÀ ABITUALE DI SPOSTAMENTO 4.5. GRADO DI SODDISFAZIONE SULLA MODALITÀ ABITUALE DI SPOSTAMENTO
5. PROPENSIONE AL CAMBIAMENTO NEGLI SPOSTAMENTI CASA-LAVORO: MOTIVI E CONDIZIONI VERSO FORME DI MOBILITÀ SOSTENIBILI
5.1. CAMBIAMENTO A FAVORE DEL TRASPORTO PUBBLICO 5.2. CAMBIAMENTO A FAVORE DELL'USO DELLA BICICLETTA 5.3. CAMBIAMENTO A FAVORE DEL CAR POOLING 5.4. CAMBIAMENTO A FAVORE DEL CAR SHARING 5.5. CAMBIAMENTO A FAVORE DELLO SCOOTER SHARING 5.6. CAMBIAMENTO A FAVORE DEL BIKE SHARING 5.7. CAMBIAMENTO A FAVORE DELL'USO DI MONOPATTINI IN SHARING

61

Figura 7 - Informazioni minime necessarie da rilevare nella redazione del PSCL, fonte Linee guida per la redazione e l'implementazione dei Piani degli Spostamenti Casa-Lavoro (PSCL) Decreto Interministeriale n. 179 del 12 maggio 2021

Facendo riferimento alle fonti dati ufficiali, descritte nel capitolo 4 "Le fonti dati sulla mobilità in Italia", risulta complesso e non immediato il confronto fra le stime di mobilità ottenute, proprio per le differenze sostanziali presenti nella rilevazione legate alle diver-

se finalità delle indagini. Questo aspetto è emerso chiaramente da un approfondimento realizzato da FS Research Centre e finalizzato all'analisi e il confronto dei dati statistici sulla mobilità tratti dalle principali fonti nazionali. Le fonti analizzate sono state il Conto Nazionale



delle Infrastrutture e della Mobilità Sostenibili (CNIMS), le indagini Istat e i rapporti sulla mobilità pubblicati da Isfort nell'ambito dell'Osservatorio "Audimob"; essendo Istat una fonte per il CNIMS, il confronto si è ridotto a due termini di paragone: CNIMS-Istat e Audimob. La prima criticità nel confronto ha riguardato il differente periodo a cui si riferiscono i due insiemi di dati: il confronto con i dati del CNIMS, tipicamente annuali, ha costretto dunque a delle elaborazioni di espansione all'anno dei dati dell'indagine svolta da Isfort, riferiti al giorno feriale/festivo medio. I dati sul giorno feriale medio e sul giorno festivo medio sono stati opportunamente sommati e riportati ad un valore annuale attraverso l'utilizzo di adeguati coefficienti di espansione. In questa operazione sono state introdotte due ipotesi: la prima, per stimare il valore di alcuni dati, per esempio i passeggeri-km distinti per modo, sul giorno festivo medio non direttamente pubblicati da Audimob; la seconda, per stabilire il numero di giorni feriali e festivi in un anno. Altro elemento di distinzione fra le fonti dati confrontate è legato all'ambito territoriale a cui si riferiscono: il fenomeno indagato dall'Osservatorio attraverso il diario giornaliero è prevalentemente quello della mobilità locale, con gli spostamenti urbani favoriti su quelli extraurbani. Per procedere

con le valutazioni desiderate si è fatto uso di un'ulteriore rilevazione di Isfort: attraverso un'apposita sezione del questionario Audimob, vengono rilevati anche gli spostamenti effettuati nei 7 giorni precedenti all'intervista di almeno 10km al di fuori del comune di normale residenza. Questi prendono il nome di Spostamenti extraurbani e vengono restituiti sotto forma di valori settimanali medi. Tali spostamenti sono complementari agli spostamenti del diario di bordo giornaliero; pertanto, nello studio di confronto qui descritto e riferito all'anno 2021, sono stati integrati nel calcolo del valore annuale complessivo, dopo essere stati riportati ad annuali attraverso opportuni coefficienti di espansione. L'inserimento dei dati extraurbani è risultato fondamentale per una corretta lettura della reale mobilità captata dalle indagini Audimob. I confronti sono stati fatti a più livelli, si è partiti da un confronto generale sulla mobilità motorizzata totale, per arrivare al dettaglio dei singoli modi di trasporto. Gli indicatori di mobilità considerati nei confronti sono stati il numero di passeggeri (o spostamenti) e i passeggeri-km. Dall'analisi, circoscritta ai dati dell'anno 2021, sono emerse differenze importanti fra i passeggeri per modo di trasporto di CNIMS-ISTAT e quelli desumibili dall'indagine di ISFORT espansi all'anno; Audimob presenta a livello generale

un valore inferiore a quello CNIMS fra il 30 ed il 60%, fatta eccezione per i passeggeri del trasporto ferroviario in cui Audimob ha un valore superiore del 10%. Le differenze fra i due insiemi possono essere motivate in parte tenendo presente il ridotto universo di osservazione Audimob ed in parte considerando la distinzione tra tratte e spostamenti fatta nello stesso questionario. Uno spostamento casa-lavoro eseguito attraverso una combinazione di mezzi di trasporto viene diviso in tratte (una per ciascun mezzo utilizzato), ma comunque conteggiato come unico spostamento; pertanto, in questi casi ciò che CNIMS conta come 2 o più passeggeri, risulta per Audimob come unico passeggero. Quest'ultima motivazione sembra trovare conferma nella riduzione delle differenze che si ha confrontando i passeggeri-km piuttosto che i passeggeri. I passeggeri-km dei mezzi motorizzati nel loro complesso stimati espandendo all'anno i valori Audimob sono circa un 30% in meno rispetto a quelli CNIMS; minori le differenze fra passeggeri-km relative ai mezzi pubblici, più marcate quelle della mobilità motorizzata privata. Nell'ultimo caso è altamente probabile che vi sia anche una sovrastima da parte del Conto, perché, come dato di partenza, viene considerato il parco veicolare definito da ACI, che, tuttavia, a sua volta chiarisce

come questo non coincida con il reale numero di veicoli circolanti sulle strade italiane, che è un dato difficilmente conoscibile.

Dall'analisi appena descritta risulta evidente che non esiste un'indagine che contenga la "verità assoluta" ma indagini che raccontano diversi aspetti e dimensioni temporali e spaziali della mobilità. Per esempio, il diario di Audimob fornisce un quadro dettagliato degli spostamenti soprattutto in ambito urbano, caratterizzando in maniera precisa la catena dei mezzi utilizzati ed i motivi dello spostamento, riferendosi al giorno ferial medio⁵; questo è prezioso per la calibrazione di modelli di generazione/distribuzione e di scelta modale differenziati per tipologia di utenza. Altre fonti dati forniscono invece, come CNIMS-Istat, una visione globale quantitativa del fenomeno di mobilità a livello nazionale, o descrivono puntualmente le condizioni di deflusso della rete veicolare e/o i livelli di servizio delle infrastrutture come i dati di sezioni di rilievo. L'uso congiunto di diverse fonti dati è peraltro una consuetudine nella costruzione dei modelli di trasporto per calibrare, validare o fornire valori dei flussi espansi dalla dimensione temporale di simulazione del modello alla dimensione necessaria per analisi di altro tipo, come per esempio Analisi Costi Benefici o analisi di impatto ambientale.



5. Nelle ultime edizioni del questionario Audimob, Isfort rileva anche gli spostamenti nei giorni festivi.

5.2 | RAFFORZARE IL POTERE INFORMATIVO DELLE INDAGINI TRADIZIONALI ATTRAVERSO L'USO DEI BIG DATA

I Big Data, e nello specifico i dati estratti dalla telefonia mobile possono contribuire insieme alle indagini tradizionali ad arricchire ed apportare nuove informazioni al puzzle rappresentativo del fenomeno con un approccio di combinazione e selezione degli elementi di

interesse in funzione della necessità, approccio già adottato per altro con le statistiche ufficiali. Alcune criticità verranno analizzate più in dettaglio nel successivo Capitolo 7 "L'estrazione dell'informazione".

5.2.1 | CASI DI STUDIO: UN APPROCCIO SPERIMENTALE.

Un'applicazione efficace dell'uso congiunto di dati da indagini tradizionali e Big Data è il modello di trasporto nazionale di lunga percorrenza sviluppato da Isfort per Ferrovie dello Stato Italiane, che vedrà prossimamente l'implementazione per i viaggi di media distanza (inferiori agli 80 km).

I quattro stadi del modello utilizzano fonti dati tradizionali come l'indagine Audimob, dati censuari ISTAT, flussi di rilievo stradali, dati di bigliettazione e saliti/discesi nelle stazioni, insieme a Big Data di vario tipo.

In particolare, sono stati utilizzati i Floating Car Data (FCD) provenienti dalle scatole nere delle compagnie assicurative e i dati di telefonia mobile la cui elaborazione per fini trasportistici ha visto la collaborazione fra l'operatore di telefonia Vodafone e il Rese-

arch Centre di Ferrovie dello Stato Italiane. Nello specifico, l'analisi dei FCD, ad integrazione del dato Audimob, ha consentito di ricostruire gli schemi dei viaggi di media e lunga percorrenza effettuati in auto e di evidenziare le differenze rispetto ai viaggi di corto raggio. Sempre nell'ottica del data fusion, i dati campionari sono stati riportati all'universo, utilizzando il parco veicolare circolante da fonte ACI 2019, per ottenere le matrici O/D dei viaggi di almeno 80 km in un giorno feriale medio. I dati telefonici sono stati utilizzati per quantificare tutti gli spostamenti, per ciascun giorno del mese di ottobre 2019, effettuati dalla popolazione residente e dagli stranieri presenti sul territorio nazionale, con riferimento a una disaggregazione su zone di traffico opportunamente individuate.



ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

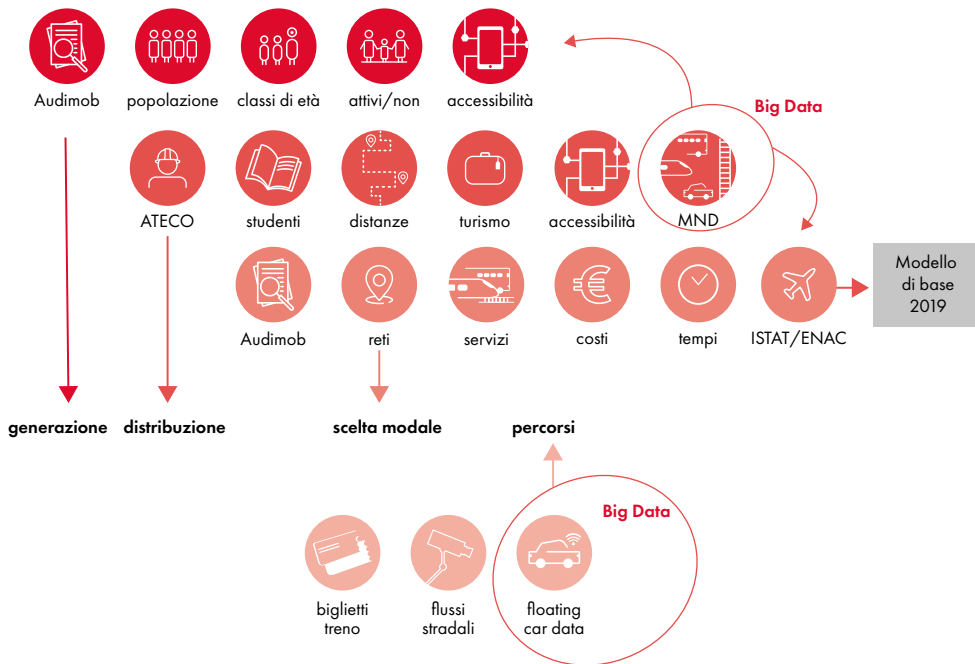


Figura 8 - Dati utilizzati nel Modello Nazionale di Lunga percorrenza (>80 km) di Isfort per Ferrovie dello Stato Italiane - Immagine FS Research Centre

Nel caso esemplificativo sopra riportato i dati telefonici sono stati quindi utilizzati per ricostruire una matrice generale degli spostamenti mentre il modello è stato utilizzato insieme a indagini preesistenti ed altri dati di contesto, per ricostruire le matrici disaggregate. Questo tipo di approccio corrisponde al primo dei due individuati da Willumsen [1] per l'utilizzo dei Big Data nelle analisi di mobilità; l'altro prevede invece di segmentare quanto più possibile la domanda di mobilità rilevata tramite i Big Data, magari utilizzando tecniche

di Data Fusion: per esempio, combinare gli spostamenti rilevati attraverso i dati telefonici con il servizio programmato o ancora meglio con quello effettivamente realizzato nei giorni di analisi dai servizi pubblici per verificare la compatibilità dello spostamento con il modo di trasporto pubblico ed eventualmente quindi attribuire allo spostamento quel modo. Questo secondo tipo di approccio è rappresentato dall'esempio descritto nel Capitolo 7 "L'estrazione dell'informazione".

5.2.2 | OPPORTUNITÀ E LIMITI INDIVIDUATI

I Big Data forniscono nuove opportunità di sviluppo per la modellistica dei trasporti:

- Potrebbero ridurre la necessità di estese e costose campagne di indagine o in alternativa la dimensione campionaria delle stesse;
- Riducono i tempi di raccolta dei dati;
- Permettono di fotografare la mobilità in maniera estesa sia territorialmente che temporalmente, dando la possibilità di cogliere la sua variabilità spaziale e/o temporale e la risposta a sollecitazioni esterne come nel caso dell'evento pandemico.

	Household Travel Surveys	Big Data
Costo	Elevato	
Dimensione del campione	al max 2% dell'area di interesse	Dal 15% al 40% nel caso dei MND in funzione della suddivisione del mercato fra le varie compagnie telefoniche.
Validità del dato raccolto	Diventano velocemente obsoleti, questo è ancora più vero per effetto dell'evento pandemico.	I provider telefonici hanno a disposizione dati relativi a lunghi periodi temporali.
Attitudine persone	le lunghe interviste aumentano il disturbo statistico; i rispondenti tendono a semplificare e omettere la descrizione degli spostamenti	Nel caso delle app, il rispondente può scegliere di non fornire dati di georeferenziazione o di fornirli solo quando l'app è in uso. Non raccolgono informazioni comportamentali a differenza delle HTS.
Tempo esecuzione	Richiedono tempi lunghi per la raccolta dei dati	In breve tempo posso avere a disposizione un grande numero di dati per un esteso periodo temporale
Precisione	Non sono esenti da errori e devono essere sottoposte a revisione post acquisizione	La distribuzione nel tempo e nello spazio dei dati raccolti dipende dall'interazione con la cella e dall'utilizzo delle app; ne segue che non è sempre possibile identificare con precisione inizio e/o fine dello spostamento.

Tabella 4 – Confronto fra indagini tradizionali e tramite Big Data - Elaborazione dei contenuti di Willumsen, L. (2021), "Use of Big Data in Transport Modelling" [1]

ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

Non sono tuttavia esenti da limiti come ben evidenziato da Willumsen (2021) [5.1] e come emerge dalle esperienze descritte nel Capi-

to 7 "L'estrazione dell'informazione "; di seguito una tabella di sintesi dei limiti evidenziati dall'autore.

Limite	Descrizione
Impossibilità di avere istante di inizio e fine di tutti gli spostamenti con precisione.	La distribuzione nel tempo e nello spazio dei dati raccolti dipende dall'interazione con la cella e dall'utilizzo delle app; ne segue che non è sempre possibile identificare con precisione inizio e/o fine dello spostamento e la sua durata.
Impossibilità di identificare gli spostamenti di breve entità	Per gli MNO è legato alla dimensione delle celle; in più si ha il fenomeno dei phantom trips o cell jumps (falsi movimenti rilevati dovuti all'aggancio a celle telefoniche differenti nel periodo di osservazione del dispositivo mobile che è in realtà fermo nella stessa posizione). Per i dati da app dalla frequenza/soglia temporale con cui collezionano gli stamps degli eventi.
Espansione del campione	Presenta numerose criticità, vengono utilizzate anche informazioni provenienti dai contratti telefonici.
Riconoscimento dello scopo del viaggio	Fatta eccezione per gli spostamenti sistematici, lavoro/studio, risulta difficile perché le aree residenziali spesso presentano una commistione di usi del suolo. Risulta difficile distinguere viaggi sistematici e non sistematici (ad esclusione di lavoro/studio).
Riconoscimento del modo di trasporto	Questa criticità si presenta soprattutto nelle aree urbane

Tabella 5 – Limiti dei Big Data provenienti dalla telefonia (MNO e app data) - Elaborazione dei contenuti di Willumsen, L. (2021), "Use of Big Data in Transport Modelling" [1]

Willumsen [5.1] individua come possibile soluzione ai limiti dei Big Data per le analisi di mobilità proprio l'utilizzo congiunto degli stessi con dati di contesto e dati di mobilità esistenti, introducendo il concetto di "Data Fusion": fra i dati di contesto individua i dati censuari, i dati di utilizzo del suolo, i punti di interesse, i dati provenienti da indagini tradizionali preesistenti, mentre per quanto riguarda quelli specifici della mobilità dati

di traffico e conteggi di persone, percorsi e frequenze dei mezzi pubblici, dati di biglietti o dati provenienti dalle smart card del trasporto pubblico.

Come evidenziato dallo stesso Willumsen [5.1] bisogna porre attenzione alle differenze fra i dati utilizzati nel Data Fusion, ad esempio con riferimento alle differenti distribuzioni degli errori presenti negli insiemi campionari che si vanno a combinare.

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

Si evidenzia inoltre che potrebbero non coincidere le definizioni base, a partire da quella di "spostamento". Nella tabella seguente si riportano le definizioni dalle Eurostat Guidelines on Passenger mobility statistics (GL⁶) dei

vari elementi considerati per le analisi di mobilità, messe a confronto con quelle utilizzate nel caso studio con i dati di telefonia mobile (MND) descritto nel successivo Capitolo 7⁷ L'estrazione dell'informazione .

Concetto	Indagini di mobilità tradizionali			Indagini di mobilità tramite MND		
	Fonte	Definizione tematica	Fattibilità (SI/NO/in parte)	Fonte MND	Definizione per algoritmo MND	Fattibilità (SI/NO/in parte) MND
Popolazione	Eurostat Guidelines on Passenger mobility statistics (GL)	Popolazione residente in Italia di età 15-84		caso studio	Tutte le persone con un telefono che corrispondono approssimativamente a chi ha almeno 12 anni di età	
Quota di Popolazione mobile	GL	Quota della popolazione di riferimento che ha effettuato almeno uno spostamento nel periodo di osservazione		caso studio	Quota della popolazione di riferimento che ha effettuato almeno uno spostamento di almeno 800m nel periodo di osservazione	
Popolazione fuori campo di osservazione	GL	Oltre alla definizione rispetto alle classi di età, non devono essere osservati gli spostamenti effettuati nell'ambito dello svolgimento di un'attività lavorativa (es. tassisti/NCC in servizio, autisti di tir durante l'attività di trasporto)		caso studio		Attualmente non applicabile perché all'interno delle SIM tracciate sono presenti anche spostamenti legati ad attività lavorative (es. autotrasportatori)

6. https://ec.europa.eu/eurostat/documents/29567/3217334/Guidelines_on_Passenger_Mobility_Statistics+%282018_edition%29.pdf/f15955e3-d7b4-353b-7530-34c6c94d2ec1?i=1611654879518

ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

Concetto	Indagini di mobilità tradizionali			Indagini di mobilità tramite MND		
	Fonte	Definizione tematica	Fattibilità (SI/NO/ in parte)	Fonte MND	Definizione per algoritmo MND	Fattibilità (SI/NO/ in parte) MND
Spostamento	GL	Movimento da un'origine a una destinazione finale, qualificate da un luogo e da una motivazione o attività		caso studio	La concatenazione di tutte le eventuali soste intermedie se di durata inferiore ad un periodo temporale prestabilito all'interno di un unico viaggio origine-destinazione	
Tappa	GL	Segmento dello spostamento caratterizzato da un'unica modalità di trasporto (anche su mezzi diversi nell'ambito della stessa modalità, es. cambio di linea TPL)		caso studio		In questo momento non è applicabile perché si tratta di spostamenti a mezzo prevalente. Da analizzare la possibilità di applicare la definizione da GL in un caso studio ulteriore.
Lunghezza della tappa/ dello spostamento	GL	Km percorsi nella tappa/nello spostamento		caso studio	Somma delle distanze percorse all'interno dello spostamento	
Durata della tappa/ dello spostamento	GL	tempo di durata in minuti della tappa/ dello spostamento		caso studio	Tempo di durata in minuti dello spostamento superiore a 800m	
Mobilità locale	GL	Insieme degli spostamenti al di sotto dei 300 km		caso studio	Insieme degli spostamenti al di sotto dei 300 km (con soglia minima 800m)	Fattibile ma non applicato nel caso studio

**ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE
ATTRAVERSO L'USO DI BIG DATA**

Concetto	Indagini di mobilità tradizionali			Indagini di mobilità tramite MND		
	Fonte	Definizione tematica	Fattibilità (SI/NO/in parte)	Fonte MND	Definizione per algoritmo MND	Fattibilità (SI/NO/in parte) MND
di cui urbana (2 definizioni alternative):				caso studio		
Mobilità urbana (1)	GL	Insieme degli spostamenti all'interno di una Functional Urban Area		caso studio	Insieme degli spostamenti all'interno di una Functional Urban Area (soglia minima 800m)	Fattibile ma non applicato nel caso studio
Mobilità urbana (2)	GL	Insieme degli spostamenti al di sotto dei 100 km		caso studio	Insieme degli spostamenti al di sotto dei 100 km (soglia minima 800m)	Fattibile ma non applicato nel caso studio
Mobilità su distanze medie	GL	viaggi la cui percorrenza è fra i 301 e i 999 Km		caso studio	viaggi la cui percorrenza è fra i 301 e i 999 Km	Fattibile ma non applicato nel caso studio
Mobilità su distanze lunghe	GL	viaggi la cui percorrenza è pari a 1000 Km e oltre		caso studio	viaggi la cui percorrenza è pari a 1000 Km e oltre	Fattibile ma non applicato nel caso studio
Numero di pernottamenti (per viaggi su distanze medio/lunghe)	GL	numero di notti trascorse tra il viaggio di andata e quello del rientro		caso studio		Potrebbe essere un tema di approfondimento per un ulteriore studio, da valutare le criticità legate agli aspetti della privacy.
Modo (o modalità) di trasporto	GL	Tipologia di veicolo utilizzato per la tappa (include la modalità 'a piedi')	in parte	caso studio		In questo momento non è applicabile perché si tratta di spostamenti a mezzo prevalente. Da analizzare la possibilità di applicare la definizione da GL in un caso studio ulteriore

ANALISI DI ALCUNI CASI STUDIO RELATIVI ALLA MOBILITÀ DELLE PERSONE

Concetto	Indagini di mobilità tradizionali			Indagini di mobilità tramite MND		
	Fonte	Definizione tematica	Fattibilità (SI/NO/in parte)	Fonte MND	Definizione per algoritmo MND	Fattibilità (SI/NO/in parte) MND
Modo (o modalità) di trasporto prevalente	GL	Tipologia di veicolo utilizzato per la tappa più lunga di uno spostamento in termini di km percorsi (include la modalità 'a piedi'); è il modo attribuito all'intero spostamento/viaggio	in parte	caso studio	Ad oggi nel caso studio si tratta solo il mezzo prevalente. I modi distinti sono treno, aereo ed altro. Affinchè un viaggio risulti compiuto in treno deve risultare che almeno la metà della distanza di viaggio sia effettuata con il treno. Ad ogni modo se nello stesso viaggio è presente anche l'aereo il modo di viaggio verrà comunque considerato solamente aereo.	Da analizzare la possibilità di applicare la definizione da GL in un caso studio ulteriore
Alimentazione dell'autovettura	GL	Tipo di carburante utilizzato per lo spostamento/viaggio con autovettura privata (conducente o passeggero)	NO	caso studio		NO
Motivazione dello spostamento	GL	Attività principale che si svolgerà a destinazione e che determina l'effettuarsi dello spostamento.	in parte	caso studio		Non è stato ancora affrontata questo tipo di analisi. Da letteratura è noto come, fatta eccezione per gli spostamenti ricorrenti come lavoro/scuola, sia di difficile determinazione.
Numero di occupanti del veicolo	GL	numero complessivo di persone che coesistono in un'autovettura.	NO	caso studio		NO

Tabella 6 – Confronto fra definizioni da normativa/linee guida per le statistiche sulla mobilità passeggeri e definizioni da indagini di mobilità tramite MND provenienti dal caso studio illustrato nel capitolo 7 "L'estrazione dell'informazione".

