

Roma, 12 maggio 2025

NUOVE FONTI PER LE STATISTICHE TERRITORIALI: RISULTATI E PROSPETTIVE

USO DI DATI AIS PER LE STATISTICHE SUL TRASPORTO MARITTIMO

NORINA SALAMONE – LUCA VALENTINO – ANGELA PAPPAGALLO

Istat | DIAE-DCAT – DIPS-DCME – DIPS-DCME

Argomenti trattati

- Normativa europea e rilevazione sul trasporto marittimo
- Fonti tradizionali e non: indagine diretta, dati amministrativi e Big Data (AIS)
- Primi risultati e aspetti critici sull'uso dei Big Data
- Metodologie di trattamento dei dati AIS
- Conclusioni

TRAMAR: un'indagine complessa

- La rilevazione sul trasporto marittimo è sotto **Regolamento UE n.1090/2010** (modifica Direttiva UE 42/2009)
- Ha **carattere censuario**: si riferisce a tutti gli arrivi e le partenze delle navi che approdano nei porti italiani
- Ha **carattere continuo**: è svolta senza soluzione di continuità, anche se restituisce dati trimestrali
- Deve fornire a Eurostat **dati con elevato dettaglio informativo**
- E' oggetto di **Quality Report** estremamente rigidi da parte di Eurostat (*coerenza interna, serie storiche, mirror check nazionale e internazionale*)

Principali output:

Dati trimestrali relativi a imbarchi e sbarchi per ogni singolo porto:

- **Peso lordo delle merci trasportate per tipo di carico** (*Merci in: rinfusa liquida, solida o altro; container distinti per unità di 20 piedi, 20-40, 40, >40; automezzi semoventi o non semoventi*) e **porto di carico** (per le merci in arrivo) o **porto di scarico** (per le merci in partenza)
- **Numero di unità per tipo di carico** (per container e automezzi)
- **Passeggeri**
- **Numero di navi arrivate** e **stazza lorda** distinti per **classe di stazza lorda** e **tipologia di nave** (*ad esempio: cargo specializzato/non specializzato, passeggeri, crociere, portacontainer, rinfusa, attività offshore, ecc.*) **Tabella F2**

Fonti alternative: Dati Amministrativi e *Big Data*

La raccolta di dati statistici si basa, per quanto possibile, sulle fonti disponibili, in modo da limitare l'onere per i rispondenti (Direttiva UE 42/2009 art. 3, comma 3).

