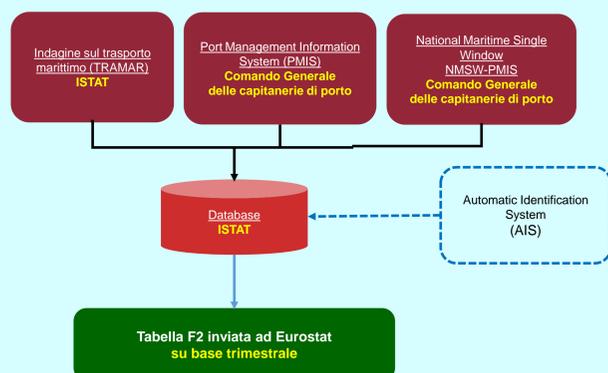


# AIS: una nuova fonte Big Data per le statistiche sul trasporto marittimo

LUCA VALENTINO - Istat luvalent@istat.it | NORINA SALAMONE - Istat norina.salamone@istat.it | ANGELA PAPPAGALLO - Istat angela.pappagallo@istat.it | MARCO DI ZIO - Istat dizio@istat.it  
FABRIZIO MARIA AROSIO - Istat arosio@istat.it | FRANCESCO PUGLIESE - Istat frpuglie@istat.it | FRANCESCO SISTI - Istat francesco.sisti@istat.it | MAURO BRUNO - Istat mbruno@istat.it  
GIULIO MASSACCI - Istat giulio.massacci@istat.it | FRANCESCO ORTAME - Istat francesco.ortame@istat.it | FRANCESCO ZARELLI - Istat zarelli@istat.it | SILVIA TALICE - Istat talice@istat.it  
AZZURRA SCIARRETTA - Istat sciarret@istat.it | ANNA EMILIA MARTINO - Istat annmarti@istat.it

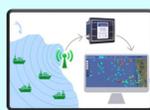
## Obiettivi

L'obiettivo dell'utilizzo della fonte Big Data, costituita dai dati satellitari AIS (*Automatic Identification System*) è disporre di una fonte aggiuntiva e complementare, rispetto alle fonti tradizionali, per ricostruire la popolazione di riferimento dell'indagine: tutte le navi in arrivo e in partenza nei porti italiani. Inoltre è una fonte di controllo e integrazione a supporto delle statistiche ufficiali sul trasporto marittimo. I dati AIS consentono, infatti, di migliorare sia la qualità che la tempestività dei dati trasmessi a Eurostat.



## Dati AIS

AIS è un Sistema di tracciamento automatico a bordo delle navi, utilizzato in ambito marittimo a scopo di sicurezza e gestione del traffico. I dati AIS costituiscono una fonte Big Data che fornisce, in tempo reale, la posizione della nave ed alcune altre informazioni relative all'identità della nave, alla navigazione e al viaggio che la nave sta effettuando.



Static data				Dynamic data				Travel related data			
IMO	MMSI	CALLSIGN	VESSEL NAME	VESSEL TYPE	TIME	COORDINATES	NAVIG. STATUS	SPEED	COURSE	DESTINATION	DRAFT
8401561	20110115	ZAD4L	FROJDI II	Cargo	04/08/2023 19:45	41.1323 16.8530	MOORED	0	258	Ravenna	null

Istat, come membro dell'*AIS Task Team on AIS Data of the UN Committee of Experts on Big Data*, può accedere ad un dataset globale (UN-AIS) di dati AIS, sia storici che real time, tramite la piattaforma UNGP (UN Global Platform). Questo data set raccoglie ogni 10 minuti osservazioni AIS di tutti i tipi di nave, con una stazza lorda di più di 300 tonnellate, a partire dal 1° dicembre 2018.



I dati nel dataset UN-AIS, oltre alle informazioni tipiche del dato AIS, forniscono anche la posizione delle navi tramite il sistema H3 (Hexagonal hierarchical geospatial indexing system <https://h3geo.org>). H3 è un Sistema di indicizzazione geo-spaziale, sviluppato da Uber Technologies, che approssima le coordinate GPS mediante esagoni, di varie dimensioni (che corrispondono a diverse risoluzioni), con cui viene ricoperta la superficie terrestre.



## Definizione Aree Portuali

Per valorizzare a fini statistici le informazioni disponibili occorre risolvere alcune criticità, quali la corretta localizzazione delle aree portuali. Innanzitutto, definiamo l'area di un porto come l'insieme degli esagoni H3, di risoluzione 8, associati alle coordinate geografiche del porto. Dopodiché, si raccolgono i dati AIS di un periodo di riferimento (e.g. 6 mesi del 2022), e si arricchisce l'area portuale con gli esagoni contenenti la posizione di navi che, più frequentemente, risultano ferme.



Esistono casi particolari, come gli attracchi al largo e piattaforme off-shore, in cui le navi risultano stazionare in mare aperto, talvolta a breve distanza dal porto più vicino. La conferma di ufficialità di tali aree viene dal confronto con le coordinate degli ancoraggi fornite da fonti ufficiali (Ministero delle infrastrutture e dei trasporti, Capitanerie di porto).