5.3 | BIBLIOGRAFIA

[5.1] Willumsen, L. (2021), "Use of Big Data in Transport Modelling", International Transport Forum Discussion Papers, No. 2021/05, OECD Publishing, Paris.

[5.2] Tartaglia M., Nourbakhsh S., Vannacci L., Chindemi A., Carbone G., Ferrara M., Sommaro W., Marino M. (2023), "Un modello multimodale per la simulazione della mobilità di media e lunga percorrenza delle persone in Italia" - Ingegneria Ferroviaria 3 (marzo 2023), 217-250.



6 | IL CICLO DI VITA DEI BIG DATA

Roberta Radini¹, Monica Scannapieco¹

1. ISTAT, Roma, Italia

6.1 | ARCHITETTURA DI RIFERIMENTO

Il ciclo di vita dei dati viene documentato dalla descrizione delle principali fasi che lo compongono. Le fasi vengono scomposte in singoli processi a partire dall' integrazione dei dati, che dalla forma iniziale, ossia quella "grezza" corrispondente alla struttura del sistema che li genera, arriva al processo di trasformazione finale che consente di estrarre l'informazione di interesse, il loro "Valore". Questo processo generale viene tarato per ogni fonte big sulle specificità di volume, valore, velocità e tutte le "V" che li caratterizzano, sulla tipologia di dato e sulla fonte che li genera⁷. Tipicamente il processo che descrive il ciclo di vita dei dati ha una struttura ciclica che consente di valutare, correggere e/o migliorare gli output finali, ma in particolare la ciclicità per i Big Data consente di ottenere un processo robusto rispetto alle variazioni della fonte che possono essere di varia natura, ma sicuramente non controllabili dall'utilizzatore finale.

Nell'ambito della statistica ufficiale è stato lanciato da Eurostat il progetto ESSnet Big Data Pilots II [6.1], che tra i suoi principali obiettivi ha avuto quello di definire uno standard di architettura dei processi di trattamento dei Big Data (Workpackage "Process and Architecture").

Sulla scorta dei risultati del progetto precedente, ESSnet Big Data Pilots I, è stata ravvisata l'esigenza di formalizzare un'architettura di riferimento con lo scopo di guidare

gli investimenti nei Big Data da parte degli istituti nazionali di statistica (NSI) e aiutare lo sviluppo di soluzioni e servizi standardizzati da condividere all'interno dell'ESS (European Statistical System) e oltre.

La modellazione delle fasi dell'intero ciclo di vita dei Big Data, predisposto nel progetto ESSnet Big Data Pilots II [6.2] come standard per la definizione di indicatori o più in generale di stime ottenibili dal trattamento dei Big Data, può essere utile per standardizzare un processo di produzione e ad analizzare le risorse necessarie alla sua gestione. Gli standard di processo non debbono diventare un vincolo ma un ausilio a identificare tutti i passi e le risorse necessarie.

Nella modellazione del ciclo di vita dei dati è posta una particolare attenzione agli *utenti coinvolti* nel processo e al loro ruolo, in particolare nel nuovo scenario di gestione dei Big Data, e con particolare attenzione verso i loro interessi di Business, come ad esempio:

- NSI mirano a introdurre l'uso dei Big Data nei loro processi produttivi;
- Organizzazioni pubbliche e private potrebbero essere interessate a seguire un modello definito e controllato di produzione di statistiche basate sui Big Data, guidate dall'esperienza della statistica ufficiale.

Viene descritto di seguito questo modello, riprendendo i risultati del progetto citato,

7. *human generated, machine generated, business generated.*

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

perché può essere utilizzato:

- Come quadro di riferimento da utilizzare da parte degli esperti di Architettura di Business e IT (secondo il modello del Enterprise Architects) per allineare le esigenze aziendali e IT delle organizzazioni interessate a implementarlo;
- Come linguaggio per descrivere i progetti di sistemi informativi che fanno uso di sorgenti Big Data;
- Come strumento per il top management per pianificare gli investimenti aziendali relativi ai progetti Big Data, tenendo conto delle economie di scala offerte dalle infrastrutture e dai servizi per i Big Data anche a livello internazionale.



74

6.1.1 | PROPOSTA DI STANDARDIZZAZIONE DI PROCESSO BIG DATA EUROSTAT: BREAL

Nell'ambito del progetto europeo ESSnet Big Data è stato formalizzato il modello BREAL (Big Data REference Architecture and Layers), sviluppato dal Workpackage "Process and Architecture", che viene di seguito illustrato e ripreso da [6.3].

La modellazione del BREAL è stata formalizzata secondo la rappresentazione a livelli dell'Enterprise Architecture [6.4] comprende la descrizione di alcuni artefatti che ne costituiscono: i principi, le funzioni, i processi del ciclo di vita dei Big Data e i possibili Stakeholder su cui si modella il Business Layer, vedi Figura 9.

In dettaglio vengono rappresentati e classificati un insieme di servizi applicativi generali, proposti per mostrare come possono

essere implementate le funzioni di business individuate. In particolare, viene descritto il flusso dei dati, vedi par.6.1.2, proposto per le Trusted Smart Statistics, che modella, su uno schema a tre livelli, detto "modello a clessidra", tutti i passi di trasformazione dei dati da grezzi a statistici.

Il Progetto ha previsto anche una fase di applicazione del modello generale su quattro specifici ambiti statistici per testarne la consistenza e l'applicabilità. I processi presi in esame sono stati:

- Online job vacancies (WPB);
- Online based enterprise characteristics (WPC);
- Smart energy (WPD);
- Tracking ships (WPE).

In seguito, è stato applicato anche alla modellazione del processo di costruzione di indicatori sulla Mobilità e alla stima della popolazione abitualmente dimorante [6.5] analizzato dal Workpackage I (WPI) focalizzato su mobile networks data.

L'architettura generale del BREAL ha seguito la tipica schematizzazione a tre livelli della Enterprise Architecture e dell'individuazione dei componenti ritenuti di interesse nella modellazione, in particolare per:

- *Business Layer* che descrive cosa si deve fare ('what') per la gestione dei Big Data, declinati in termini di "artifacts" che rappresentano: principi, funzioni di business, ciclo di vita dei dati dello specifico

processo di produzione, attori coinvolti e utilizzatori finali;

- *Application Layer*, che descrive come si possono realizzare ('how') le funzioni e i servizi che compongono il processo del ciclo di vita dei dati;
- *Information Layer*, che descrive come si possono realizzare ('how') i modelli dati o gli algoritmi (operational model) implementativi delle trasformazioni ed elaborazione dei dati. La classificazione dei dati da "nano-dati" e "micro-dati" a "macro-dati"⁸

Il livello Tecnologico che descrive le soluzioni tecniche a livello di SW e HW non rientra nella modellazione del BREAL perché legato ad aspetti tecnici e non di modellazione.

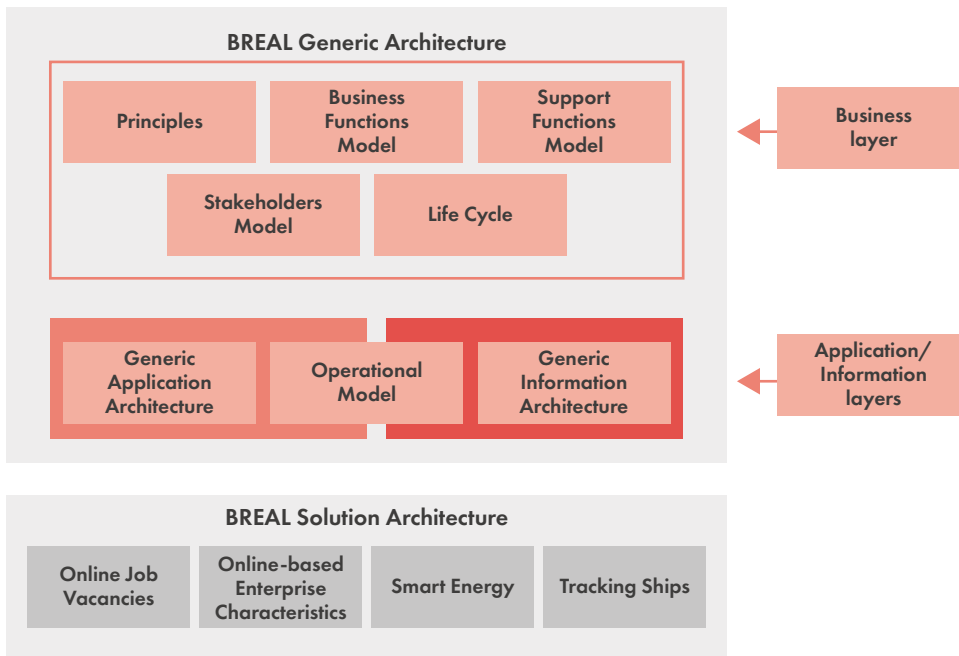


Figura 9: Schema generale dei livelli architetturali del BREAL.

8. Il termine "nano-dati" si riferisce in generale a fonti di dati in cui i dati si riferiscono a eventi (ad es. traiettorie, tratte e viaggi) definiti con una granularità più fine rispetto a quella dei "micro-dati", cfr. <https://doi.org/10.1017/dap.2020.7>. Mentre i record di micro-dati generalmente si riferiscono "uno a uno" a unità statistiche (ad esempio, singole persone o famiglie), nei set di nano-dati il numero di dati riferiti a una singola unità statistica può essere "uno a molti". I nano-dati sono chiamati in letteratura anche "granular data" o "behavioural data", ossia che descrivono i comportamenti di una singola entità statistica.

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

Il ciclo di vita modellato da BREAL, come mostrato in Figura 10, si focalizza sulla rappresentazione delle funzioni di business che compongono i tre processi principali, come:

- *Development and Information Discovery* – processo che si occupa dell'esplorazione della fonte Big Data, della sua integrazione con altri dati e della "scoperta" delle informazioni, in modo induttivo e/o deduttivo dall'elaborazione della fonte stessa;
- *Production* – processo che modella i passi di creazione di prodotti statistici attraverso l'uso di sorgenti Big Data e non solo;
- *Continuous Improvement* – processo di monitoraggio e valutazione della qualità dell'utilizzo delle sorgenti Big Data, con particolare attenzione alle problematiche di copertura della popolazione di riferimento e alla validità dei modelli utilizzati. Questa è la fase che chiude il ciclo, perché eventuali evidenze di correzione o di miglioramento danno avvio alla progettazione di un nuovo processo di Development and Information Discovery e/o Production.

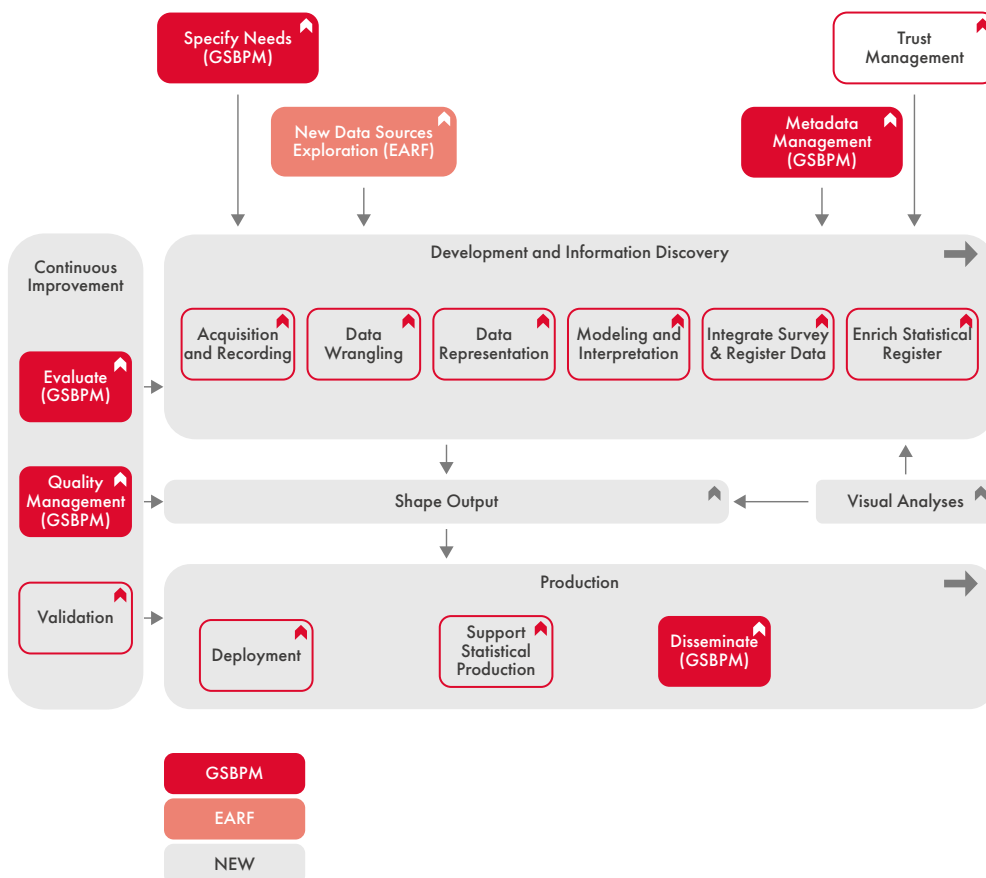
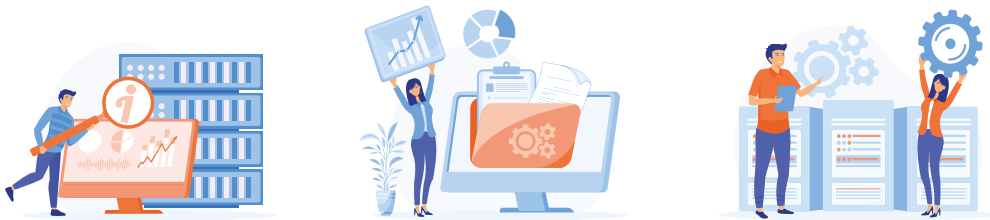


Figura 10: Ciclo di vita BREAL.

Come è evidenziato dai colori in Figura 10, alcuni dei processi di Business presi in considerazione sono già stati modellati e descritti da altri modelli di standardizzazione come GSBPM e ERF, ma nell'analisi di questo nuovo modello questi standard non sono stati

giudicati completi rispetto alla modellazione di tutte le attività da mettere in campo nell'analisi del ciclo di vita dei Big Data. Il processo è stato schematizzato secondo il linguaggio ArchiMate che è un linguaggio open per la modellazione della Business Architecture.



6.1.2 | ARCHITETTURA DEL FLUSSO DEI DATI

Come detto il processo di gestione dei dati prevede, in corrispondenza di ogni step del processo, uno "stato" dei dati che si trasformano da grezzi a dati statistici. Questo modello formalizzato come Generic Information Architecture for Big Data (GIAB) è costituito da tre livelli rappresentati dal "modello a clessidra" proposto per le Trusted Smart Statistics. I tre livelli della clessidra sono costituiti da: la base della clessidra dove sono modellati i dati grezzi, nel mezzo i dati che rappresentano il punto di convergenza, ossia di trasformazione da grezzi a dati statistici, mentre il livello in alto rappresenta i dati statistici su cui vengono effettuate le analisi statistiche e il calcolo di indicatori di interesse. In particolare, i tre livelli si possono descrivere come segue:

- Il Raw Data Layer include i dati acquisiti e archiviati dalla funzione "Acquisition and Recording". Nel modello GIAB si descrivono i concetti e non viene fornito alcun dettaglio su formati o altre specifiche tecniche che possono essere utili per l'acquisizione e la memorizzazione dei dati grezzi.

- Il Convergency Layer contiene i dati riferiti alle unità di interesse per le analisi. Questi dati sono prodotti come risultato delle funzioni del ciclo di vita BREAL di "Data Wrangling" e "Data Representation".
- Lo Statistical Layer include quei concetti che sono gli obiettivi dell'analisi. Questi dati sono prodotti principalmente da "Modeling and Interpretation", "Integrate Survey and Register Data", "Enrich Statistical Registers" e "Shape Output".

Nella modellazione dei tre livelli viene effettuata una descrizione dei dati e dei metadati. In particolare, per i Big Data ricoprono una grande rilevanza i metadati di provenienza (Lineage), ossia non solo quelli specifici della singola fonte, ma anche le particolarità legate ai dati forniti da un particolare provider (es: i dati di telefonia sono legati alla rete telefonica che pur essendo standardizzata ha delle particolarità legate ai singoli provider di telefonia). Questo modello è stato declinato nell'elaborazione dei dati di telefonia mobile nel "Reference Methodological Framework" (RMF) [6.6].

6.2 | RUOLI E ATTORI

I Big Data possono essere prodotti da un unico fornitore, ossia da un'unica piattaforma che li genera (es: Twitter, Facebook,...), oppure da diversi provider proprietari di diverse piattaforme che li generano secondo tecnologie condivise ma con formati diversi (es: dati di telefonia mobile, di siti WEB, ...).

Contrariamente ai sistemi di dati tradizionali che tendono a essere archiviati, sviluppati e distribuiti per una sola organizzazione, come i dati amministrativi o i dati economici di aziende, i sistemi di Big Data possono essere distribuiti da una o più organizzazioni e utilizzati da utilizzatori finali diversi, che in funzione dei rispettivi interessi, producono prodotti diversi.

Quindi a parità di ruolo, gli attori nei sistemi Big Data possono provenire da diverse organizzazioni. Il modello BREAL classifica gli attori che partecipano al ciclo di vita dei Big

Data come segue:

- IT & Statistical Pipeline Actors - figure caratterizzate da una sinergia di competenze che spaziano dalla statistica all'informatica;
- Capacity Providers - ossia i fornitori dei dati;
- Global roles - un insieme di attori istituzionali, come gli istituti di statistica che possono mettere a disposizione le esperienze in materia di Big Data, e attori che siano in grado di gestire gli aspetti di controllo e di gestione di un progetto di produzione, ma anche i cittadini stessi;
- Audit, Control, and Compliancy Actors - un complesso di attori, spesso esterni all'organizzazione che svolgono funzioni di controllo rispetto a diversi aspetti.



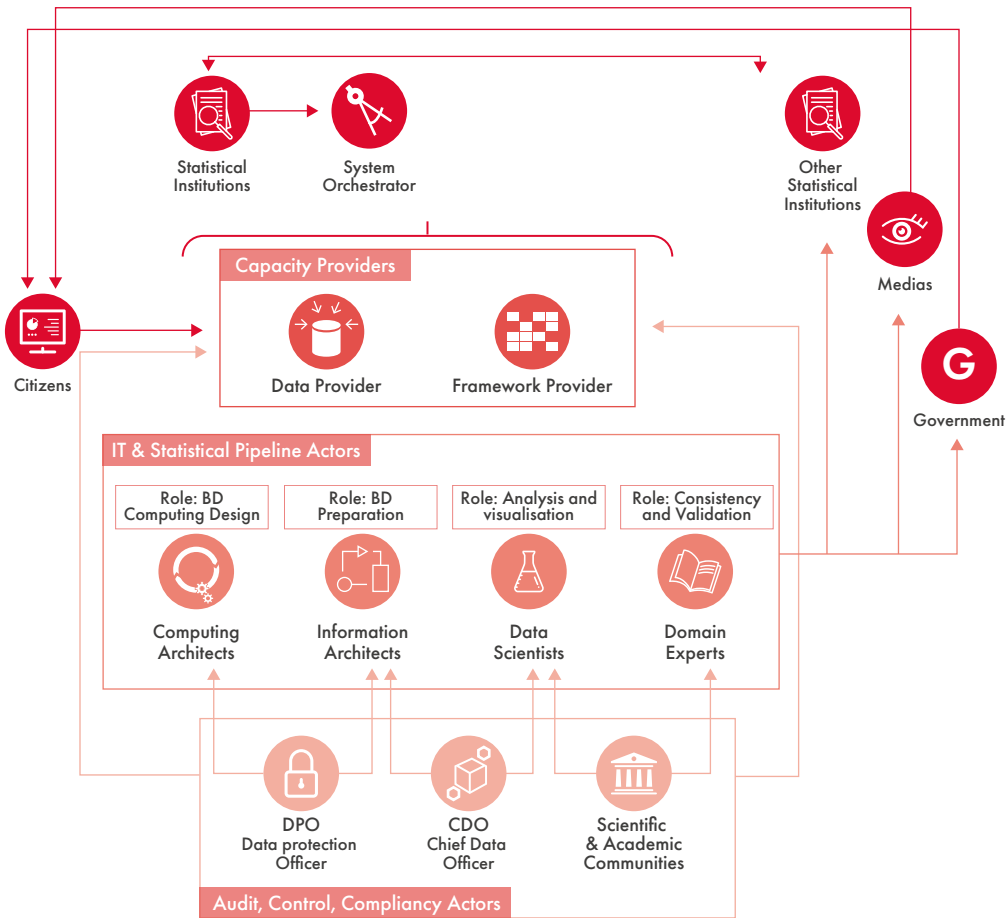


Figura 11: BREAL actors.