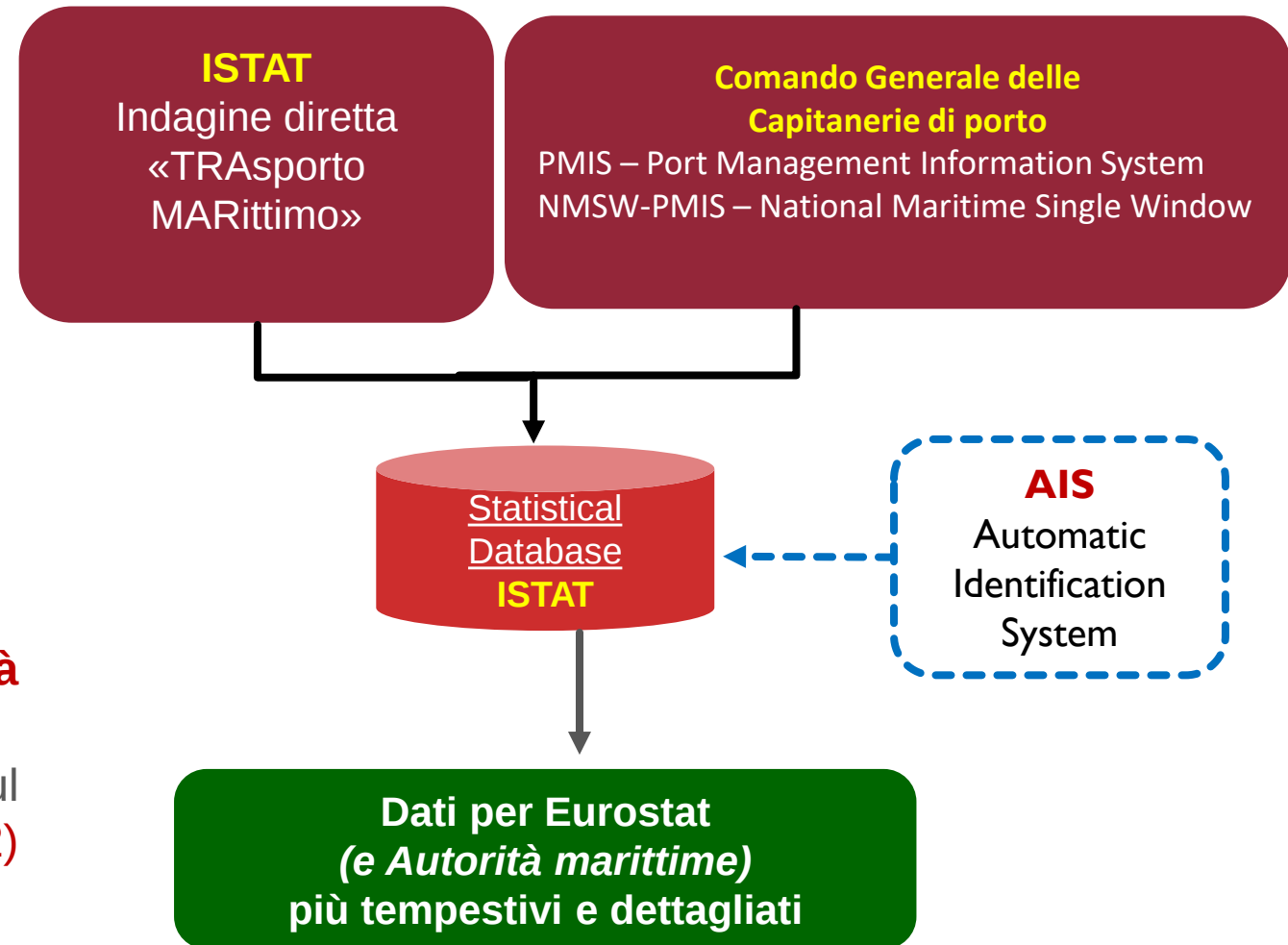
- **Fonti amministrative:** da anni si utilizzano i dati amministrativi delle Capitanerie di Porto (in passato venivano raccolti e inseriti manualmente in un database chiamato Ades, i «Modelli 61»), ora si fa riferimento ai sistemi informativi PMIS e NMSW-PMIS, del Comando Generale delle Capitanerie di porto, che l'Istat acquisisce in maniera automatica *Machine to Machine* (dal 2020 i 38 porti PMIS; dal 2023 70 porti del NMSW-PMIS).
- ***Big Data – AIS (Automatic Identification System)***: si tratta di un sistema di tracciamento automatico delle navi, utilizzato nel mondo marittimo principalmente per scopi di sicurezza per evitare le collisioni tra le unità in navigazione. I dati trasmessi, oltre a dare un allarme in caso di rotta di collisione, consentono alle autorità marittime di monitorare i movimenti delle navi.

Perché investire in una nuova fonte **Big Data**

- ✓ Perché sussistono ancora **disallineamenti** nelle informazioni del trasporto marittimo tra le diverse fonti amministrative e statistiche
- ✓ Perché una terza fonte rappresenta un necessario **riscontro** per dirimere incongruenze tra fonti statistiche e fonti amministrative
- ✓ Perché AIS fornisce informazioni estremamente **aggiornate** in *tempo reale* e **georeferenziate**

AIS è un'opportunità per migliorare la **tempestività** delle statistiche e la **qualità dei dati**:

1. Per la trasmissione a Eurostat dei dati sul numero di navi arrivate nei porti italiani (**Tabella F2**) con cadenza trimestrale;
2. Per la **ricostruzione delle rotte**.



Da dove siamo partiti: analisi a livello macro

CONFRONTO DATI AIS CON TABELLA F2 (NUMERO DI NAVI ARRIVATE NEI PRINCIPALI PORTI ITALIANI – IV TRIM 2021)

La Tabella F2 considera i **porti maggiori**: per merci (trasportano più di 1 milione di ton. di merci in un anno), o per passeggeri (trasportano più di 200.000 passeggeri in un anno), o per merci e passeggeri.

Per il IV trimestre 2021, da un primo confronto tra i dati AIS e quanto pubblicato nella tabella F2 per Eurostat, complessivamente abbiamo riscontrato una corrispondenza tra fonti del 66%

Un discreto punto di partenza, ma non ottimo.