## Metodologia

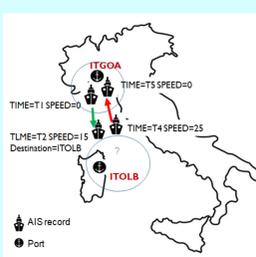
Questa procedura di elaborazione di dati AIS ha l'obiettivo di ricostruire tutti i viaggi con partenza o arrivo in un porto italiano. Le variabili fondamentali che definiscono il viaggio sono quelle mostrate nella tabella di esempio.

IMO	VESSEL TYPE	DEPARTURE PORT	ARRIVAL PORT	DEPARTURE DATE	ARRIVAL DATE
8401561	Cargo	ITBRI (Bari)	ITRAN (Ravenna)	04/09/2023	05/09/2023
9483712	Passenger	ITGOA (Genova)	ILOLB (Olbia)	05/09/2023	06/09/2023

Da questi micro dati è possibile poi ottenere statistiche come il numero di arrivi per porto. L'idea è che per ricostruire il viaggio di una nave sia sufficiente individuare la partenza e l'arrivo e dunque siano necessarie le sole osservazioni in cui la nave ha velocità 0 e si trova nell'area di un porto.



UNECE stesso avverte che la base dati AIS è affetta da errori che possono causare problemi all'algoritmo. In particolare si osservano periodi di tempo in cui la nave non ha osservazioni. Se in questo periodo c'è l'arrivo in un porto, il viaggio viene perso. Nell'esempio a sinistra la perdita dell'arrivo a ITOLB (Olbia) causa la perdita di due viaggi (ITGOA-ITOLB e ITOLB-ITGOA).



## Imputazione dati mancanti

Quando la mancanza di alcuni dati AIS non consente di individuare il porto di arrivo o di partenza di una rotta, questo può essere imputato. Esistono due possibilità per imputare, ad esempio, il porto di arrivo:

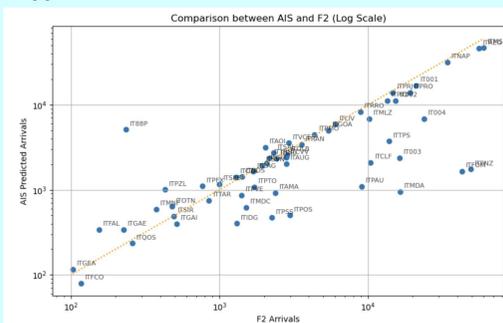
1. Usare l'informazione contenuta nel campo 'destinazione' dell'ultimo dato AIS disponibile della rotta;
2. predire il porto di arrivo mediante un algoritmo di *Deep Learning (DL)*.

Nell'ambito del DL, il miglior modello attualmente sperimentato per questo task risulta essere la rete convolutiva (CNN). Il modello, sfruttando la storia pregressa dei viaggi, è in grado di determinare il porto di arrivo più probabile anche in mancanza del segmento finale, come nei casi mostrati nelle figure a destra. Questa imputazione permetterà di ricostruire il viaggio anche per quelle rotte con lunghi tratti mancanti che l'algoritmo principale non è in grado di calcolare interamente (circa il 10% dei casi).



## Risultati

La principale tabella prodotta per Eurostat riporta il totale delle navi, che svolgono operazioni commerciali, arrivate nei 56 principali porti italiani: quelli che in un anno movimentano più di 1 milione di tonnellate di merci e/o 200.000 passeggeri. Il grafico sottostante, usando una scala logaritmica, confronta gli arrivi rilevati per ogni porto dall'indagine Istat (asse delle ascisse) con quelli ottenuti da AIS (asse delle ordinate) relativi all'anno 2022. I punti vicini alla diagonale rappresentano i porti per i quali il numero di viaggi rilevati da Istat e AIS coincidono o si discostano di poco.



Sotto la diagonale compaiono i porti per i quali gli arrivi di AIS sono inferiori a quelli Istat (soprattutto porti con viaggi brevi e frequenti). Sopra la diagonale si trovano molti porti collegati con le piattaforme offshore.

Un altro contributo dei dati AIS è stato quello di rappresentare il traffico di navi nelle piattaforme offshore, non facilmente rilevabile con le fonti tradizionali.

**Conclusioni:** l'esperimento di sfruttare i Big data a supporto delle fonti più tradizionali (sia amministrative che statistiche) è particolarmente utile e opportuno perché va a compensare le difficoltà e i limiti di queste nel riuscire a rappresentare in modo esaustivo un fenomeno estremamente dinamico, "fluidico" e incerto come il traffico marittimo, di cui non riescono a ricostruire con certezza l'universo effettivo di riferimento e le unità di analisi (le navi, i viaggi e le tratte).

**Prossimi passi:** Un ulteriore contributo informativo, che si potrebbe realizzare in futuro, è la modellizzazione georeferenziata delle tratte per la produzione di stime, utili nel caso di dati mancanti riferibili a tratte ricorrenti.