Nella Figura 11 sono rappresentati i ruoli dei diversi attori e come si relazionano. Nei paragrafi seguenti vengono descritti in dettaglio rispetto alle funzioni che sono chiamati ad implementare all'interno del modello di ciclo di vita BREAL proposto da EUROSTAT. Si mostra come le organizzazioni che vogliono utilizzare i Big Data nei loro processi di produzione si debbano organizzare: acquisendo nuove figure professionali come i *Data*

Scientist, far crescere il proprio personale in campi conoscitivi nuovi e su metodologie e tecnologie innovative, oltre che dotarsi di una organizzazione che copra diversi ambiti, tra cui quello della gestione della sicurezza informatica e della privacy dei dati. In generale si deve essere consapevoli che l'uso dei Big Data richiede grandi investimenti non solo in termini tecnici ma soprattutto di acquisizione di conoscenze e di risorse umane qualificate.

6.2.1 | NUOVI PROFILI CON UNA SINERGIA DI COMPETENZE STATISTICHE E INFORMATICHE

Il primo attore che interviene nel ciclo di vita dei Big Data si occupa dello studio e della "scoperta" delle informazioni che si potrebbero ottenere dai Big Data grezzi e dello sviluppo delle prime elaborazioni di essi. Per questo tipo di attore sono richieste diverse competenze che gli consentono di realizzare le funzioni del processo di "Development and Information Discovery". In particolare, sono in grado di definire algoritmi di elaborazione dei dati grezzi, necessarie per acquisire e utilizzare i Big Data in un processo statistico. Questo ruolo è etichettato nel processo come "Big Data Computing Design", ricoperto da figure denominate "Computing Architects", ossia da esperti di cloud computing, di gestione di storage di grandi volumi di dati e, più in generale, di esperti di architettura IT. Gli architetti di sistemi informatici coinvolti da subito nella fase di "acquisizione e registrazione", rappresentano un ruolo trasversale, che si occupa di curare i prerequisiti tecnici sia per le fasi principali di "Discovery" che di "Produzione".

Una volta che il framework informatico generale è stato impostato e fornito, in tutte le sue componenti sia hardware che software, un altro ruolo cruciale è quello dedicato alla "Preparazione dei Big Data", ossia gli "Information Architects". Questi attori sono principalmente dedicati alle fasi di trattamento dei dati ("Data Wrangling") e rappresentazione ("Data Representation"). Garantiscono i servizi di estrazione dei dati e di rappresentazione attraverso reportistica (ad esempio la distribuzione dell'intero set di dati e le diverse visualizzazioni delle analisi sui dati grezzi). Le principali attività di pre-trattamento garantite sono:

- Convalida dei dati (ad es. controlli del formato) e pulizia dei dati (ad es. eliminazione di record danneggiati, deduplicazione, selezione di quelli di interesse);

- Conversione dei dati (ad es. standardizzazione, riformattazione e incapsulamento);
- Aggregazione e sintesi dei dati.

Una volta che i dati sono disponibili e gestiti correttamente, altri profili come i *data scientist*, attraverso diversi metodi di "Analisi e visualizzazione", sono in grado di identificare il potenziale utilizzo di questi dati per le statistiche e costruire la pipeline statistiche di analisi. Queste attività sono realizzate nella fase di "Modellazione e interpretazione" e nella fase "Analisi visiva".

I diversi profili devono collaborare lungo tutta la pipeline, poiché il trattamento dei Big Data per la loro complessità richiede un pool di competenze, ad esempio la collaborazione dei *data scientist* e degli architetti dell'informazione sono essenziali nelle fasi "Data Wrangling" e "Data Representation". Compito dei *Data Scientist*, nell'ambito dell'analisi dei Big Data è essenziale e si muove su diverse dimensioni di analisi, attività e tecniche, come ad esempio:

- l'analisi "Human-in-the-loop" (ad es. scoperta di informazione, ipotesi di analisi o definizione del fenomeno, verifica dell'ipotesi);
- la scelta del metodo statistico per l'implementazione (es. Machine Learning, deep learning, elaborazione del linguaggio naturale, elaborazione delle immagini e reti neurali);
- lo sviluppo degli algoritmi e la loro ottimizzazione.

Quindi i *data scientist* dovrebbero concentrarsi sulla *scoperta del potenziale statistico* dei Big Data (ovvero eseguire rapidi cicli di test di ipotesi) per trovare il "valore informativo dei dati", combinando due approcci: *analisi e visualizzazione*.

La visualizzazione può servire a due scopi: da un lato aiuta a comprendere grandi volumi

di dati e a controllare rapidamente in che modo dovrebbero essere eseguite ulteriori esplorazioni; dall'altro aiuta a creare una rappresentazione semplificata dei risultati, atta ad aiutare una decisione o a comunicare le conoscenze acquisite (attraverso semplici visualizzazioni o infografiche).

Pertanto, i Data Scientist gestiscono anche alcune attività secondarie come:

- Visualizzazione dei dati esplorativi per la comprensione dei dati stessi (ad es. navigazione, rilevamento dei valori anomali, condizioni al contorno);
- Visualizzazione esplicativa dei risultati analitici (ad es. conferma, presentazione quasi in tempo reale delle analisi, interpretazione dei risultati analitici);
- Visualizzazione esplicativa per "raccontare la storia" (es. business intelligence).

Nelle fasi di scoperta e di validazione, la collaborazione tra *data scientist ed esperti di dominio* è essenziale per dare un senso ai risultati. Gli esperti di dominio dovrebbero essere coinvolti nelle prime fasi in cui si definiscono le esigenze di analisi e le definizioni dei concetti base del dominio preso in considerazione. Quindi gli esperti di dominio oltre ad intervenire nella definizione degli Use Case, debbono essere coinvolti nelle fasi di convalida, ossia nel valutare se i dati ottenuti possano realmente supportare l'analisi di interesse. Particolare rilievo assume l'attività di valutazione della rappresentatività dei dati nel dominio di analisi, sia per quanto riguarda la copertura (rappresentatività dei dati) della popolazione di riferimento che delle definizioni dei singoli concetti.

81

6.2.2 | FORNITORI DATI

Il fornitore di dati acquisisce i dati primari da fonti proprie o di terzi (ad es. browser Web, dispositivi mobili, sensori). Si occupa della persistenza dei dati, spesso accessibili tramite meccanismi come i servizi web. Inoltre, il fornitore, in funzione del titolo di possesso dei dati o di gestione di questi, ne definisce la politica di accesso, nonché il livello di dettaglio dei dati messi a disposizione.

Spesso il fornitore non è il *proprietario* dei dati ma ne è il *custode*, e gestisce i livelli di accesso in funzione dei livelli di privacy che ha dichiarato di garantire e anche in funzione delle autorizzazioni ricevute dal cittadino se i dati sono personali.

I provider possono essere: imprese, ad esem-

pio società web, operatori di rete o agenzie che forniscono servizi di estrazione dati da sistemi di sensori, ma potrebbero anche essere agenzie pubbliche.

In generale, è necessario definire un "Framework Provider" in cui sono configurate le risorse o i servizi di accesso ai dati, ma anche in alcuni casi i servizi di creazione dell'applicazione specifica di estrazione dei dati. Possiamo dire che "Framework Provider" gestisce tre attività: *framework di infrastruttura*, *framework di piattaforma dati* e *framework di accesso e elaborazione*, ciascuna attività in relazione con gli attori del processo descritti in precedenza.

6.2.3 | ATTORI DEL PROCESSO INTERNI ED ESTERNI ALL'ORGANIZZAZIONE

Per integrare le funzioni *interne*, di chi è interessato a costruire prodotti statistici, ed *esterne* ossia di quanti esterni all'organizzazione

produttrice degli output statistici sono chiamati a partecipare al processo, è necessaria una funzione globale, denominata nel BREAL

come *System Orchestrator*. Questa funzione è dedicata a fornire i *requisiti generali* che il sistema deve soddisfare, inclusi politica, governance, architettura, risorse, nonché monitoraggio o auditing per garantire che il sistema sia conforme a tali requisiti. In generale, è un pool di attori che sovrintende al contesto aziendale in cui opera il sistema, compresa la specificazione degli obiettivi aziendali, i contratti del Data Provider, la conduzione delle negoziazioni con il Framework Provider e la progettazione del piano di assunzione delle

risorse umane (bisogni interni in termini di figure professionali come: architetti informatici, architetti dell'informazione e data scientist). È necessario che questo insieme di figure lavorino in sinergia nell'ambito dell'azienda per gestire e formalizzare i bisogni, garantire la disponibilità di risorse generali e anche di collegare il processo dei Big Data a funzioni generiche come "Supporto alla produzione" e "Diffusione", che non differiscono sostanzialmente da quelle svolte per altri tipi di dati che non siano i Big Data.

6.2.4 | ATTORI DELL'ORGANIZZAZIONE CON FUNZIONI SPECIFICHE

82

I processi di produzione che utilizzano i Big Data, come tutti gli altri hanno bisogno di implementare e specificare tre ulteriori funzioni trasversali: Gestione generale dei dati, Sicurezza e privacy, Rilevanza scientifica dei dati. L'uso dei Big Data cambia l'approccio: i dati vengono ora raccolti per puntare a un obiettivo. Pertanto, i dati raccolti devono essere coerenti tra loro. Questo è il compito del *Chief Data Officer* (abbreviato CDO), il quale si occupa di semplificare il modo di accedere ai dati e di individuare i dati utili agli scopi conoscitivi dell'organizzazione.

In particolare, nella funzione del ciclo di vita dei Big Data "*Esplorazione di nuove fonti di dati*", è svolta dal *Chief Data Officer* che è un attore interno all'organizzazione ed è il responsabile della gestione complessiva dei dati.

In futuro, questa funzione dovrà rispondere sempre più prontamente alle esigenze degli utenti, occupandosi di identificare, studiare e analizzare nuove fonti di dati con il maggiore potenziale informativo rispetto le esigenze

di analisi di interesse. Una delle caratteristiche dei Big Data è la tempestività, ossia la capacità di rappresentare eventi e di fornire informazioni su questi in tempi brevissimi. Poiché molte delle fonti di questo tipo possono riguardare dati personali o persino dati sensibili, c'è bisogno di coinvolgere nel processo anche il *Data Protection Officer* (abbreviato DPO) che si occupa di costruire un insieme di misure di salvaguardia per tutte le tipologie di piattaforme Big Data, come ad esempio i controlli di sicurezza e privacy per proteggere sia le operazioni e le risorse critiche ed essenziali delle organizzazioni, che la riservatezza delle informazioni individuali. L'obiettivo ultimo è verificare che il sistema informativo soddisfi i requisiti di sicurezza e privacy definiti dalle leggi, nonché le misure complementari decise dal *System Orchestrator*.

Inoltre, le comunità scientifiche e accademiche possono contribuire al processo per migliorarlo e arricchirlo di tutte le tecniche e metodologie più avanzate applicabili a queste fonti di dati in continua evoluzione.

6.3 | LA STANDARDIZZAZIONE DELL'ACCESSO AI DATI

Molti Big Data sono detenuti da soggetti privati e sono pertanto denominati *privately-held data* (PHD). Per i PHD è estremamente importante affrontare il tema dell'accesso in tutte le sue dimensioni: legale, organizzativa e tecnica. Da un punto di vista *legale*, la strategia europea dei dati, ed in particolare il Data Act⁹, [6.7] rappresenta certamente un grosso passo in avanti per la regolamentazione delle relazioni business-to-government e l'accesso quindi ai *privately-held data*. Il Data Act sottolinea l'importanza di principi come la condivisione obbligatoria dei dati per alcune finalità pubbliche, naturalmente nel rispetto della privacy e di altri diritti sui dati, e anche la standardizzazione delle modalità di accesso ai dati, garantendo condizioni eque, ragionevoli, proporzionate, trasparenti e non discriminatorie.

Da un punto di vista *organizzativo*, in molti casi, i dati potrebbero risiedere dal detentore (holder) o dal fornitore (provider) degli stessi. Questo comporta la necessità di mettere in piedi dei processi di accesso ai dati che possono essere di elevata complessità. Si tratta di un vero e proprio cambio di paradigma¹⁰ passando da un approccio in cui i dati sono portati all'interno di una organizzazione (*pulling data in*) ad un approccio in cui i dati risiedono presso il detentore/fornitore ed è, invece, l'elaborazione che viene spinta fuori dall'organizzazione (*pushing computation out*). L'impatto organizzativo di tale cambio di paradigma di accesso ai dati è notevole e, oltre a prevedere la citata progettazione di nuovi processi, include anche un investimento sulla creazione di una nuova cultura del dato, che

non è più in possesso dell'organizzazione che produce un servizio a valore aggiunto a partire da questo, ma può essere reso disponibile in diversi punti di una *supply chain* che coinvolge tanti attori.

Infine, da un punto di vista *tecnico*, è necessario investire in sistemi dedicati che consentano di effettuare

- i. accessi secondo modalità definite dal detentore/fornitore oppure
- ii. processamento di dati presso la sede del detentore/fornitore.

Nel primo caso, laddove non ci siano scambi dati dedicati, un'importante modalità di accesso è tramite le API (*Application Programming Interface*) messe a disposizione dai detentori/fornitori; un esempio è la public API di Twitter. Nel secondo caso, invece, sono di aiuto le cosiddette *privacy-enhancing technologies* (PET) che includono: tecniche di *homomorphic encryption*, *Secure Multy-party Computation* (SMC), *Trusted Execution Environment* (TEE), generazione di dati sintetici. Mediante le PET, è possibile fare elaborazioni che garantiscano la riservatezza (*privacy preserving*) presso il detentore/fornitore, senza vedere in chiaro i dati ma soltanto il risultato dell'elaborazione. Queste tecniche rappresentano dunque un importante elemento di realizzazione del citato approccio di "*pushing computation out*". In sintesi, cominciano ad essere disponibili soluzioni legislative, organizzative e tecniche che realizzano l'accesso ai *privately-held data*. Sono però ancora aperti molti fronti di investimento, il più importante dei quali, è, probabilmente, quello di regolamentazione nell'ambito della strategia nazionale italiana.

9. European Commission (23 February 2022). Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data). Brussels, Belgium: European Commission, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

10. <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji190584>

6.4 | BIBLIOGRAFIA

- [6.1] ESSnet Big Data Pilots II: https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page
- [6.2] Eurostat: ESS Enterprise Architecture Reference Framework: https://ec.europa.eu/eurostat/cros/content/WPF_Process_and_architecture_en
- [6.3] Deliverable F1 "BREAL: Big Data REference Architecture and Layers Business Layer": https://ec.europa.eu/eurostat/cros/sites/default/files/WPF_Deliverable_F1_BREAL_Big_Data_REference_Architecture_and_Layers_v.03012020.pdf
- [6.4] Statistical Enterprise Architecture: <https://joinup.ec.europa.eu/collection/statistical-enterprise-architecture>
- [6.5] Deliverable I.6 "A proposal for a statistical production process with Mobile Network Data": https://cros-legacy.ec.europa.eu/system/files/wpi_deliverable_i6_a_proposal_for_a_statistical_production_process_with_mobile_network_data_18_03_2021_final.pdf
- [6.6] "Towards a Reference Methodological Framework for the processing of mobile network operator data for official statistics", Presentato a "Mobile Tartu 2018". https://ec.europa.eu/eurostat/cros/system/files/rmf_mobiletartu2018_ricciato_printout.pdf
- [6.7] European Commission (23 February 2022). Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data. Brussels, Belgium: European Commission, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113



7 | L'ESTRAZIONE DELL'INFORMAZIONE

Lorenzo Vannacci¹[0000-0001-9587-7611], Martina Farsi¹[0000-0002-3132-8071], Mauro Capurso²[0000-0001-7599-4431], Massimo Cerquenich², Elia Pianelli²

1. FS Research Centre, Firenze, Italia

2. Ferrovie dello Stato Italiane, Trenitalia, Strategie, Roma, Italia

7.1 | ANALISI SPERIMENTALE

7.1.1 | I DATI UTILIZZATI

I dati di telefonia mobile utilizzati per il caso studio descritto in questo capitolo sono stati forniti da un unico operatore telefonico che conta sul territorio italiano circa 23 milioni di SIM tipo Human¹¹.

A livello spaziale, la precisione di georeferenziazione dei dati raccolti dipende in primo luogo dall'infrastruttura di diffusione del segnale radiomobile; la **granularità di base** è legata al numero di celle telefoniche ed alla loro distribuzione. L'operatore in questione è presente sul territorio nazionale con oltre 200.000 celle telefoniche che coprono, con il segnale 4G, circa il 99% della popolazione. La **granularità spaziale avanzata** è invece legata ad ulteriori dispositivi affiancati all'infrastruttura di base; sono ascrivibili a questa seconda tipologia i ripetitori di segnale 4G e 4.5G, dette Crowdcell, e i Distributed Antenna System (DAS), tecnologia ancora in fase di consolidamento, costituita da una serie di piccole antenne generalmente installate all'interno dei luoghi di interesse allo scopo di migliorare la copertura del servizio radiomobile. Un ulteriore miglioramento della precisione spaziale delle misure rispetto alle coperture delle celle telefoniche viene ottenuto accedendo a sorgenti dati e metodologie complementari di maggiore dettaglio. Tra queste, figurano sorgenti basate su dati generati dalle applicazioni presenti nei dispositivi mobili, che forniscono una geo-localizzazione con

accuratezza a pochi metri (AGPS), a differenza delle antenne telefoniche e relative coperture che, invece, sono basate su aree; nel caso studio in questione circa dieci milioni di dispositivi permettono questo tipo di **taratura spaziale addizionale**. A partire da queste fonti, sempre opportunamente aggregate, è quindi possibile costruire mappe probabilistiche che permettono di distribuire statisticamente gli utenti con effettivo aumento di dettaglio spaziale sul territorio rispetto all'analisi basata unicamente su celle telefoniche.

L'altro aspetto fondamentale rispetto all'accuratezza e alla capacità descrittiva del fenomeno della mobilità è la **granularità temporale** dei rilievi: la frequenza con cui vengono campionate le posizioni delle SIM dell'operatore è di fondamentale importanza per abilitare il processo di profilazioni successive ed analisi accurate. Questo aspetto è decisivo soprattutto nei casi in cui si renda necessario conoscere l'effettiva presenza o passaggio in aree geografiche limitate, come stazioni ferroviarie, caselli autostradali, valichi di frontiera, etc... Un campionamento rado, ad esempio ogni trenta minuti o più, non permetterebbe di garantire l'effettiva presenza dei Clienti in zone geografiche limitate e, quindi, diminuirebbe drasticamente l'attendibilità delle profilazioni e della maggior parte degli insight. La registrazione temporale degli eventi avviene tramite due di-

11. SIM Human sono quelle SIM che permettono traffico Voce o Voce + Dati

stinti processi di acquisizione: il primo è quello relativo al traffico Voce e SMS, detto Circuit Switch (CSD), che mediamente produce circa dieci rilevazioni/giorno per SIM/dispositivo, l'altro è legato invece alla rete Packet Switch (PS o TMA) relativa al traffico dati e all'utilizzo delle APP e di altre attività dei dispositivi. Questa seconda fonte permette una frequenza ben sotto il minuto, circa duemila rilevazioni al giorno per SIM/dispositivo.

In questo contesto si arrivano a produrre fino a trenta miliardi di posizioni referenziate nel tempo e nello spazio in un giorno, anonimizzate in maniera irreversibile.

Come per ogni indagine statistica devono essere implementate tecniche e metodologie per azzerare o limitare gli eventuali errori. In particolare, tutte le indagini campionarie risentono di due tipologie di errore:

- Errore Campionario che deriva semplicemente dal fatto che stiamo osservando soltanto una parte della popolazione;
- Errori non Campionari che dipendono dai possibili "bias" relativi all'ambiente di indagine. Questi si possono diminuire o eliminare se si conoscono informazioni sulla loro natura.

Per quanto riguarda l'errore campionario, questo può essere ritenuto irrilevante a fronte della grande dimensione del campione oggetto di indagine e della frequenza di rilevazione delle informazioni geografiche (più volte al minuto). D'altra parte, l'errore non campionario necessita di essere costantemente verificato e ridotto grazie alle informazioni che si riscontrano sul campo.

I dati raccolti, attraverso un insieme opportuno di algoritmi proprietari e tecniche di espansione, vengono riportati all'intera popolazione italiana e straniera presente sul territorio italiano. Il processo per espandere il campione al totale della popolazione di riferimento (tutte le persone con un telefono, che corrispondono approssimativamente a chi ha almeno dodici anni di età) è basato su un modello di Machine Learning che considera un insieme di variabili, le principali delle quali sono:

- Market share locale dell'operatore ottenuto da analisi di mercato interne all'azienda;
- Market share per tipologia di SIM (business/consumer) ottenuto da studi di mercato e report ufficiali, quali ad esempio l'osservatorio sulle telecomunicazioni o da AGCOM;
- Caratteristiche sociodemografiche degli utenti dalle anagrafiche proprietarie. Su questi indicatori vengono applicati due livelli di raffinamento del dato:
 - Utilizzo delle informazioni del "reale utilizzatore": questo avviene attraverso i punti di interfaccia dell'operatore con il cliente (contact-center, shop, ecc) attraverso i quali vengono aggiornate le informazioni anagrafiche del reale utilizzatore della SIM a seguito dell'interazione con i clienti stessi. Il reale utilizzatore è la persona che effettivamente utilizza la SIM a prescindere da chi ha sottoscritto il contratto;
 - Correzione di eventuali distorsioni legate alla differente penetrazione di mercato fra le diverse fasce d'età della popolazione utilizzando statistiche ufficiali quali i dati di censimento ISTAT.