Port name	Unlo Code	F2	AIS	Diff%	Port name	Unlo Code	F2	AIS	Diff%
Isola di Ischia	IT001	3686	3539	-4.0%	Napoli	ITNAP	7215	6261	-13.2%
Isola d'Elba	IT002	2489	1715	-31.1%	Olbia	ITOLB	479	53	-88.9%
Isole Egadi	IT003	3391	1238	-63.5%	Ortona	ITOTN	70	54	-22.9%
Isole Eolie	IT004	679	3759	453.6%	Palau	ITPAU	3607	2	-99.9%
Ancona	ITAOI	426	340	-20.2%	Porto Foxi (Sarroch)	ITPFX	205	172	-16.1%
Augusta	ITAUG	523	320	-38.8%	Piombino	ITPIO	2231	1760	-21.1%
Brindisi	ITBDS	321	122	-62.0%	Palermo	ITPMO	1000	807	-19.3%
Bari	ITBRI	494	331	-33.0%	Porto Nogaro	ITPNG	76	64	-15.8%
Cagliari	ITCAG	488	335	-31.4%	Ponza	ITPNZ	495	125	-74.7%
Chioggia	ITCHI	81	85	4.9%	Capri	ITPRJ	2072	2050	-1.1%
Carloforte	ITCLF	2638	1168	-55.7%	Procida	ITPRO	4338	2632	-39.3%
Calasetta	ITCLS	1353	656	-51.5%	Porto Santo Stefano	ITPSS	512	4	-99.2%
Catania	ITCTA	368	248	-32.6%	Porto Torres	ITPTO	324	116	-64.2%
Civitavecchia	ITCVV	588	467	-20.6%	Portoscuso (Porto Vesme)	ITPVE	1329	552	-58.5%
Falconara Marittima	ITFAL	43	52	20.9%	Pozzallo	ITPZL	241	206	-14.5%
Fiumicino	ITFCO	15		-100.0%	Oristano	ITQOS	96	35	-63.5%
Gaeta	ITGAE	96	51	-46.9%	Ravenna	ITRAN	853	195	-77.1%
Golfo Aranci	ITGAI	95	16	-83.2%	Reggio Calabria	ITREG	13263	8539	-35.6%
Gela	ITGEA	45		-100.0%	Sorrento	ITRRO	1032	776	-24.8%
Gioia Tauro	ITGIT	509	343	-32.6%	Salerno	ITSAL	588	375	-36.2%
Genova	ITGOA	1332	925	-30.6%	Siracusa	ITSIR	109	134	22.9%
Isola del Giglio	ITIDG	509	1	-99.8%	La Spezia	ITSPE	217	238	9.7%
Livorno	ITLIV	1428	854	-40.2%	Savona	ITSVN	430	781	81.6%
La Maddalena	ITMDA	3402		-100.0%	Taranto	ITTAR	265	140	-47.2%
Marina di Carrara	ITMDC	133	128	-3.8%	Trapani	ITTPS	3180	887	-72.1%
Milazzo	ITMLZ	827	1202	45.3%	Trieste	ITTRS	535	425	-20.6%
Monfalcone	ITMNF	108	107	-0.9%	Venezia	ITVCE	630	1827	190.0%
Messina	ITMSN	13449	8815	-34.5%	Totale		84908	56027	-34.0%

Primi risultati: differenze e punti critici

- Produzione della tabella 'arrivi nei porti' con dati AIS equivalente alla tabella F2 prodotta per Eurostat
- Confronto dei dati aggregati tra AIS e F2 e analisi delle principali differenze tra i due dataset di base per individuare casistiche di errori sistematici nell'elaborazione dei dati AIS
- Diversa copertura tra Eurostat e AIS delle unità di osservazione: **navi** in arrivo/partenza nei **porti**
- Confronto dei dati a livello micro tra AIS e TRAMAR per tipo di nave (a partire dalle navi da crociera)
- Problemi nei micro dati di AIS: **record incorretti e mancanza di dati**

Diversa copertura tra Eurostat e AIS

Si è reso necessario un lavoro di coerenza dell'universo di riferimento, partendo dalle richieste di Eurostat.

Classificazione delle navi e riduzione del campo di osservazione dei dati AIS al tipo nave richiesto da Eurostat (per esempio sono escluse dall'indagine le navi da pesca).