Le combinazioni dei fattori di calibrazione hanno portato ad individuare diverse classi di algoritmi inferenziali a seconda delle tipologie di analisi, creando in questo modo una libreria suddivisa per ambiti ben specifici, quali, ad esempio, la presenza indoor, la modalità ad alta velocità, la presenza territoriale, il passaggio per luoghi circoscritti. Particolare attenzione è stata posta nella metodologia di verifica e taratura continua del modello inferenziale. Infatti, vengono analizzate a cadenza periodica le situazioni in cui è disponibile un riferimento numerico sicuro di confronto.

È disponibile uno storico dei diciotto mesi passati e questo permette di poter confrontare la fenomenologia, considerando un completo ciclo di stagionalità, ed offre la possibilità di misurare trend e differenziali anno su anno.

7.1.2 | INTERAZIONE PER L'ESTRAZIONE DEL DATO

7.1.2.1 | ESIGENZE LEGATE ALLE ANALISI DI MOBILITÀ

I dati necessari per le analisi e studi di mobilità possono essere suddivisi in dati che descrivono l'offerta di mobilità, dati socioeconomici e dati che descrivono le abitudini di mobilità degli utenti.

I dati relativi all'offerta sono generalmente dati già in possesso delle amministrazioni pubbliche e dei gestori dei servizi, come per esempio quelli che riguardano la rete infrastrutturale e l'offerta programmata dei servizi (GTFS). I dati socioeconomici derivano dalle raccolte statistiche nazionali, come il censimento della popolazione, o da altre indagini specifiche. In questo ambito vengono raccolte anche informazioni relative alle abitudini di mobilità, di grande interesse per lo sviluppo di analisi e modelli di simulazione dei trasporti. Sebbene l'indagine sia una buona base di partenza, non si riescono a soddisfare, anche a causa della limitata granularità territoriale con cui si riesce a rappresentare il fenomeno, il fabbisogno informativo desiderato. Ne segue la necessità di intercettare o creare ulteriori fonti, che prima dell'arrivo dei Big Data, erano costituite unicamente da indagini progettate ad hoc per gli studi da svolgere, ad esempio le rilevazioni dei flussi veicolari in determinate sezioni o quelle dei passeggeri di servizi pubblici di interesse specifico.

Per le amministrazioni territoriali, i dati relativi alle abitudini di mobilità sono alla base di una buona pianificazione degli interventi infrastrutturali e delle politiche gestionali. Dal punto di vista progettuale, tali dati sono necessari per la realizzazione dei modelli econometrici, che permettono la valutazione della domanda di trasporto al variare dell'offerta. Essi entrano in gioco sia in fase di costruzione del modello, che nella fase di valutazione della capacità del modello di rappresentare le dinamiche reali di mobilità (validazione). Dal punto di vista gestionale, permettono di monitorare l'efficacia di una determinata politica

relativa al settore trasporti, per esempio l'istituzione di una *congestion charge*, l'applicazione di una nuova tariffa dei titoli di viaggio o di un insieme di azioni, come è richiesto ad esempio per i PUMS.

Per il gestore del trasporto, i dati di mobilità permettono la definizione della desiderata della domanda di trasporto al fine di implementare un'offerta adeguata e possibilmente ampliare il bacino di utenza dei servizi. I dati di offerta e di domanda descritti, diventano imprescindibili, se si progettano sistemi di Mobility as a service (MaaS): al fine di creare un sistema integrato di trasporti è necessaria la condivisione e la raccolta di tutte le informazioni, sia di natura statica (origine, destinazione, fasce orarie di picco, etc.) che dinamica (stato della rete e dei servizi in tempo reale, tempi di percorrenza stimati, etc.). I dati di cui si necessita nel campo dei trasporti sono quindi di varia natura. Rimanendo nell'ambito dei dati descrittivi della mobilità delle persone, le informazioni necessarie sono:

- Organizzazione spaziale degli spostamenti. Origine e destinazione degli spostamenti e numero di persone/veicoli per ciascuna coppia OD;
- Distribuzione temporale all'interno della giornata degli spostamenti;
- Frequenza dello spostamento;
- Motivo dello spostamento, lavoro, studio, affari, occasionale da cui può dipendere anche la disponibilità a pagare per effettuare lo spostamento;
- Tipologia di utente (età, occupazione, etc.);
- Mezzo con cui viene effettuato lo spostamento.

In funzione dell'analisi da svolgere, altre informazioni interessanti sono i dati relativi all'utilizzo di terminali a supporto dello spostamento, come stazioni, fermate e banchine,

come avviene negli studi relativi ai livelli di servizio.

In concomitanza della pandemia, è apparsa poi più evidente la necessità di avere istantanee della mobilità in un determinato momento, che permettano rapidi confronti con le condizioni standard; questo elemento è peraltro di interesse anche tutte le volte che si vogliono monitorare gli effetti di azioni e politiche gestionali nuove nell'ambito dei trasporti, in accordo con la filosofia dei Piani Urbani della Mobilità Sostenibile. MaaS e Smart Mobility, inoltre, non possono prescindere dalla conoscenza in tempo reale delle condizioni di deflusso della rete, delle perturbazioni e dei tempi stimati di percorrenza.

Per poter essere utilizzati come input o set

di confronto per i modelli di simulazione o permettere di definire una serie storica di analisi, i campioni utilizzati per la raccolta dei dati di mobilità devono poter essere ripetibili con caratteristiche simili, o riconducibili alle precedenti raccolte dati. Ne è un esempio l'attribuzione spaziale degli spostamenti; nei modelli di trasporto si definiscono unità di base univocamente definite e che hanno una variabilità limitata quali sezioni di censimento ed unità amministrative (Comuni, Province, etc.).

Soprattutto nell'ambito della modellistica dei trasporti, gli individui inclusi nei campioni devono essere classificabili da un punto di vista socioeconomico, per permettere l'espansione all'universo e l'analisi dei risultati.

7.1.2.2 | CREAZIONE DI UN LINGUAGGIO COMUNE FORNITORE-ACQUIRENTE FUNZIONALE ALL'ESTRAZIONE DEI DATI DI MOBILITÀ

L'utilizzo dei dati telefonici al fine di rispondere alle esigenze informative sulla mobilità necessita di trattamenti preliminari: i Big Data infatti forniscono grandi moli di informazioni, che devono però essere "tradotte" negli indicatori tipici delle indagini e degli studi in ambito trasportistico. Per questo motivo, nei casi di studio e nelle sperimentazioni applicative, è necessaria un'interazione fra utilizzatore del dato e fornitore, per individuare e definire le grandezze che si vogliono ricavare dai dati in questione, affinando le metodologie di estrazione delle stesse testandone i risultati.

Il primo contesto di utilizzo concreto dei dati di telefonia che si presenta, è relativo a uno studio a livello nazionale della mobilità, integrativo alle fonti dati ufficiali, quali le indagini ISTAT, il Conto Nazionale dei Trasporti del MIT e l'indagine Audimob. In funzione di questo tipo di analisi, sono state concordate le caratteristiche della fornitura dati con la società incaricata delle elaborazioni.

Il primo passo è stato definire l'entità "spostamento": per lo studio applicativo qui

descritto, per "spostamento" si è inteso la concatenazione di tutte le eventuali soste intermedie se di durata inferiore a un'ora all'interno di un unico viaggio origine-destinazione. Ad esempio, la catena casa-sosta per fare benzina e poi andare al lavoro è stata considerata come un unico spostamento. La **distanza** associata allo spostamento è stata stimata come somma delle distanze percorse all'interno del viaggio.

Per il riferimento spaziale degli spostamenti è stata elaborata un'apposita zonizzazione: la suddivisione dell'area di studio, il territorio nazionale, è stata sviluppata a partire dai comuni e dalla zonizzazione ANAS. Quest'ultima è stata definita internamente ad ANAS ai fini della costruzione di un modello di trasporto, sempre su scala nazionale, per la previsione dei flussi di traffico sulla rete ANAS e su parte della restante rete nazionale. A sua volta la zonizzazione ANAS è coerente con le zone SIMPT del MIT (anno 2004) e con i confini della suddivisione del territorio nazionale in Sistemi Locali del Lavoro (SLL – anno 2011).

Utilizzando questi riferimenti, si è proceduto quindi all'aggregazione di piccoli comuni, mentre per i comuni di grandi dimensioni in area metropolitana è stata adottata una

zonizzazione più disaggregata sulla base di quartieri, circoscrizioni o loro aggregazioni (Figura 12). È quindi stata definita una zonizzazione composta da 3000 zone.

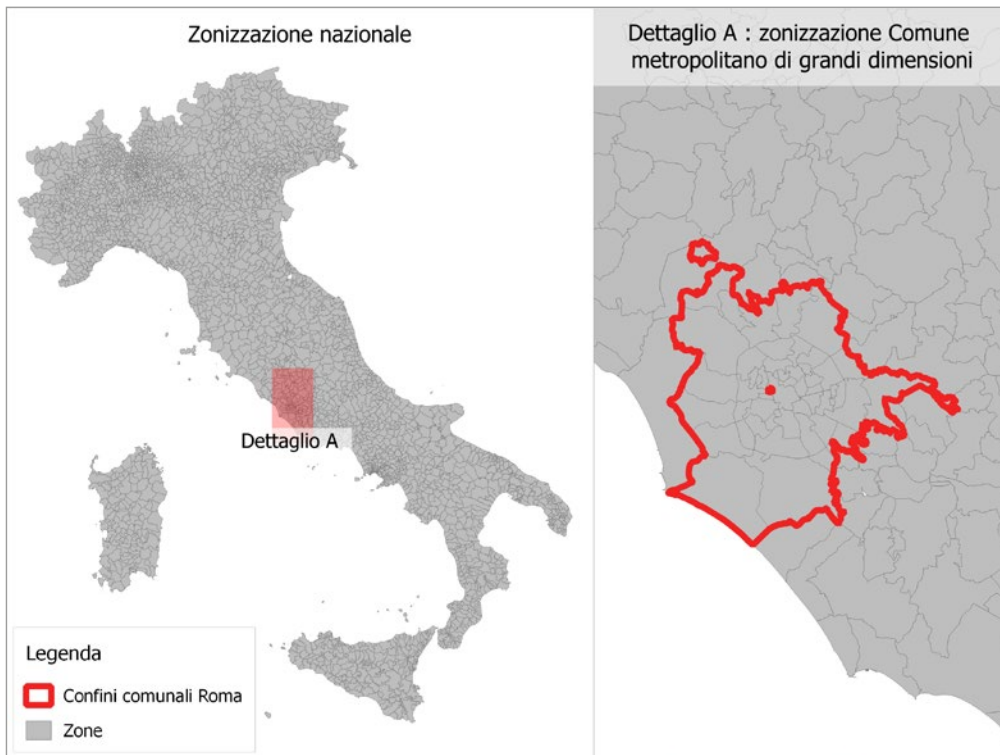


Figura 12 - Zonizzazione alla base degli spostamenti da dati telefonici

I principali dati estratti da questo tipo di analisi sono:

- La ripartizione modale per i modi treno, aereo, altro;
- la distribuzione degli utenti dei modi treno, aereo, altro rispetto alle distanze percorse;
- la distribuzione mensile e settimanale e oraria degli spostamenti;
- la sistematicità o meno degli spostamenti;
- le percentuali di popolazione mobile per regione;
- la distribuzione spaziale (O/D) e modale per gli spostamenti a livello provinciale e nel dettaglio per alcune città metropolitane.

L'altro tipo di utilizzo dei dati telefonici è stato quello relativo alla definizione delle presenze all'interno di quattro stazioni ferroviarie principali: Roma Termini, Milano Centrale, Napoli Centrale, Genova Piazza Principe. In particolare, è stato indagato il comportamento giornaliero degli utenti a seconda delle attività svolte, del giorno della settimana e del mezzo di arrivo e di partenza dalla stazione. Per ognuna delle quattro stazioni sono analizzate anche alcune destinazioni specifiche come gli aeroporti. Per questo tipo di attività sono state sviluppate le modalità di individuazione delle tipologie di utenze:

- Modalità treno AV: l'utente classificato come passeggero treno, è stato visto in

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

un'altra stazione collegata dal servizio alta velocità, in un intervallo di tempo coerente con il tempo di viaggio in alta velocità;

- Modalità Aeroporto Express: l'utente classificato come passeggero treno origina o termina il viaggio in treno all'aeroporto di Fiumicino per Roma o Malpensa per Milano;
- Modalità treno REG/IC: l'utente classificato come passeggero treno non rientra

nelle classificazioni precedenti;

- Modalità altro: l'utente non è classificato come passeggero treno;
- Co-visita: l'utente oltre alla stazione visita anche altri punti di interesse.

Da questo tipo di analisi dei dati telefonici è stato possibile estrarre le presenze totali mensili, le presenze medie per giorno della settimana, la distinzione delle tipologie di utenze sulla base delle definizioni suddette.

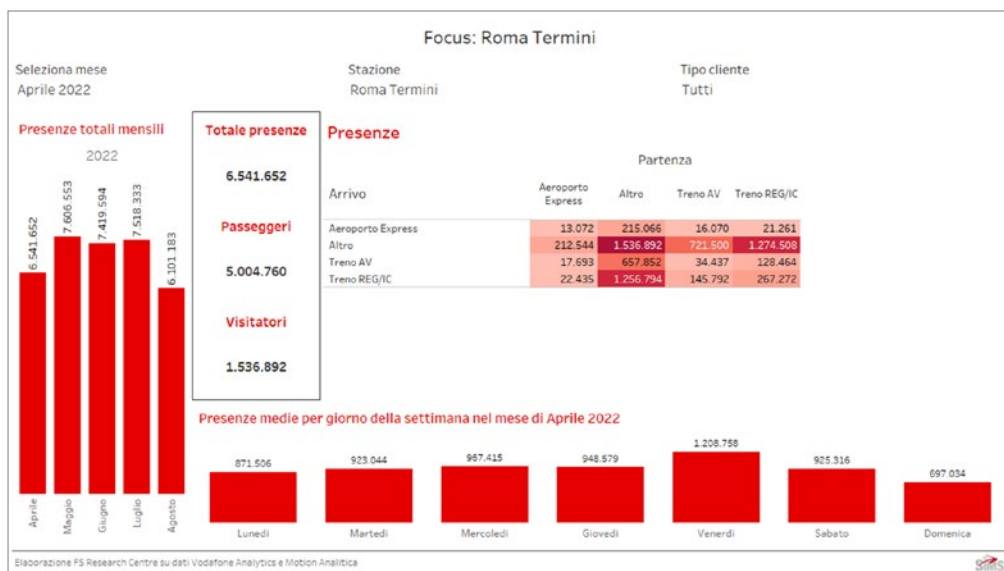


Figura 13 - Esempio di analisi delle presenze in stazione attraverso i dati di telefonia- fonte SIMS

Sempre riferendosi alle stazioni, un altro tipo di impiego è stato quello relativo all'analisi della loro *catchment area*, sia in origine (da dove provengono gli utenti che utilizzano la specifica stazione per accedere alla rete ferroviaria), sia in destinazione (dove vanno successivamente gli utenti che sono arrivati in treno nella stazione di studio); gli spostamenti per ogni stazione studiata sono stati aggregati secondo la zona di origine/destinazione, il modo di accesso/egresso alla stazione, il mese e la sistematicità dello spostamento *Figura 14*. Secondo questa definizione, in analogia con quanto avviene con il calcolo

delle isocrone di accesso/egresso verso gli hub di trasporto, la *catchment area* non è stata ricondotta ad una specifica O/D percorsa dal viaggiatore, ma alle singole stazioni. A titolo esemplificativo, la città di Pavia è stata genericamente considerata all'interno della *catchment area* della stazione di Genova Piazza Principe; tuttavia, mentre questa definizione potrebbe risultare appropriata con riferimento ai servizi Genova-Roma, potrebbe non esserlo quando si analizzano i servizi Genova-Milano, essendo la stazione di Pavia essa stessa fermata per numerosi servizi sulla relazione Genova-Milano.

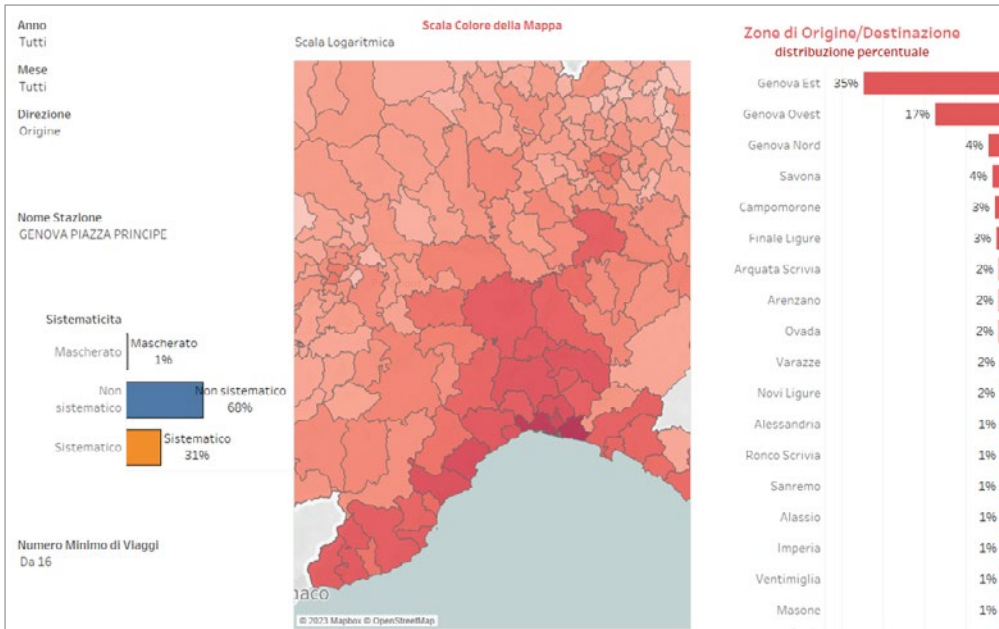


Figura 14 - Esempio di catchment area per la stazione di Genova Piazza Principe

Come evidenziato nel capitolo 5 "Analisi di alcuni casi studio relativi alla mobilità delle persone", i Big Data e nel caso particolare i dati di telefonia mobile, permettono analisi su lunghi periodi temporali e l'individuazione della variabilità temporale dei fenomeni di mobilità. L'espansione temporale estesa ha permesso per esempio di fotografare l'andamento della mobilità in periodo pre-covid e post-covid, con la copertura totale del territorio italiano, analisi impensabile disponendo delle indagini tradizionali anche solo

per l'eccezionalità e l'imprevedibilità dell'evento pandemico. La disponibilità di questi dati in quasi real time e per tutto il territorio nazionale, in aggiunta a fonti più tradizionali come survey sulle intenzioni di viaggio, ha permesso anche una adeguata proposizione di servizi di trasporto nel periodo pandemico, soprattutto di quelli a lunga distanza che sono stati drasticamente ridotti e successivamente gradualmente ripristinati con la ripresa degli spostamenti.

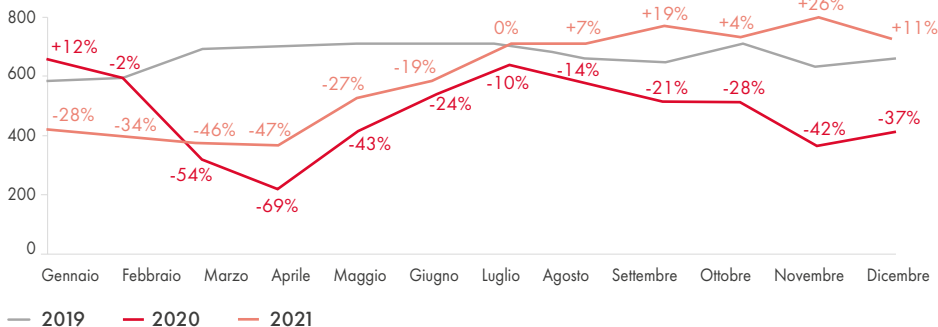


Figura 15 - Andamento della mobilità in Italia in tre anni differenti ,pre e post pandemia

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

Un altro esempio di questo tipo di utilizzo è rappresentato dallo studio effettuato sugli spostamenti relativi ad alcune coppie origine-destinazione di interesse; per due mesi significativi di periodi diversi, invernale ed estivo, sono stati estratti gli spostamenti medi per giorno della settimana, come si può

vedere dalla Figura 16. Questa tipologia di analisi è funzionale alla rapida valutazione della domanda potenziale di trasporto su una O/D di interesse e la sua distribuzione stagionale, tra i giorni della settimana e fascia oraria (Figura 17).

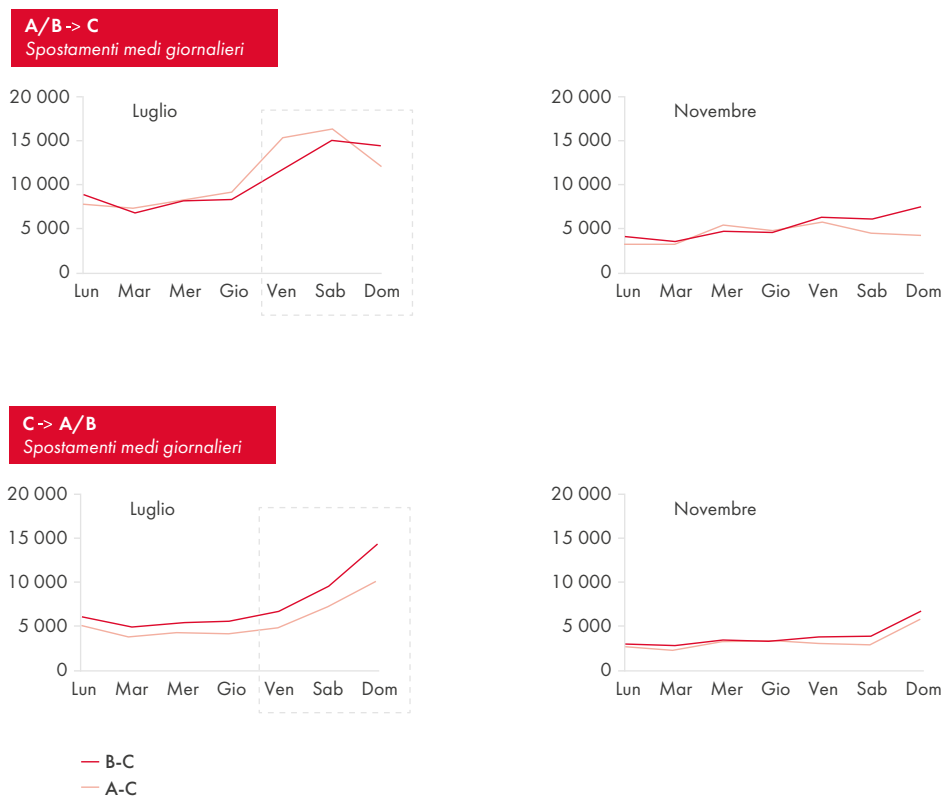


Figura 16 - Esempio di caratterizzazione temporale degli spostamenti su OD di interesse

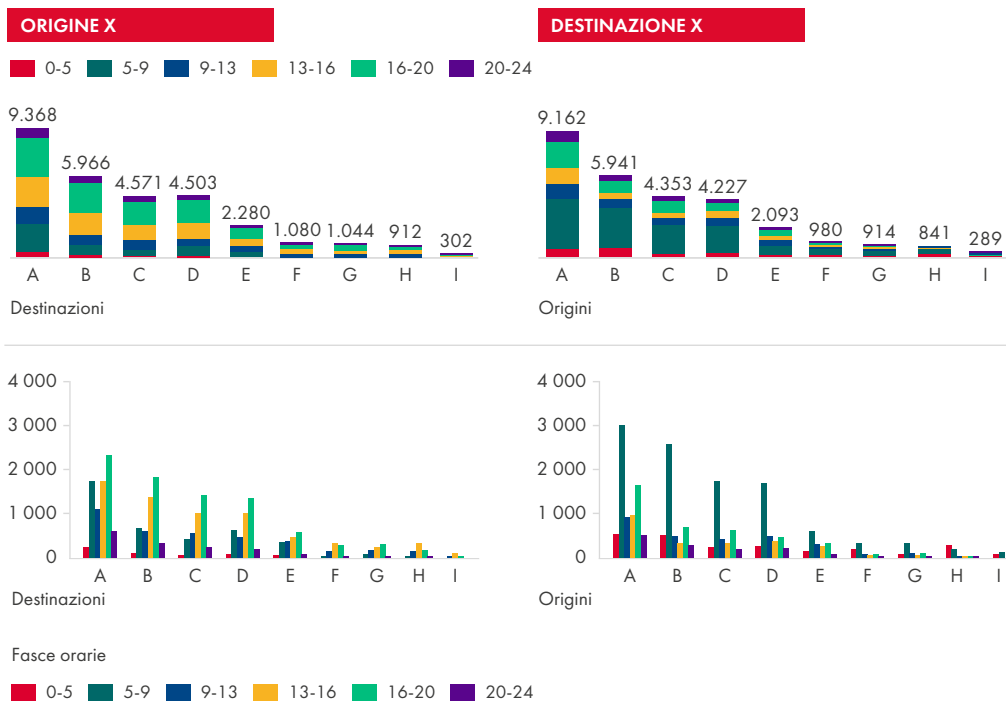


Figura 17 - Esempio di analisi di distribuzione oraria degli spostamenti

Grazie alla possibilità di riconoscere i dispositivi mobili in roaming, sono state effettuate analisi per origine/destinazione disaggregate

per utenti residenti in Italia e stranieri come quelli riportati in Figura 18.

Spostamenti per mese
(K spostamenti)

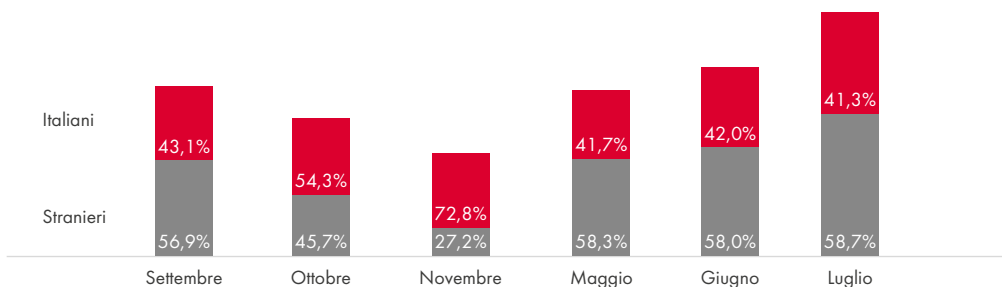


Figura 18 - Disaggregazione degli spostamenti fra SIM italiane e straniere

7.1.3 | CRITICITÀ EMERSE E CRITICITÀ SUPERATE

Nell'utilizzo dei Big Data telefonici per le analisi di mobilità si cerca di soddisfare in primo luogo la quantificazione degli elementi fondamentali di un'indagine o di uno studio trasportistico: origine e destinazione, modo e motivo dello spostamento, distribuzione oraria giornaliera, distribuzione stagionale. Rispetto alle indagini convenzionali, quali interviste telefoniche o via web, indagini al cordone, conteggi e quanto altro, l'utilizzo dei Big Data genera campioni di dimensioni superiori di qualche ordine di grandezza, il che farebbe pensare a un'indubbia convenienza di questa scelta. Per contro, si rende necessario affrontare alcuni limiti legati all'interpretazione del dato grezzo, secondo un processo che non permette l'analisi del singolo caso, come può

succedere in indagini di dimensioni esigue, ma necessariamente deve essere standardizzata secondo un algoritmo generale per far fronte alle dimensioni campionarie.

Il primo elemento di "taratura" nel passaggio dal dato grezzo all'informazione opportunamente elaborata per uso trasportistico, è legato alla definizione stessa di spostamento. Si è reso necessario trascurare gli spostamenti inferiori agli 800 metri, per eliminare alcuni problemi legati ad elementi tecnici di generazione dei segnali, ad esempio i fenomeni di "rimbalzo" tra le celle telefoniche: un utente fermo può infatti essere agganciato in maniera alternata a celle telefoniche diverse che si sovrappongono sull'area in cui staziona l'utente in questione.

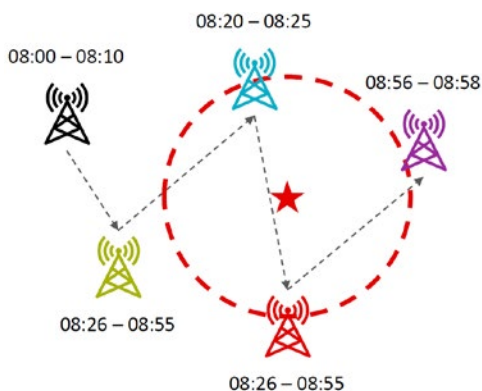


Figura 19 - Definizione del raggio di mobilità

Per ovviare a questo problema, è stata calcolata la media pesata (sul tempo di permanenza) delle distanze delle celle telefoniche visitate; se tale media è per più di un'ora consecutiva inferiore a 800 metri l'utente è considerato come stazionario, viceversa in movimento. Per gli utenti stazionari, è inoltre definita anche la "cella prevalente", ovvero quella dove l'utente è agganciato per la maggior parte del tempo e nella quale si assume localizzata la sua posizione. Questo tipo di



problematica è particolarmente importante in ambito urbano, dove potrebbero andare perse informazioni relative agli spostamenti di piccola entità.

La seconda problematica emersa è quella legata alla definizione del modo di trasporto utilizzato per lo spostamento. Si possono distinguere due ambiti, urbano ed extraurbano. Nel primo, ad oggi, non si è ancora definita una metodologia che permetta di distinguere quale sia con esattezza il mezzo utilizzato,



in quanto in ambito urbano mezzi diversi possono condividere oltre ai percorsi anche velocità di percorrenza analoghe a causa dei fenomeni di congestione; spostamenti con autobus urbano possono essere confusi con spostamenti effettuati in bici e viceversa, per esempio. In ambito extraurbano per il caso oggetto di studio, attraverso la collaborazione fra analisti ed utilizzatori finali del dato telefonico **è stato sviluppato un algoritmo di definizione del modo**. I modi individuati nello specifico studio sono:

- aereo;
- treno;
- altro.

La modalità **"aereo"** viene assegnata a tutti i viaggiatori che compiono un viaggio origine-destinazione, effettuando una partenza da un aeroporto dopo essere stato rilevato nella zona di origine, ma prima di raggiungere la destinazione e, inoltre, effettua un arrivo in aeroporto dopo aver effettuato la partenza e prima di raggiungere la destinazione. Al fine di determinare partenze e arrivi si utilizza la metodologia del "teletrasporto", cioè si analizza se l'utente viene registrato in una delle celle associate all'aeroporto di partenza e, dopo un tempo congruo, viene registrato nelle celle associate all'aeroporto di arrivo senza mai essere stato rilevato dalla rete telefonica fra questi due eventi.

La modalità **"treno"** viene assegnata a tutti i viaggiatori che compiono un viaggio origine-destinazione e che abbiano effettuato un viaggio in treno che cominci dopo aver

lasciato l'origine e prima di aver raggiunto la destinazione. I viaggi in treno vengono individuati analizzando utenti che iniziano e terminano il viaggio in una stazione monitorata. L'attribuzione del modo treno si articola in fasi successive:

1. Il primo passo della metodologia prevede di associare ad ognuna delle stazioni monitorate una serie di stazioni radio base di riferimento. Tale operazione permette di individuare gli utenti che si connettono alle celle associate all'area della stazione.
2. Il secondo passo prevede di individuare una serie di punti denominati "tornelli ferro" corrispondenti ad una serie di punti lungo la rete ferroviaria la cui copertura telefonica non raggiunge le principali arterie stradali (autostrade e strade statali principali). L'identificazione di tali punti ha lo scopo di evitare di creare discontinuità nel tracciato telefonico degli utenti che percorrono tratte lunghe. Tale comportamento, infatti, potrebbe essere causato dalla mancanza di stazioni «reali» nel percorso. In questo modo è possibile quindi garantire la "continuità" al viaggio dell'utente, assimilando quindi i "tornelli ferro" a delle stazioni intermedie senza però creare ambiguità fra utenti in movimento lungo la rete ferroviaria e quelli in mobilità su gomma. Nella seguente immagine è possibile vedere una serie di stazioni reali (rosso) e i tornelli ferro (blu) identificate sulla rete ferroviaria.

**ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE
ATTRAVERSO L'USO DI BIG DATA**

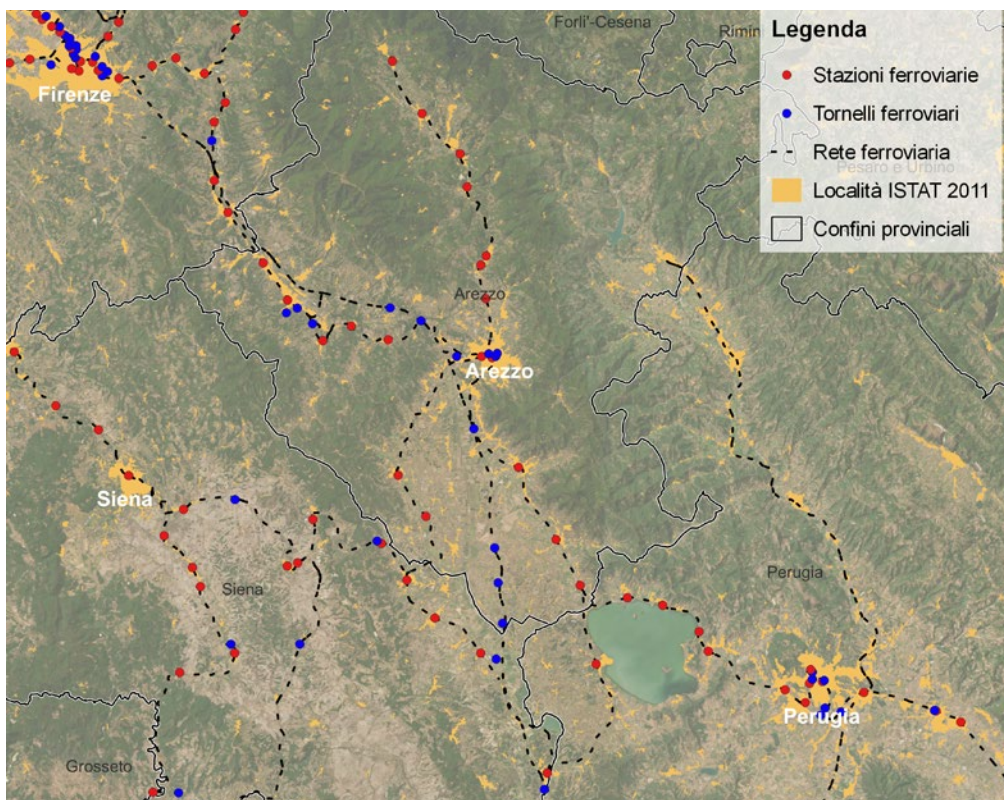


Figura 20 – Individuazione dei tornelli “ferro” funzionali all’identificazione degli spostamenti effettuati in treno

- Una volta indentificati i “tornelli ferro”, il passo successivo prevede l’identificazione della stazione di inizio e di fine del

viaggio, nonché tutte le stazioni intermedie toccate dall’utente.

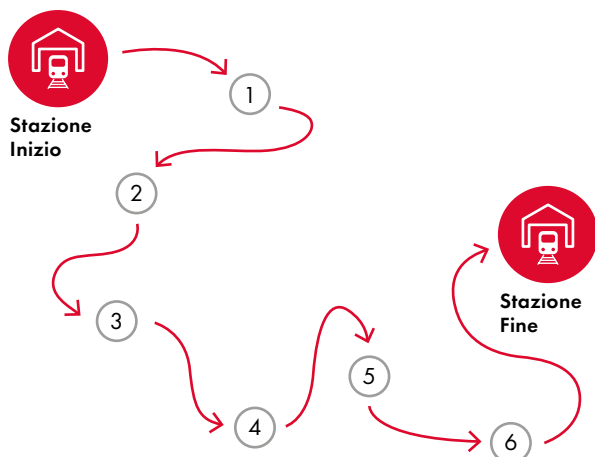


Figura 21 - Rappresentazione dello schema di identificazione del modo treno

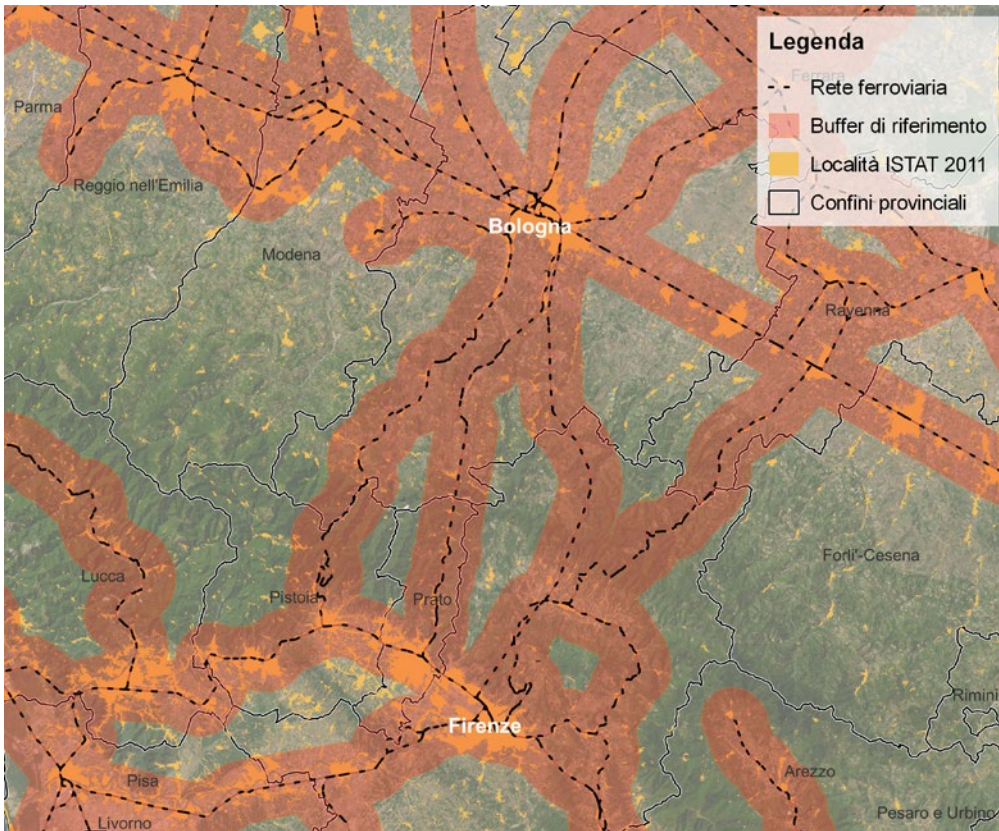


Figura 22 Individuazione buffer rete ferroviaria

Uno spostamento-treno viene definito tale quando la prima stazione (stazione d'inizio spostamento) e l'ultima stazione (stazione di fine spostamento) in cui viene rilevato l'utente sono diverse e l'utente transita attraverso un numero minimo di stazioni reali o tornelli intermedi. Il numero minimo di punti di transito intermedi attraversati dall'utente nello spostamento viene definito in funzione della distanza chilometrica tra la stazione di origine e quella di destinazione.

Una volta effettuata questa prima classificazione dei viaggiatori-treno, vengono applicati **tre ulteriori livelli di filtro** al fine di raffinare

ulteriormente la classificazione degli stessi:

1. Il primo livello di filtro aggiuntivo consiste nell'identificazione dei "tornelli gomma", ossia i punti in cui non si transita in treno. L'utente treno ideale non è visto transitare in alcun "tornello gomma".
2. Il secondo livello di raffinamento prevede l'utilizzo di una metodologia data-driven per la validazione finale dei percorsi. Per validare la sequenza di stazioni è stato costruito il grafo dell'intera rete ferroviaria nel territorio nazionale che viene confrontato con i possibili cammini "teorici" tra due stazioni.

3. Infine, l'ultimo livello di raffinamento, anch'esso data-driven, per l'identificazione dei "passeggeri treno" prevede che per ogni coppia origine-destinazione si identifichi un insieme di stazioni intermedie valide. L'utente treno ideale è visto transitare in tutte le stazioni di controllo intermedie.

Combinando in un modello statistico le informazioni provenienti dai filtri precedentemente descritti, si ottiene il subset di utenti che, verosimilmente, compiono un viaggio in treno. Inoltre, affinché venga associata la modalità "treno" ad un viaggio è necessario che la distanza di spostamento in treno (dalla stazione di partenza alla stazione di arrivo) copra almeno la metà della distanza di spostamento dell'utente tra l'area di origine e quella di destinazione. Qualora un utente avesse compiuto un viaggio multimodale utilizzando sia l'aereo sia il treno, la modalità di viaggio verrà comunque considerata solamente "aereo". Infine, tutti gli spostamenti tra un'area di origine e una di destinazione che non rispettano le condizioni per essere classificati nella modalità "aereo" o "treno", vengono classificati con modalità "altro".

La definizione dell'algoritmo per l'individuazione del modo di spostamento sopra descritto è in realtà solo il primo passo di un processo con ampie prospettive di miglioramento. Un primo perfezionamento della metodologia sopra descritta potrebbe derivare dall'integrazione degli orari schedulati dei mezzi di trasporto pubblico, grazie anche alla disponibilità nel formato GTFS, in modo da poter associare le tracce registrate ai servizi orari, anche se si apre immediatamente la questione legata alla rappresentatività dello schedulato rispetto al circolato effettivo. Proprio nella direzione del Data Fusion è orientato un aggiornamento in corso dell'algoritmo modale per tenere conto dello scheduling aereo, in modo da associare correttamente gli utenti che gravitano nelle zone aeroportuali ai passeggeri.

L'altro tema emerso durante l'applicazione del metodo sopra descritto, ed ancora in via di risoluzione, è quello della presenza di viaggi aerei fra province adiacenti, ovvero di errori di attribuzione del modo dovuti probabilmente alla cella aeroportuale che copre altre destinazioni o al comportamento del personale aeroportuale.

È inoltre importante osservare che nelle analisi condotte fino ad oggi non sono mai stati presi in considerazione gli spostamenti via mare perché al momento non interessanti per lo studio in corso; l'algoritmo di identificazione modale dovrebbe quindi essere integrato con un processo specifico per l'individuazione di questi utenti modali.

Un'altra criticità è legata all'anonimizzazione del dato telefonico; quando gli utenti che effettuano un certo tipo di spostamento non raggiungono un numero minimo definito a priori, vengono "mascherati" per evitare che si possa risalire alla loro identità, in un'ottica di protezione della riservatezza. Questo ha chiaramente delle ripercussioni sulla definizione dell'insieme reale delle coppie O/D dove sono registrati degli spostamenti non nulli e sulle modalità di trasporto.

A livello di espansione del campione all'universo, dallo studio è emerso che su alcune relazioni Origine/Destinazione non valgono le stesse regole applicabili in generale. In particolare, per la quantificazione degli spostamenti-treno, su alcune O/D (ad esempio Roma- Milano), utilizzando la proporzione standard di SIM business rispetto alle ordinarie e confrontando i risultati ottenuti con i dati di bigliettazione ferroviaria, emergevano forti discrepanze, legate presumibilmente alla natura della stessa relazione O/D, che vede rispetto alla media una composizione diversa dell'utenza, a favore di spostamenti legati a lavoro/affari rispetto ad altri motivi. Altro aspetto relativo all'espansione ed alla rappresentatività del campione è legato all'inclusione di SIM di autisti di servizi di trasporto di media/lunga percorrenza e soprattutto

di autisti di mezzi di trasporto merci. Questi spostamenti vengono erroneamente ascritti al mondo degli utenti standard ma sono in realtà relativi a spostamenti di tipo logistico. È in corso un approfondimento metodologico per comprendere se esiste la possibilità di filtrare tali segnali.

Rimane un tema del tutto da esplorare quello relativo al motivo dello spostamento. Da letteratura¹² è noto che le motivazioni di spostamento individuabili con maggiore affidabilità a partire dai Big Data telefonici, siano quelli degli spostamenti casa/lavoro, casa/scuola e ritorno a casa, in aggiunta alla classificazione in spostamenti ricorrenti e spostamenti non ricorrenti. Alcune proposte vanno nella direzione dell'utilizzo congiunto di altre fonti di dati, come le immagini satellitari relative all'uso del suolo e la localizzazione dei Points

of Interest; anche queste ipotesi, tuttavia, trovano un limite nella forte commistione di funzioni e servizi in ambito urbano.

Altri spunti, per arrivare ad un maggiore dettaglio rispetto allo scopo del viaggio, sono quelli relativi all'uso dei Big Data derivanti dai Weblog e dalle Picocells: con la prima tipologia si utilizza la lista dei siti recentemente consultati (*log*) dall'utente, per avere una caratterizzazione di maggior dettaglio delle attività che compie; le Picocells sono invece celle telefoniche di dimensioni più piccole rispetto a quelle standard, poste generalmente in luoghi con elevate concentrazioni di persone come i centri commerciali, i cui segnali consentirebbero di identificare con maggiore verosimiglianza i luoghi frequentati e metterli quindi in relazione con le probabili attività svolte dall'utente.

7.2 | CONCLUSIONI

Si riporta di seguito una tabella di sintesi dell'esperienza sopra descritta, in cui si legano le opportunità ed i limiti individuati nello studio condotto utilizzando i Big Data

telefonici; opportunità e limiti sono strutturati in funzione delle esigenze di base delle analisi di mobilità.



12. Luis Willumsen, "Use of Big Data in Transport Modelling-Discussion paper" – International Transport Forum, 2021

**ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE
ATTRAVERSO L'USO DI BIG DATA**

ESIGENZA TRASPORTISTICA	OPPORTUNITÀ	AMBITO DEL LIMITE
DISTRIBUZIONE SPAZIALE DEGLI SPOSTAMENTI O/D	Utilizzando le georeferenziazioni proveniente dai dati delle app si ha una definizione molto precisa dell'origine/destinazione dello spostamento.	Spostamenti di piccola entità, urbani.
DISTRIBUZIONE TEMPORALE/ FREQUENZA DELLO SPOSTAMENTO	Numerosità dei dati e disponibilità per estesi periodi temporali permettono di definire le caratteristiche mensili, settimanali, orarie degli spostamenti.	
MOTIVO DELLO SPOSTAMENTO		
MODO SPOSTAMENTO	Attraverso la realizzazione di uno specifico algoritmo di attribuzione modale si possono individuare in ambito extraurbano i tre modi: aereo, treno, altro. Possibile implementazione dell'algoritmo con il modo Nave, non affrontato nello studio oggetto di descrizione.	Spostamenti urbani Spostamenti extraurbani
ESPANSIONE DEL DATO ALL'UNIVERSO	La numerosità del campione nello spazio e nel tempo permette di considerare irrilevante l'errore campionario. Gli operatori telefonici possono poi contare sulle informazioni provenienti dai contratti legati alle SIM e da quelle reperite tramite contatto diretto con l'utente tramite interfacce commerciali.	

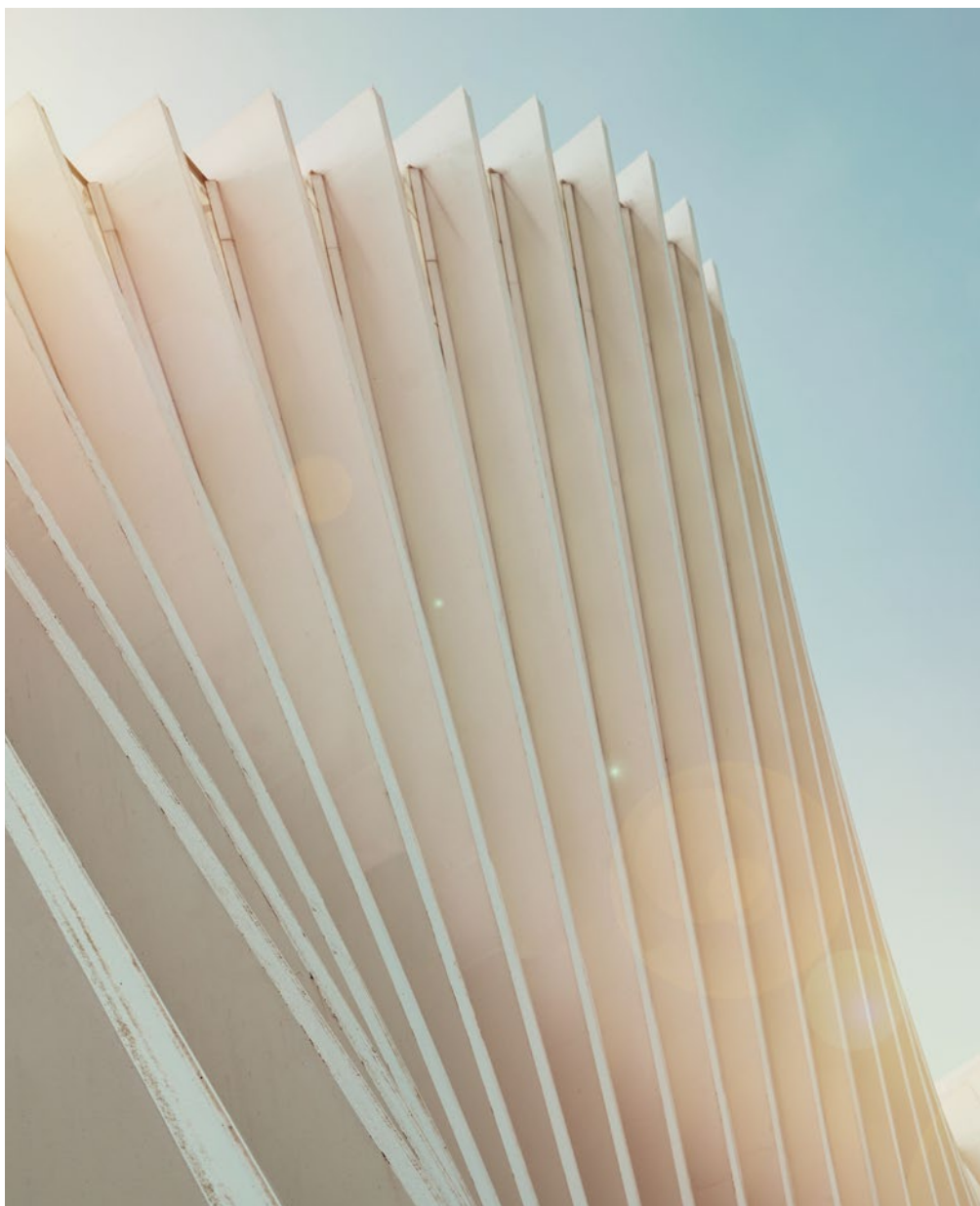
Tabella 7 - Sintesi delle potenzialità e dei limiti dei Big Data telefonici rispetto allo studio descritto nel capitolo

LIMITE	SOLUZIONE/ POSSIBILE SOLUZIONE
Non tutti i devices e gli utenti permettono la georeferenziazione da app.	
Problema di protezione della privacy.	
Basandosi sui dati di posizione legati alle celle telefoniche si ha il fenomeno dell'utente fermo che risulta però in movimento perché aggancia celle diverse che coprono la stessa area (cell jumps).	Media pesata delle distanze delle celle telefoniche visitate nel periodo di riferimento
Non è stato ancora affrontata questo tipo di analisi specifica nel caso studio. Da letteratura è noto il problema dell'attribuzione del motivo in particolare per gli spostamenti non sistematici.	Data fusion con uso del territorio e altri tipi di Big Datt come POI
<p>Difficile la ricostruzione del modo di spostamento a causa delle possibili sovrapposizioni delle varie modalità sia in termini di velocità che di sede fisica in cui avviene lo spostamento.</p> <p>Non è possibile distinguere fra Auto e Trasporto pubblico collettivo.</p> <p>Esistono attribuzioni dubbie rispetto al modo aereo.</p> <p>Esistono utenti per cui non è possibile risalire al modo per le regole di protezione della privacy.</p>	Data fusion con GTFS
Non è possibile distinguere fra utenti privati e utenti "commerciali", come autisti, mezzi commerciali e trasporto collettivo.	Data fusion/Confronto con altre fonti dati, nel caso in questione bigliettazione ferroviaria
Alcune O/D presentano peculiarità (alta presenza business) per cui si rende necessario un'espansione all'universo distinta dalle altre.	
Le matrici finali saranno matrici di persone al netto di trasformazioni successive tramite coefficienti di riempimento mezzi privati.	

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

Come evidente dalla Tabella 7, i Big Data telefonici oggetto di studio sono potenzialmente una risorsa preziosa per le analisi di mobilità; almeno ad oggi però non sono dati pronti all'uso e richiedono la costruzione di un know-how che necessita della stretta interazione fra società di analisi a valle dell'o-

peratore telefonico e tecnici del settore dei trasporti e delle statistiche in materia, al fine di individuare le grandezze, definire le modalità di estrazione e le eventuali criticità presenti ed ottenere quindi il dato utile alle analisi ed alle statistiche trasportistiche.



8 | ANALISI E MODELLIZZAZIONE

Tartaglia Mario¹[0000-0003-3216-8150], Lorenzo Vannacci¹[0000-0001-9587-7611], Martina Farsi¹[0000-0002-3132-8071]
 1. FS Research Centre, Firenze, Italia

8.1 | FORMATI E STRUMENTI DI ARCHIVIAZIONE, GESTIONE ED ANALISI DEI BIG DATA

8.1.1 | STRUMENTI PER L'ARCHIVIAZIONE E LA GESTIONE DEI BIG DATA

In questo paragrafo si illustrano sinteticamente gli strumenti a disposizione per l'utilizzo dei Big Data, soffermandosi soprattutto sui prodotti open-source in ragione della loro facile reperibilità, convenienza e maggiore interoperabilità (Antonioni et al., 2019)[8.1]. I dati grezzi possono presentarsi in formati molto differenti fra loro; possono essere infatti strutturati ed organizzati secondo lo schema classico delle tabelle definite in maniera chiusa per righe e colonne e adatti all'utilizzo nei database "di tipo tradizionale", semi-strutturati come Json, GeoJson, HTML e XML ed infine non strutturati come foto, video, immagini.

Per l'archiviazione, l'interpretazione e l'analisi dei contenuti dei diversi formati è possibile avvalersi di strumenti come linguaggi di scripting e software di analisi, che permettono in primo luogo di estrarre le informazioni utili, trasformarle nel formato desiderato e adatto alle analisi ed infine importarle nel database di archiviazione.

Python, linguaggio di programmazione di tipo scripting ed *object-oriented*, è uno degli strumenti open-source che permette le operazioni sopra descritte; grazie all'utilizzo di librerie specifiche consente l'estrazione dei dati, l'importazione in database e l'analisi statistica. È inoltre utilizzato nei processi di Machine Learning e sviluppo di applicativi web.

R è un altro strumento open source utilizzabile

per l'analisi di dati; questo linguaggio, rispetto a Python, è orientato in misura maggiore sull'analisi statistica e la restituzione grafica dei dati ed è utilizzato ampiamente per il data mining. Per contro Python risulta più versatile e più adatto alle attività di programmazione in generale.

Infine, **Matlab**, software commerciale, consiste in una piattaforma di programmazione e calcolo numerico per lo sviluppo di algoritmi, l'analisi dati e la creazione di modelli. Spesso assimilato a **R**, come software commerciale presenta però aspetti vantaggiosi, come una dettagliata documentazione e pacchetti di sviluppo maggiormente robusti, aspetti che hanno la loro controparte nel costo di acquisto del software. Per quanto riguarda la programmazione, Python viene ritenuto maggiormente flessibile rispetto a Matlab [8.1]. Per quanto riguarda infine l'aspetto dell'archiviazione, organizzazione dei dati ed interazioni con gli stessi è necessario disporre di un sistema di gestione del database e definire il tipo di rappresentazione dei dati effettuata dal database (Data model). La tipologia di database più comune è quello relazionale, **RDBMS** (*Relational Database Management System*) con il quale si può interagire con un linguaggio SQL. I software di gestione dei database di questo tipo sono molteplici: fra quelli open source troviamo una delle versioni di MySQL, PostgreSQL che combinato a PostGIS permette la gestione di dati spaziali,

analogamente a SQLite insieme a Spatialite, mentre fra quelli commerciali si annoverano Microsoft SQL Server e Microsoft Access. Altra filosofia è quella dei database **No-SQL**, definiti anche "Not only SQL" o ancora database non relazionali, che memorizzano i dati in maniera diversa rispetto alle tabelle relazionali. Sebbene siano nati anch'essi negli anni '60-'70' come i RDBMS, hanno visto uno sviluppo consistente negli anni 2000 in concomitanza con l'avvento dei Big Data e delle applicazioni web real-time e contestualmente all'evoluzione della capacità di archiviazione dei dati. Con i database NoSQL non è più necessario definire una strutturazione specifica dei dati da immagazzinare e si prestano ad evoluzioni dei dati nel tempo; questo li rende più flessibili e più veloci nelle interrogazioni rispetto ai RDBMS. Inoltre, permettono una scalabilità di tipo orizzontale (scaling out): invece di aumentare la potenza della macchina per gestire moli di dati sempre più grandi (scalabilità verticale, scaling up), si distribuiscono i dati su più macchine. Per contro si può avere perdita di consistenza dei dati nell'accezione che il risultato di una query sul dataset potrebbe non riflettere esattamente i cambiamenti più recenti avvenuti nei dati, non si ha un

linguaggio sviluppato come SQL per l'integrazione con il database e conseguentemente potrebbe essere più difficoltoso eseguire le queries e le joins, se possibili, più complesse. In aggiunta, vengono meno le caratteristiche delle transazioni individuate dall'acronimo **ACID** (atomicità, coerenza, isolamento, durabilità) proprie dei databases relazionali. I NoSQL databases seguono invece un principio detto **BASE**: *Basic Availability, Soft State, Eventual consistency*. Il principio BASE deriva dal **teorema di CAP**: è impossibile per un sistema informatico distribuito garantire contemporaneamente *Consistency* (tutti i nodi vedono gli stessi dati nello stesso momento), *Availability* (la garanzia che ogni richiesta riceva una risposta su ciò che sia riuscito o fallito), *Partition tolerance* (il sistema continua a funzionare nonostante perdite arbitrarie di messaggi). I principali tipi di database NoSQL utilizzati sono i database documentali (MongoDB, COuch DB) i database di valori chiave (Membase, Redis, Amazon Dynamo), gli archivi wide column (Cassandra, Hbase), i database a grafo (Neo4J, Infogrid). L'impiego dei databases NoSQL è ormai ampiamente diffuso; social networks, Google Earth, Maps, eBay, Amazon e molti altri applicativi si basano su questo tipo di tecnologia.

8.2 | ANALISI DEI BIG DATA E UTILIZZO PER LA PREVISIONE DI FENOMENI DI INTERESSE

Una volta definiti sinteticamente e non esaurientemente gli strumenti a disposizione per la gestione di questi grandi quantità di dati, si descrivono le tecniche di estrazione dell'informazione di interesse, attività definita come **Data mining**. Le tecniche tradizionali potrebbero essere inefficaci o necessitare di tempi lunghi per analizzare grandi ed eterogenee moli di dati; si interviene quindi utilizzando strumenti informatici e tecniche innovative per ricercare strutture che si ripetono nei dati a disposizione (**Data Mining descrittivo o in-**

terpretativo) e/o prevedere eventi o probabilità di accadimento (**Data Mining predittivo**). Il Data Mining si inserisce nel processo di elaborazione dei Big Data, che si compone di molteplici fasi: si parte dall'identificazione dell'obiettivo che si vuole raggiungere attraverso l'utilizzo dei dati selezionati, si passa quindi alla preselezione dei dati in funzione dell'obiettivo individuato ed infine alla pulizia degli stessi dati. Il trattamento applicato ai dati dipende dal loro tipo e dimensione, ma anche dallo scopo dell'analisi [8.7]: diverso

sarà infatti lo strumento adottato, a seconda che il dato serva per alimentare un sistema interattivo come le smart cities o venga utilizzato a fini statistici o per analisi di mercato. Oltre alla classificazione generale in base allo scopo (descrittivo e/o predittivo), le tecniche di Data Mining vengono suddivise secondo le seguenti principali categorie, che al loro interno comprendono diversi metodi ed algoritmi [8.7]:

- **Classification (Classificazione)** - si tratta di un approccio di Machine Learning supervisionato che utilizza un gruppo iniziale di oggetti come dati di "training". Il gruppo iniziale viene utilizzato per creare un modello esplicativo della relazione fra un attributo target - in funzione del quale sono definite le classi - e gli altri attributi degli oggetti dell'insieme di allenamento. Una volta testato, il modello così elaborato viene utilizzato per attribuire gli oggetti alle classi predefinite. Sono catalogati come tali i seguenti metodi: Bayesian network, Support Vector Machine (SVM), k-nearest neighbour (KNN).
- **Clustering (Raggruppamento)** - si basa sul Machine Learning non supervisionato: i gruppi sono creati a partire dalle caratteristiche distintive e significative dei dati analizzati, sulla base delle quali è possibile definire una funzione di "vicinanza" che permette di attribuire gli oggetti a un gruppo piuttosto che a un altro. Fanno parte di questo gruppo le metodologie di partitioning, hierarchical clustering, Co-occurrence.
- **Association rule mining (Regole associative)** - in questa classe rientrano i metodi che sono specializzati nell'identificare e creare regole basate sulla frequenza di accadimento di dati numerici e non numerici. Questa tipologia di data mining è utilizzata soprattutto nelle analisi di mercato e nelle strategie di business;
- **Prediction categories (Serie tempo-**

rali) - fanno parte di questo gruppo i metodi che utilizzano serie storiche per definire le tendenze ed il comportamento dei dati. In questo contesto vengono utilizzati SVM e algoritmi di logica fuzzy che permettono di identificare le relazioni fra variabili dipendenti e indipendenti e ricavare le curve di regressione per le previsioni.

- **Regressione** - generalmente è utilizzata con algoritmi di apprendimento supervisionato. Il risultato di queste tecniche è la definizione di un output caratterizzato da continuità, in contrapposizione al risultato degli approcci di classificazione e clustering. Oltre alle classiche tecniche di regressione, semplice e multipla, si parla nel campo del Machine Learning di ulteriori tipologie come la regressione logistica e Bayesiana.

Il Data Mining viene anche classificato in funzione del campo di specializzazione; si parla per esempio di Process Mining per la branca focalizzata sulla scoperta e monitoraggio dei processi aziendali, di text mining per il data mining volto a estrarre parti di testo per la ricerca di informazioni utili sul WEB, libri, recensioni, articoli ed ancora di Mining Predittivo rispetto alle tecniche adottate per definire trend e quindi adottare decisioni di business o marketing.

Risulta evidente quindi lo stretto legame fra l'analisi e l'utilizzo dei Big Data e la disciplina del Machine Learning, di cui di seguito si riportano sinteticamente i metodi più utilizzati nell'ambito specifico degli studi sulla mobilità.

8.2.1 | MACHINE LEARNING E METODI PIÙ UTILIZZATI NELL'AMBITO DELL'USO DEI BIG DATA PER LA RAPPRESENTAZIONE DELLA MOBILITÀ

Il Machine Learning combina scienze statistiche, ottimizzazione e scienze informatiche [8.1]. Alla sua base troviamo la necessità "storica" di ricercare una relazione che leghi due insiemi di dati ed eventualmente permetta di fare previsioni su uno dei due insiemi a partire dagli elementi dell'altro. Rispetto alla classica ricerca della forma funzionale e dei rispettivi coefficienti che minimizzano l'errore fra valore stimato e valore misurato, il Machine Learning entra in gioco con sistemi complessi, in cui le stesse variabili indipendenti sono a loro volta funzioni di cui stimare il valore e talvolta fra loro interconnesse, mentre le forme funzionali utilizzate nell'algoritmo rimangono ignote all'utente.

I due principali macro-insiemi in cui si organizza la modellizzazione sono il Machine Learning supervisionato e quello non supervisionato. Nel primo caso si ha a disposizione un certo numero di relazioni note (per esempio variabile dipendente-variabile indipendente), per cui è possibile valutare il modello elaborato dal Machine Learning e restituire al sistema un feedback che quantifichi la qualità di riproduzione del fenomeno indagato. Nel Machine Learning non supervisionato la variabile target non è invece presente; in questo caso l'analisi si concentra sull'individuazione di strutture comuni ai dati analizzati ed eventualmente gruppi di "similitudine".

Il primo esempio di Machine Learning risale agli anni '50 del 1900 [1]; successivamente questa disciplina ha visto fasi alterne di sviluppo, fino ad affermarsi definitivamente dal 2010 con le reti neurali. L'avvento dei Big Data ha quindi definitivamente sancito la necessità di disporre di tecniche avanzate per la gestione, la visualizzazione e l'analisi dei dati, che ha trovato il suo corrispettivo nelle tecniche di Machine Learning in continua evoluzione. Si riportano di seguito i quattro filoni

di sviluppo principali per il Machine Learning negli anni 2000:

1. **Rule-based systems.** Questo gruppo di metodi prevede regole predefinite, alberi delle decisioni, programmazione logica; pertanto, sono le più interpretabili fra le varie tecniche di Machine Learning.
2. **Kernel-based algorithms.** Si basano sul concetto di vicinanza: per risolvere il problema si cercano k situazioni simili. Le questioni base sono il criterio di similarità delle situazioni e il numero k . Tali metodi sono detti anche non-parametrici, perché all'utente non è richiesta la definizione di alcuna forma funzionale. La debolezza di questo tipo di modelli è legata all'importante costo computazionale connesso all'inversione di matrice, necessaria per la loro applicazione. Per questo motivo negli ultimi anni a questi si preferiscono i DNN che permettono di aggiungere dati di training senza dover necessariamente riutilizzare tutto il set di dati.
3. **DNN (Deep Neural Networks).** Deep perché hanno numerosi layer "sovrapposti", sono un'evoluzione delle reti neurali iniziali, in cui c'era un solo layer e tutti i neuroni erano connessi a tutta la rete. Anche in questo caso si ha bisogno di numerosi dati di training proprio per "accordare" i numerosi layer di neuroni. Il limite principale è legato però alla scarsa interpretabilità dei parametri in uscita dal sistema; oltre a ciò, mancano di un fondamento teorico applicando un approccio principalmente empirico.
4. **Statistica bayesiana.** La teoria della probabilità bayesiana permette di prevedere la distribuzione di una variabile, assegnato un set di osservazioni. Questo tipo di approccio richiede di definire la struttura del modello o da parte dello

stesso operatore o in maniera automatica attraverso, per esempio, il Bayesian Network structure learning. Il principale beneficio di questo approccio è quello di combinare una formulazione basata su un dominio noto e definito attraverso funzioni note con le tecniche del data-driven, superando quindi il problema dell'“opacità” delle reti neurali. Questa tipologia di approccio si è sviluppata molto a partire dal 2010 con l'avvento dei PGM (Probabilistic Graphical Models).

Per quanto riguarda il campo dei trasporti, l'analisi condotta da Kaffash, Nguyen e Zhu (2021) [8.9], volta a raccogliere, catalogare e sintetizzare tutto il materiale bibliografico relativo ai Sistemi di Trasporto Intelligenti (ITS), all'utilizzo dei Big Data e del Machine Learning per scopi trasportistici, evidenzia come le tecniche di Machine Learning più ricorrenti per l'analisi e l'impiego dei Big Data per studi trasportistici siano le reti neurali e il deep learning. Dallo studio emerge come i confini fra le varie tecniche siano sempre più labili e si tenda alla combinazione dei vari algoritmi. Altra evidenza importante è la limitazione del campo di impiego, dei dati e degli strumenti sopra individuati, in via quasi esclusiva alla previsione dei flussi stradali, della loro velocità, dei tempi di viaggio e delle valutazioni su congestione ed incidenti; l'85% degli articoli analizzati è relativo alle applicazioni nel

campo delle previsioni suddette. Dagli articoli portati in evidenza nella ricerca bibliografica citata, si evince inoltre che si tratta molto spesso di previsioni a breve termine (5-15 minuti) dei flussi stradali, utili quindi nell'ottica dell'informatizzazione e/o gestione smart della città, in cui i dati dei sensori vengono elaborati e utilizzati per rispondere istantaneamente alla gestione dei flussi veicolari, ottimizzando cicli semaforici o suggerendo percorsi alternativi agli utenti. Gli altri campi di applicazione nel settore ITS (Intelligent Transportation System) individuati dall'articolo sono quelli legati a Recognition, Detection, Safety e Optimization. Per “Recognition” si intende l'attività di identificare vari oggetti, quali targhe, veicoli, comportamenti dei guidatori. Per “Detection” si intende invece la possibilità di automatizzare il rilevamento di incidenti, il riconoscimento della segnaletica stradale ma anche l'identificazione di comportamenti degli utenti che potrebbero causare incidenti o problemi di congestione. L'altro campo di applicazione è poi la “Safety” con la quale si intende sia la sicurezza del veicolo ma anche la sicurezza delle strade; i Big Data vengono utilizzati per ricavare informazioni sullo scenario di accadimento degli incidenti e quindi calibrare i modelli di incidentalità. Infine, altre applicazioni sono quelle legate alla “Optimization” intesa come ricerca dell'“ottimale” in termini di percorso, velocità, consumo energetico, tempo di attesa.

8.2.2 | VISUALIZZAZIONE DEI BIG DATA

La visualizzazione dei dati è un passaggio indispensabile per la loro interpretazione, sia nella fase di preparazione per individuare eventuali elementi outlier da eliminare o gruppi di unità affini, sia nelle fasi successive al trattamento, per la valutazione dei risultati. La visualizzazione dei Big Data pone due ordini di questioni: il primo, che si presenta anche con gli insiemi tradizionali di dati, è relativo alla necessità di avere una buona rap-

presentazione, che fornisca tutti gli elementi utili per la lettura dell'immagine (legenda, etichette, individuazione delle misure rappresentate, leggibilità del grafico); il secondo riguarda le caratteristiche dei nuovi dati in questione, in quanto per un singolo record si ha a che fare con una mole ingente di informazioni. La multidimensionalità dei nuovi dati, che rende inapplicabili le tecniche classiche di visualizzazione come gli istogrammi ed i

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

semplici diagrammi a dispersione bidimensionali, impone l'utilizzo di tecniche nuove, anch'esse in evoluzione come le tecniche di Machine Learning.

Nel capitolo "Data Science and Data Visualization" di "Mobility patterns, Big Data and Transport analytics" (Antoniou et al., 2019) [8.1] vengono individuati due gruppi principali di tecniche di visualizzazione. Appartengono al primo gruppo i metodi che permettono la visualizzazione contemporanea di più dimensioni, raggruppate a due a due come nel caso della matrice Scatterplot, oppure utilizzano molteplici assi dimensionali, come nel caso della visualizzazione per Coordina-

te Parallele, in cui sono presenti diversi assi verticali cui corrispondono diverse dimensioni. Altre tipologie appartenenti a questo primo gruppo sono le cosiddette "mappe di calore", in cui la terza dimensione viene rappresentata con il colore, le visualizzazioni "pixel-oriented" in cui si sfrutta l'unità pixel per rappresentare quante più dimensioni possibili rispetto a un dato record, ed ancora le visualizzazioni di tipo "gamminder", ovvero rappresentazioni bidimensionali in cui dimensioni aggiuntive rispetto a quelle degli assi cartesiani sono inserite sfruttando colori, dimensioni e forme del punto rappresentativo del record.

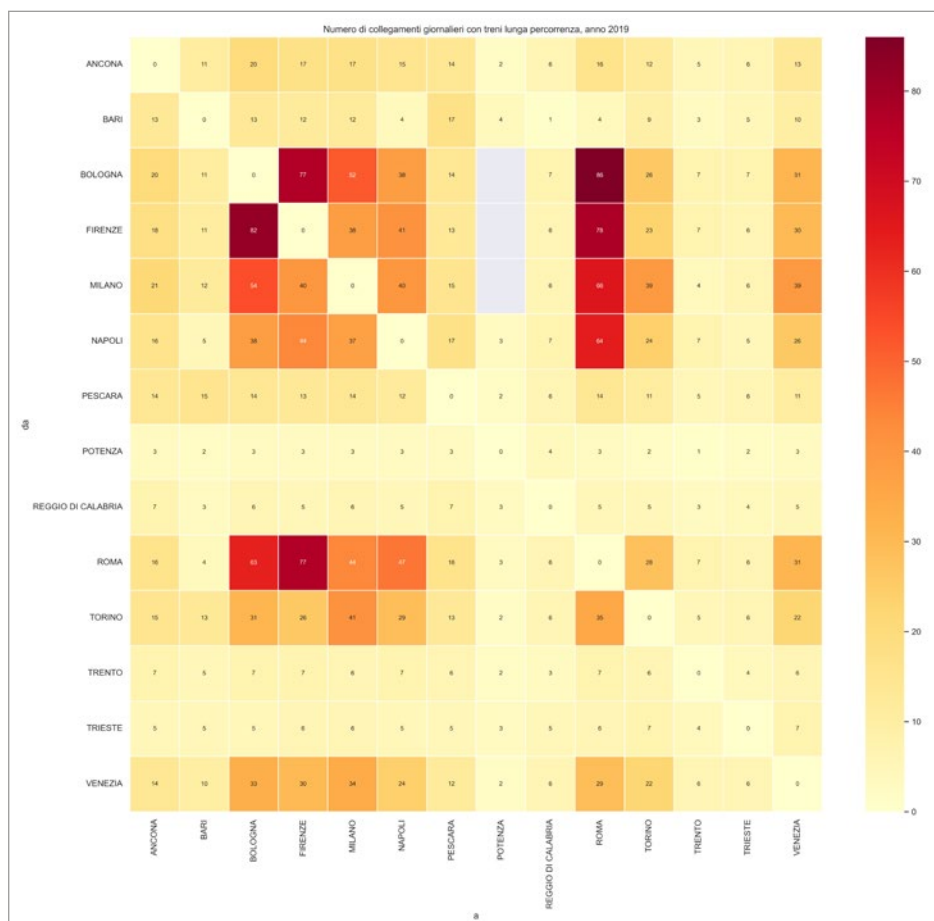


Figura 23- Esempio di Heatmap (realizzato con Seaborn-Python [8.11]), rappresentativa del numero di collegamenti lunga percorrenza presenti giornalmente fra le principali relazioni italiane.

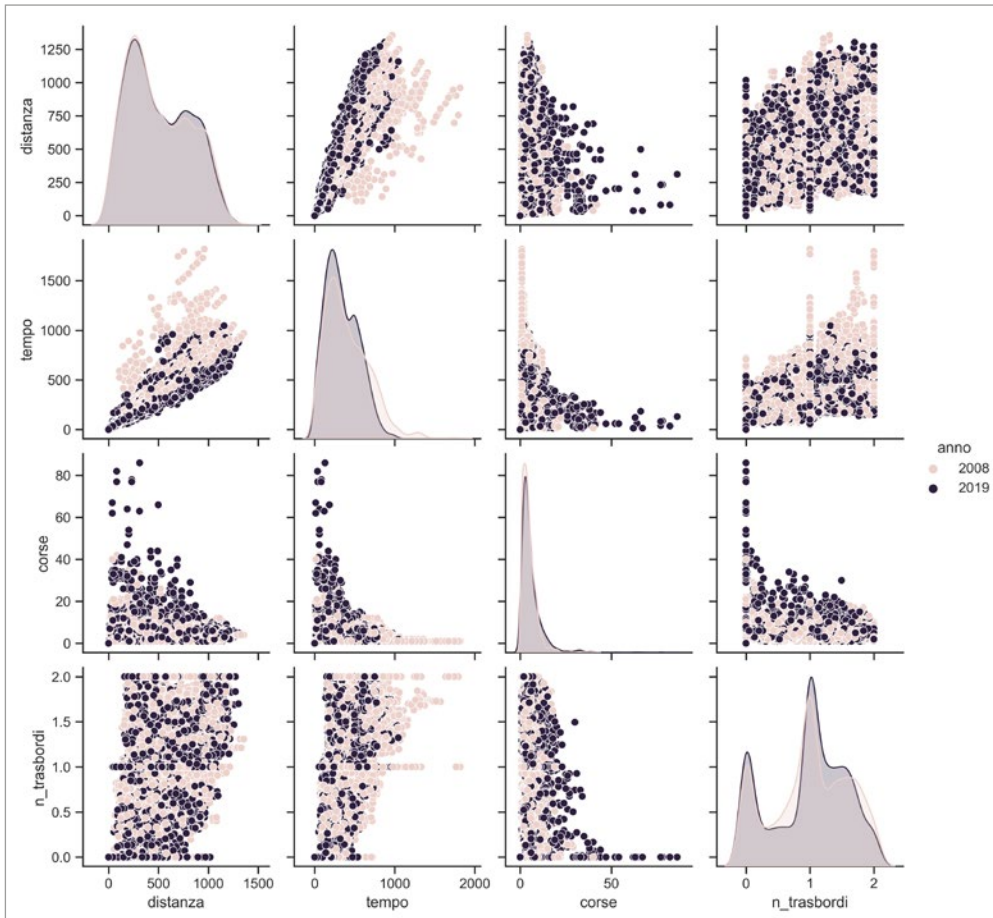


Figura 24- Esempio di scatterplot matrix (realizzato con Seaborn-Python[8.11]), utilizzando i dati 2008 e 2019 delle OD lunga percorrenza.

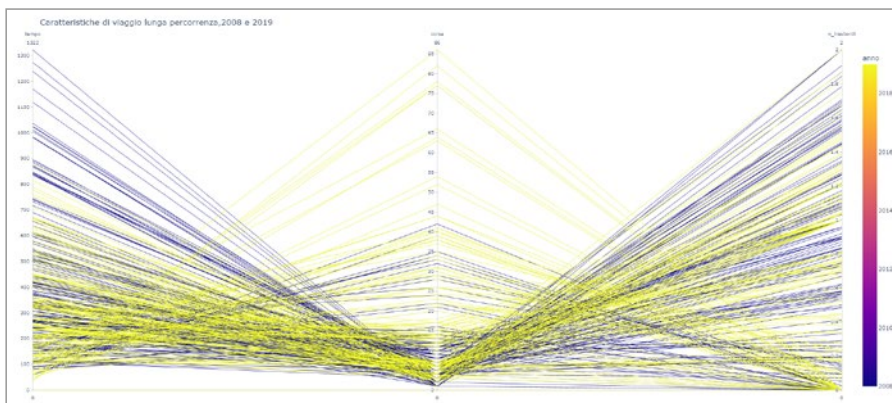


Figura 25- Esempio di coordinate parallele per la visualizzazione delle caratteristiche di viaggio 2008 (blu) e 2019 (giallo) delle OD lunga percorrenza (realizzato con Plotly-Python).

ELEMENTI METODOLOGICI PER L'ANALISI DELLA MOBILITÀ DELLE PERSONE ATTRAVERSO L'USO DI BIG DATA

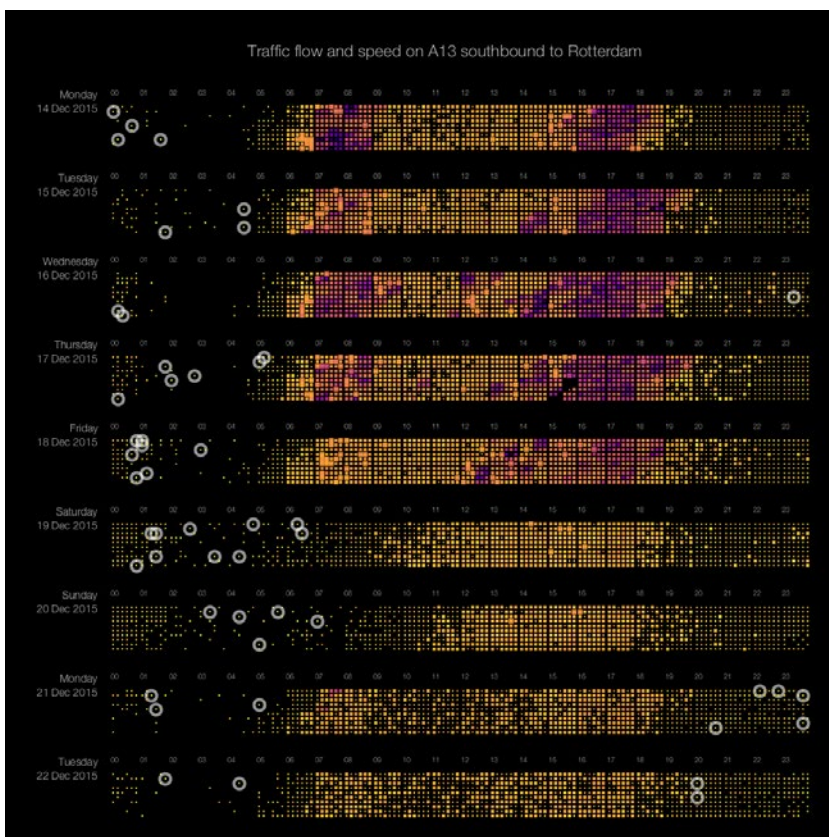


Figura 26- Pixel-plot dei dati di traffico della A13 a Rotterdam. I colori indicano la velocità media del flusso, la dimensione dei rettangoli il valore del flusso, mentre i cerchi individuano valori elevati di velocità (>150 km/h) - Fonte Pixel-based visualization of traffic data, by Erik Boertjes, 2017 [8.10].

Il secondo gruppo di tecniche di visualizzazione è invece costituito dai metodi di riduzione dimensionale, ulteriormente suddivisi in metodi che utilizzano relazioni di combinazione lineare fra le variabili e metodi che utilizzano tecniche di combinazione non lineare, al fine di arrivare ad un insieme di analisi ridotto rispetto a quello di partenza [8.12]. In questo gruppo, i metodi di maggior interes-

se sono: l'analisi in componenti principali (PCA), lo scaling multidimensionale (MDS), la mappatura delle caratteristiche isometriche (ISOMAP), Locally Linear Embedding (LLE) e le sue varianti come la Hessian LLE, t-distributed stochastic neighbor embedding (t-SNE), la Kernel PCA, la Local Tangent Space Alignment (LTSA) e le tecniche di autoencoders che si basano sulle reti neurali [8.1],[8.12].

8.3 | UTILIZZO E PROSPETTIVE FUTURE DEI BIG DATA E MACHINE LEARNING NELL'AMBITO DEI TRASPORTI

Come messo in evidenza da Kaffash, Nguyen e Zhu (2021) [8.9], la maggior parte delle applicazioni che utilizzano i Big Data nel campo dei trasporti finora sperimentate, sono relative alla previsione dei flussi e dei tempi di percorrenza veicolari. Le opportunità nell'utilizzo dei Big Data in combinazione con le tecniche del Machine Learning sono però molteplici e spaziano dall'ottimizzazione dei flussi in termini di percorsi, velocità, consumi di carburante e tempi di attesa, alla sicurezza stradale con l'individuazione delle situazioni di potenziale pericolo grazie al riconoscimento di stati della rete perturbati o ad elementi che determinano incrementi di pericolosità nella circolazione, quali per esempio particolari condizioni meteo o la presenza di mezzi di dimensioni diverse dallo standard. Nell'ambito della modellistica dei trasporti risulta interessante la ricerca illustrata dall'articolo "A deep gravity model for mobility flows generation" di Simini, Barlacchi, Luca, Pappalardo (2021) [8.13] che illustra come Big Data e Machine Learning vengono utilizzati per la costruzione di un modello di distribuzione su diverse aree in Inghilterra, Italia e nello stato di New York.

Partendo dalla forma tradizionale del modello distributivo basato sulla legge gravitazionale, si ipotizza il problema della distribuzione come un problema di classificazione, in cui ogni flusso registrato deve essere attribuito alla classe corretta, corrispondente con la destinazione, ed il modello di distribuzione viene quindi accostato ad un classificatore con due variabili esplicative, popolazione e distanza. A partire da queste considerazioni sullo stato di fatto del modello distributivo si implementa una rete neurale profonda, con strati nascosti. Una parte delle osservazioni disponibili viene utilizzata per addestrare il sistema, mentre la restante parte viene utilizzata per valutare le prestazioni del modello. La rete è alimen-

tata oltre che dai flussi noti e dalle distanze geografiche, dalle caratteristiche di origini e destinazioni disponibili da OpenStreetMap, come uso del suolo, infrastrutture di trasporto, servizi sanitari e scolastici, punti di ristoro. L'output della rete è costituito da un punteggio per ciascuna coppia O/D: maggiore è il punteggio, maggiore la probabilità che si abbia un flusso su quella relazione. Due aspetti importanti valutati nell'applicazione del metodo sono la trasferibilità dei risultati nell'ottica di poter utilizzare il modello in contesti non previsti nel set di addestramento e la valutazione delle caratteristiche che influenzano i risultati ottenuti. Per quanto riguarda il primo elemento, l'utilizzo dei dati a disposizione è stato progettato appositamente per poter valutare la risposta del modello in ambienti "sconosciuti", con buoni risultati finali. Il secondo elemento è strutturale rispetto all'utilizzo del Machine Learning nell'ambito delle applicazioni trasportistiche; la non conoscenza delle regole che determinano la risposta della rete neurale profonda costituisce il limite nell'utilizzo di queste nuove tecniche, perché impedisce l'interpretazione dei risultati e la stima dell'affidabilità del modello. In questo studio invece, attraverso strumenti che a loro volta si basano sulla teoria dei giochi, viene valutata l'influenza delle variabili immesse nel modello sui risultati, sicché l'individuazione delle tendenze al variare degli elementi permette una valutazione critica di quanto ottenuto e riflessioni sulla trasferibilità del sistema in altri contesti.

Altra prospettiva interessante nell'ambito della modellistica dei trasporti, che si ha grazie ai Big Data e alle ampliate capacità computazionali e di storage dei dati, è legata ai modelli activity-based; questo tipo di modelli è basato sul principio che la domanda di trasporto viene "generata" dalle persone nello svolgimento delle loro attività quotidiane. Si

passa quindi dalla modellizzazione tipica dei modelli trip-based, incentrata sui viaggi compiuti dagli individui, alla modellizzazione della sequenza delle attività compiute da un individuo e quindi dei viaggi necessari per compierle. Sebbene l'origine dei modelli activity-based risalga agli anni 70', sono stati utilizzati in piccola misura rispetto ai trip-based; uno dei principali limiti è proprio legato all'elevata quantità di dati utilizzati per la calibrazione, in quanto è necessario ricostruire la catena delle attività dei singoli individui nell'arco della giornata, sequenza che poi determinerà le scelte di trasporto. I Big Data sembrano quindi essere la risposta a questa esigenza ed in questa direzione si muove lo studio "Mobile phone records to feed activity-based models: MATSim for studying a cordon toll policy in Barcelona" [8.14]. In questo studio la domanda di trasporto viene generata da individui ("agents") che compiono le attività quotidiane; ognuno di essi cerca di massimizzare l'utilità delle attività giornaliere programmate, minimizzando il

tempo totale di viaggio. Come input iniziale il modello ha bisogno di una domanda di trasporto costituita da una popolazione di "attori" e dalla pianificazione delle loro attività. Il diario delle attività giornaliere dei singoli utenti viene ricostruito attraverso i dati telefonici di tipo "Call Detail Records" (CDR), registrati nel momento in cui l'utente compie una telefonata, manda un messaggio o effettua una connessione dati. I risultati della calibrazione e della validazione di questo modello danno buone speranze rispetto alla combinazione dei modelli Agent-Based con i Big Data, in particolare quelli provenienti dalla telefonia; fra i possibili vantaggi di un modello Activity-Based ci sono i diversi livelli a cui si può svolgere l'analisi, aggregato o disaggregato, e la possibilità di indagare come particolari politiche gestionali, come nel caso dello studio suddetto in cui si valutano gli effetti di una Congestion Charge, impattino sulle abitudini di mobilità del singolo utente in funzione del tipo di attività generalmente svolte all'interno di una giornata tipo.

8.4 | VALUTAZIONI CONCLUSIVE

Quanto descritto nei paragrafi precedenti, la veloce evoluzione dei metodi computazionali e la grande disponibilità di dati "secondari", esistenti cioè a prescindere dal quesito scientifico cui si vuole rispondere (in contrapposizione con i dati primari provenienti dalle indagini tradizionali), può influire anche sull'approccio che si applica nell'analisi dei dati e in particolare nella costruzione dei modelli. L'approccio classico parte dai dati fornisce delle regole per poi creare un modello che simula i fenomeni e restituisce dei risultati. Può essere quindi utilizzato per prevedere anche risultati relativi a situazioni non ancora in essere, noti i dati caratterizzanti l'evento futuro che potranno essere utilizzati come input del modello. L'approccio alternativo è quello Data Driven, in cui si attua un processo inverso rispetto a

quello dell'approccio classico: l'intelligenza artificiale di qualsiasi tipo viene istruita su quali sono i dati e quali sono i risultati che da essi derivano. Il modello impara e successivamente crea le regole, che possono essere non note all'analista. Le regole poi vengono impiegate con altri dati di input, diversi da quelli di "training", e con l'applicazione dello stesso modello si producono dei risultati. Nell'approccio classico, i dati sono principalmente provenienti da indagini tradizionali mentre nel secondo approccio sono prevalentemente Big Data, proprio per la necessità di avere insieme numerosi per istruire il modello. La questione principale posta dal secondo tipo di approccio è legata alla non conoscenza delle regole applicate al fenomeno: se nella descrizione dello stato di fatto della realtà

può essere un elemento accettabile, non si può dire altrettanto per l'utilizzo del modello in ambito di previsione. I modelli trasportistici classici vengono spesso utilizzati proprio per prevedere gli effetti di interventi gestionali o infrastrutturali non ancora in essere, modellando il comportamento dell'utente attraverso i dati disponibili raccolti da indagini mirate e talvolta progettate proprio con l'obiettivo

di indagare la propensione al cambiamento, come le indagini *Stated Preference*. I Big Data che fotografano lo stato attuale a prescindere dalle necessità modellistiche non permettono questo tipo di approfondimento. Offrono però l'opportunità di integrare e supportare l'approccio tradizionale classico, oltre a rendere possibile l'uso dell'intelligenza artificiale per scopi modellistici.

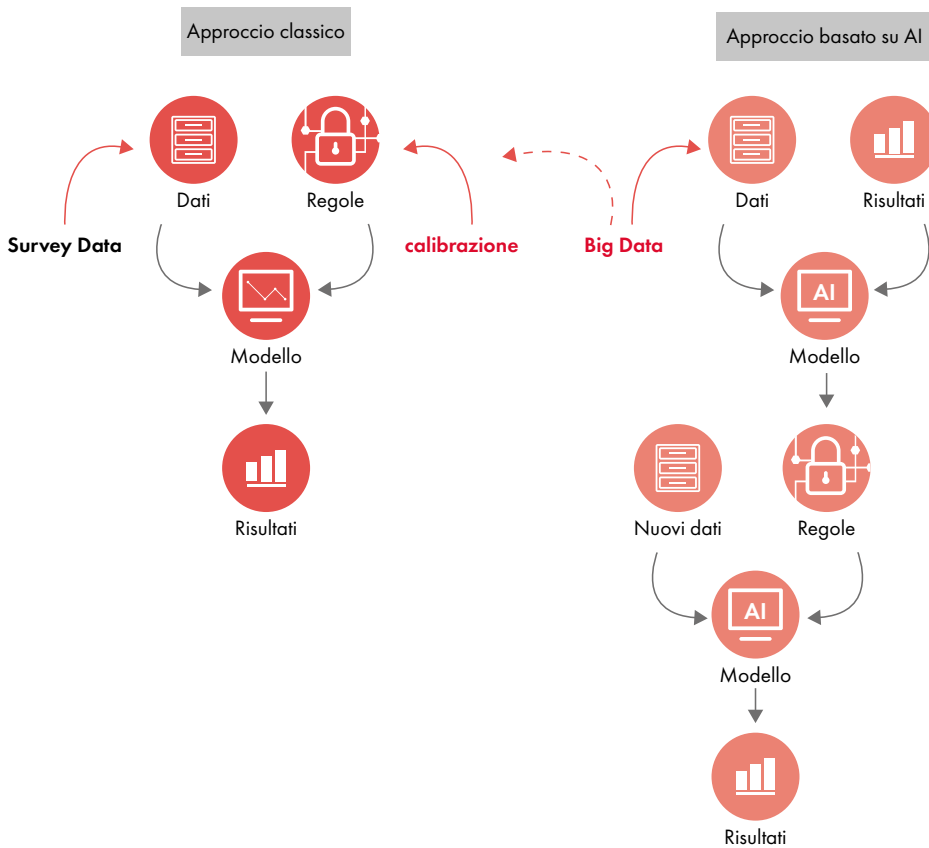


Figura 27 - Approccio classico e approccio Data Driven, il ruolo dei Big Data

8.5 | BIBLIOGRAFIA

- [8.1] Constantinos Antoniou, Loukas Dimitriou, Francisco Pereira, 2019 "Mobility Patterns, Big Data and Transport Analytics. Tools and Applications for Modeling". Elsevier
- [8.2] <https://www.mongodb.com/it-it/nosql-explained>
- [8.3] <https://it.wikipedia.org/wiki/NoSQL>
- [8.4] <https://www.rackone.it/blog/news/big-data-con-database-nosql-unintroduzione-pratica/>
- [8.5] <https://www.next04.it/limportanza-dei-database-nosql-per-lutilizzo-dei-big-data-nel-mondo-connesso-dellinternet-of-things/>
- [8.6] Big Data, NoSQL e Machine Learning: un'applicazione di prediction e recommendation basata sulle API di Amazon
- [8.7] Mohsen Marjani, Fariza Nasaruddin, Abdullah Gani, Ahmad Karim, Ibrahim Abaker Targio Hashem, Aisha Siddiq, And Ibrar Yaqoob, 2017 "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges"- IEEE Access
- [8.8] <https://www.bigdata4innovation.it/data-science/data-mining/data-mining-cose-perche-conviene-utilizzarlo-e-quali-sono-le-attivita-tipiche/>
- [8.9] Sepideh Kaffash, An Truong Nguyen, Joe Zhu, 2021 "Big Data algorithms and applications in intelligent transportation system: A review and bibliometric analysis". International Journal of Production Economics-Elsevier.
- [8.10] Erik Boertjes, Pixel-based visualization of traffic data, <https://social-glass.tudelft.nl/visualising-traffic-data/>, 2017
- [8.11] Waskom M., 2021, "seaborn: statistical data visualization", Journal of Open Source Software, 6:60-3021, <https://doi.org/10.21105/joss.03021>
- [8.12] F.S. Tsai , 2010. Comparative Study of Dimensionality Reduction Techniques for Data Visualization. Journal of Artificial Intelligence, 3: 119-134. <https://scialert.net/abstract/?doi=jai.2010.119.134>
- [8.13] Filippo Simini, Gianni Barlacchi, Massimiliano Luca, Luca Pappalardo, 2021, " A Deep Gravity model for mobility flows generation", Nature Communications
- [8.14] Alex Bassolas, José J. Ramasco, Ricardo Herranz and Oliva G. Cantú-Ros, 2018, "Mobile phone records to feed activity-based models: MATSim for studying a cordon toll policy in Barcelona" Transportation Research Part A: policy and Practice, volume 121 56-74

9 | CONCLUSIONI E RACCOMANDAZIONI

Giovanna Astori¹, Roberta Radini¹, Lorenzo Vannacci²[0000-0001-9587-7611]

1. ISTAT, Roma, Italia,

2. FS Research Centre, Firenze, Italia

Il fenomeno della mobilità delle persone richiede una rappresentazione statistica qualitativa e quantitativa complessa e articolata. Parallelamente, lo studio dei modelli trasportistici necessita di informazioni sempre più dettagliate, per garantire una rappresentazione verosimile dei comportamenti degli utenti.

La normativa internazionale e italiana presenta più di un input in questo senso, richiedendo o normando la produzione di indicatori di vario tipo e con diverse disaggregazioni, non sempre sovrapponibili fra loro.

Le indagini statistiche tradizionali, anch'esse numerose nella produzione attuale e diversificate per obiettivi, risultano impegnative dal punto di vista delle risorse da mettere in campo, e il relativo carico statistico per i rispondenti è abbastanza elevato.

Parallelamente, viviamo una fase di velocizzazione del progresso tecnologico in tutti gli aspetti della vita quotidiana, per cui vengono prodotti continuamente 'segnali' di vario tipo che potrebbero essere utilizzati per aiutare la rappresentazione di vari fenomeni socioeconomici, fra cui appunto la mobilità.

Come riportato in più di un riferimento in letteratura, i dati di per sé non esistono finché non si crea uno strumento per misurarli e raccogliarli. Questo vale sia per le indagini tradizionali, dove gli strumenti sono il campione, il questionario, la tecnica di raccolta, codifica, elaborazione, e via dicendo, sia per i 'segnali' citati, che hanno origine dalla tecnologia con obiettivi totalmente indipendenti dalla rappresentazione dei fenomeni. È pertanto necessario sviluppare metodologie e strumenti idonei a trasformare le enormi moli di segnali (Big Data), e successivamente a trattarli per produrre informazione statistica.

L'obiettivo ambizioso su cui dovrebbero

convergere le numerose esperienze di ricerca in questo campo, è quello di individuare una serie di possibili 'fonti' o segnali per poter integrare le indagini tradizionali, implementando il loro 'riuso' a scopi statistici per ridurre la pressione sui rispondenti e contemporaneamente sfruttarne il potere informativo, che le rende anche idonee all'applicazione di tecniche di elaborazione avanzate nel campo dell'intelligenza artificiale.

Una delle fonti più sondate, nel capo dei Big Data, è rappresentata dai dati telefonici, come ampiamente illustrato. Per loro natura, essi sono ampiamente diffusi e capillari rispetto allo spazio e al tempo, ma strettamente legati alla struttura tecnologica che li genera. È pertanto un lavoro molto complesso e non privo di istanze ancora irrisolte quello di provare a tradurli in 'dati', in informazione fruibile per rappresentare fenomeni come quello della mobilità.

L'impostazione della ricerca in campo statistico, nell'ultimo decennio e a livello internazionale, è quella di implementare dei modelli di Trusted Smart Statistics, in cui non sia più l'istituto di ricerca ad acquisire il dato base appena 'tradotto' dai segnali, per poi elaborarlo e farne degli indicatori, ma che si acquisisca una meta-informazione già parzialmente elaborata dal 'proprietario' dei segnali, che spesso è un soggetto privato. In questo modo si cerca di superare sia un problema di risorse, che di sostenibilità, che di protezione della riservatezza individuale. Inoltre, se si definiscono degli algoritmi comuni, diversi soggetti potranno ricevere dai gestori/detentori un patrimonio informativo standardizzato.

In questa visione, il primo passo è quello della definizione degli algoritmi, per cui si rende

necessario attivare un dialogo fra proprietari dei segnali e ricercatori, attraverso il quale trovare un terreno comune per attuare questa traduzione in informazioni accessibili e che siano il più possibile rispondenti a una rappresentazione del fenomeno di interesse. L'organizzazione del lavoro deve necessariamente prevedere dei pool di esperti di vario tipo (data analyst, trasportisti, statistici ecc.), che riescano sinergicamente a trovare soluzioni per rendere fruibili le nuove fonti di dati in base agli obiettivi di ricerca.

Nei casi di studio presentati, sono state individuate una serie di possibili soluzioni ad alcuni problemi, ma molti ne restano aperti. Tra essi, l'identificazione più disaggregata del mezzo di trasporto utilizzato, al momento distinguibile solo in 'aereo/treno/altro' ma di complessa soluzione per l'ambito urbano; la ricerca della motivazione dello spostamento; la distinzione fra mobilità individuale e spostamenti nel campo della logistica e dei servizi di trasporto persone (attività di autisti/autotrasportatori/

tassisti etc).

È stato evidenziato, inoltre, un insieme di variabili necessarie e previste dai riferimenti normativi e procedurali, per le quali l'utilizzo dei dati telefonici presenta attualmente delle criticità più o meno profonde, e in questo senso sarebbe pertanto necessario approfondire e portare avanti un lavoro di ricerca di soluzioni anche mettendo in campo l'utilizzo di più fonti informative.

In conclusione, il lavoro presentato vuole fornire una panoramica sulle necessità informative, le fonti small e Big Data attualmente disponibili, i metodi di trattamento e utilizzo delle diverse fonti, nonché alcune importanti criticità che devono ancora essere risolte dalla ricerca sul campo. I numerosi spunti emersi possono costituire dei punti di partenza per nuovi casi di studio, con l'obiettivo a tendere di realizzare una sistematizzazione del trattamento e dell'impiego delle fonti individuate, per fornire una rappresentazione sempre più soddisfacente della mobilità.

9.1 | PROSPETTIVE FUTURE

Si apre un nuovo scenario per le analisi di mobilità, si amplia la possibilità di misurare questo fenomeno attraverso nuove fonti dati. Le esperienze fatte dai vari autori di questo lavoro portano necessariamente nella direzione di utilizzare i big data come parte di una strategia di Data Integration che include altre fonti dati che permettano di sciogliere i nodi ancora irrisolti. In questo documento si è voluto quindi descrivere come le fonti di dati tradizionali, come le rilevazioni attraverso questionari, le fonti amministrative, ad oggi rappresentano una piccola porzione dei dati disponibili, comunque importanti. Una grande quantità di dati, che si integrano e completano i precedenti, è invece detenuta dal settore privato (privately-held data). Quest'ultimo,

insieme agli Istituti di statistica, le istituzioni pubbliche e i cittadini, stanno iniziando a partecipare in modo attivo ad un ecosistema di scambio e gestione dei dati, e non ultimo di interessi e aspettative che determinano i comportamenti e le relazioni tra di essi.

Di questa evoluzione si è presa coscienza tra gli istituti di statistica europei già nel 2014, a seguito del memorandum di Scheveningen¹³. Da quel momento, il Sistema statistico europeo (ESS) ha lanciato diversi progetti come la Big Data Task Force nel 2016 e poi nel 2018 per sviluppare competenze metodologiche sui big data e alcuni risultati sono stati riportati nel capitolo 6. Inoltre, iniziative simili hanno avuto luogo in altri progetti a livello delle Nazioni Unite.

13. https://ec.europa.eu/eurostat/cros/content/scheveningen-memorandum_en.

Attualmente tutti gli istituti di statistica sono investiti nella definizione delle Trusted Smart Statistics (TSS) che rappresentano la frontiera dalla statistica. Un nuovo sistema di statistiche che sfruttino le fonti Big Data con i nuovi metodi di elaborazione, utilizzando anche modelli come quelli forniti dall'intelligenza artificiale (Smart Statistics), il tutto costruendo una infrastruttura di dati e di risultati all'insegna della "fiducia" (Trusted). Un sistema di statistiche che vede coinvolti gli stakeholder esterni agli istituti di statistica (cittadini, aziende private, autorità pubbliche), in una condivisione di compiti che vanno dalle elaborazioni, ai controlli dei risultati, del codice degli algoritmi, della condivisione dei risultati, tutto nel rispetto della privacy del cittadino e dei legittimi interessi di business delle aziende. Quindi una condivisione di compiti nel processo di costruzione dei risultati statistici senza condividere i dati grezzi di input, ma consentendo a tutti gli attori di avere esplicita visione di come siano stati ottenuti i prodotti statistici di interesse¹⁴.

Le TSS si prefiggono di costruire un Framework di costruzione degli output statistici che consenta di tracciare la qualità dei vari passi che dai dati grezzi portano all'ottenimento dell'output statistico, anche quando l'intera pipeline di elaborazione non viene eseguita tutta dagli istituti di statistica come avviene per le altre fonti. Questo modello può essere applicato anche da altri enti interessati ad ottenere output da questo tipo di dati. A tale scopo Eurostat nel 2022 ha bandito un progetto, detto Multy MNO (Mobile Network Operator), rivolto ad un consorzio di privati

e Istituti di statistica per definire un processo di produzione generalizzato che consenta di realizzare output statistici in ambito demografico, della mobilità e del turismo utilizzando i dati di telefonia. In particolare, questo progetto ha lo scopo di definire la metodologia di trattamento dei dati e della costruzione di un framework di scambio di dati che vengono trattati in parte dagli MNO e in parte dagli istituti di statistica, nel rispetto della privacy e della trasparenza degli algoritmi di trattamento oltre che del tracciamento della qualità dei prodotti statistici. ISTAT è tra i partner vincitori della gara e il progetto terminerà i lavori entro il 2025.

Nel 2023 Eurostat ha lanciato un progetto ESSNET, a cui partecipa anche ISTAT, per definire strategie di integrazione di altre fonti con i Big Data, in particolare con i dati di telefonia, con una particolare attenzione alla definizione di metodi che consentano di eseguire conteggi di popolazione attraverso questi dati.

Tutto questo fermento di progetti e di finanziamenti è volto a creare soluzioni nell'ambito delle TSS, applicabili ai processi di produzione. Soltanto condividendo le esperienze condotte nei vari ambiti sia del mondo della ricerca universitaria che statistica, delle pubbliche istituzioni, e degli enti privati, ed i risultati con esse conseguiti, è possibile ottimizzare gli sforzi e formalizzare le varie esigenze dei diversi settori di produzione. Il futuro, quindi, non farà scomparire le indagini tradizionali ma le renderà più funzionali ad integrazione di quanto già rilevato da altre fonti.

14. <https://cros-legacy.ec.europa.eu/system/files/sji190584.pdf>





10 | ACRONIMI E DEFINIZIONI

- ACI** = Automobile Club d'Italia
ACID = Atomicità, Coerenza, Isolamento, Durabilità
AGCOM = Autorità per le Garanzie nelle COmunicazioni
AIS = Automatic Identification System
ANPR = Anagrafe Nazionale della Popolazione Residente
API = Application Programming Interface
AV = Alta Velocità
AVQ = indagine Multiscopo Aspetti della Vita Quotidiana
BASE = Basic Availability, Soft State, Eventual consistency
BD = Big Data
BES = Rapporto su Benessere Equo e Sostenibile
BREAL = Big Data REference Architecture and Layers
CAP = Codice di Avviamento Postale
CAP = Consistency, Availability, Partition tolerance
CAPI = Computer Assisted Personal Interviewing
CATI = Computer-Assisted Telephone Interviewing
CAWI = Computer Assisted Web Interviewing
CDO = Chief Data Officer
CDR = Call Detail Records
CNIT = Conto Nazionale delle Infrastrutture e dei Trasporti
CNIMS = Conto Nazionale delle Infrastrutture e della Mobilità Sostenibili
CSD = Circuit Switch
Data Fusion = Integrazione di molteplici origini dati
DAS = Distributed Antenna System
DPO = Data Protection Officer
EMSWe = European Maritime Single Window environment, interfaccia unica marittima a livello europeo
ESS = European Statistical System
FCD = Floating Car Data
FUA = Functional Urban Area, Area Urbana Funzionale
GIAB = Generic Information Architecture for Big Data
GPS = Global Positioning System
GTFS = Global Transit Feed Specification
IMR = Indici di mobilità rilevata
IOT = Internet of Things
ITS = Intelligent transportation system
LAC = Liste Anagrafiche Comunali
LG = Linee Guida
LLE = Locally Linear Embedding
LTSA = Local Tangent Space Alignment
MaaS = Mobility as a Service
MDS = Scaling Multidimensionale
MIT = Ministero delle Infrastrutture e dei Trasporti
MM = Mobility Manager
MND = Mobile Network Data detti anche MPD
MPD = Mobile Phone Data detti anche MND
NSI = Istituti nazionali di statistica
NUTS = Nomenclatura comune delle unità territoriali statistiche europee
OBU = On Board Unit
O/D -OD = Origine/Destinazione
Pax-km = Passeggeri km
PCA = Analisi delle Componenti Principali
PET = Privacy-Enhancing Technologies
PHD = Privately-Held Data
Pkm = Passeggeri km
POI = Point of Interest
PS = Packet Switching
PSCL = Piani Spostamenti Casa Lavoro
PSN = Piano Statistico Nazionale
PUM = Piano urbano della Mobilità
PUMS = Piano Urbano della Mobilità Sostenibile
PUT = Piano Urbano del Traffico
REG/IC = Regionale/Intercity
RDBMS = Relational Database Management System
RMF = Reference Methodological Framework

SDG = Sustainable Development Goals,
Obiettivi di Sviluppo Sostenibile delle Nazioni
Unite

Sharing mobility = mobilità condivisa (bike/
car/scooter sharing e altri tipi di mezzi in
sharing)

SIM = dispositivo a microprocessore per la
telefonia mobile, che consente di archiviare il
numero univoco associato a tutti gli utenti di
telefonia mobile

SIMPT = Sistema Informativo Monitoraggio
Pianificazione Trasporti

SiQual = Sistema informativo della Qualità
delle indagini

SIR = Sistema Integrato dei Registri

SISTAN = Sistema Statistico Nazionale

SLL = Sistemi Locali del Lavoro

Smart survey= Indagine statistica effettuata
attraverso dispositivi informatici di uso comu-
ne, generalmente APP dei cellulari mobili

SMC = Secure Multy-party Computation

SNAI = Strategia Nazionale per le Aree
Interne

SQL = Structured Query Language

TEE = Trusted Execution Environment

TGMA = Traffico Giornaliero Medio Annuo

t-SNE = t-distributed stochastic neighbor
embedding

TSS = Trusted Smart Statistics

TPL = Trasporto Pubblico Locale

TUS = Indagine sull'uso del tempo ISTAT

Vkm= Veicoli km

ZTL = Zona a Traffico Limitato

WPI = Popolazione abitualmente dimorante



In questo volume dei Technical Papers si presenta l'analisi svolta in ambito SISTAN da FS, Istat e MIT che indaga la possibilità di utilizzo dei Big Data, e in particolare di quelli telefonici, per studiare la mobilità delle persone.

fsitaliane.it