Classificazione dei porti, seguendo le definizioni di Eurostat per: **statistical port, main port, NAT. STAT. GROUP**

PRINCIPALI DIFFERENZE TRA EUROSTAT E AIS

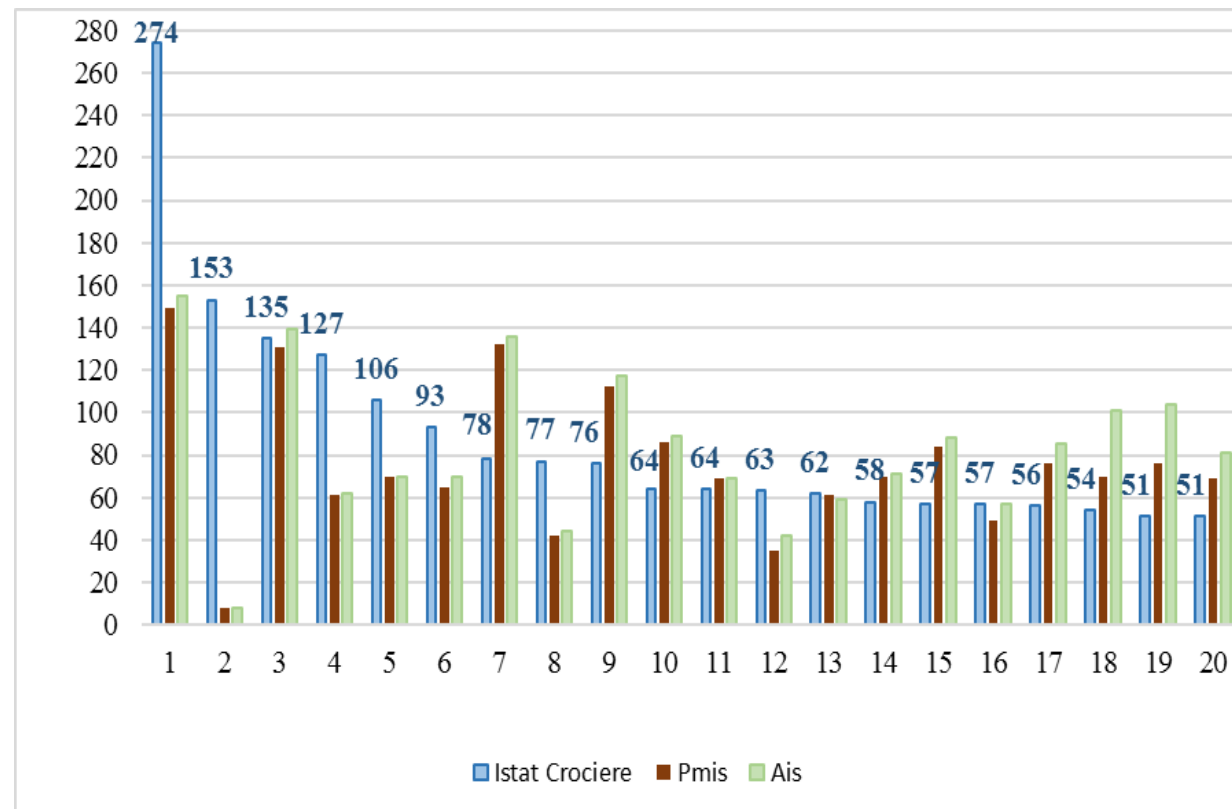
	EUROSTAT	AIS
NAVI	Sono obbligatori i dati delle navi commerciali (merci e passeggeri) con GT > 100 (Direttiva 2009/42/CE)	AIS è obbligatorio per le navi commerciali con GT>300 e per tutte le navi passeggeri , indipendentemente dalle dimensioni.
PORTI	Il porto statistico è costituito da uno o più porti, non vengono conteggiati i viaggi intragruppo (esempio Isole Eolie)	In AIS sono stati sommati gli arrivi di tutti i porti inclusi nella stessa aggregazione di porti (esempio Isole Eolie)

Uso dei dati AIS in produzione: caso delle crociere

Usare AIS come **fonte di controllo**

- **Anno 2023:** con i dati AIS per le crociere sono stati individuati facilmente i dati **sovrastimati** dalla nuova indagine sulle Crociere di fonte Istat (Anno 2023)
- **Anno 2024:** sono stati corretti altri duplicati di dati sulle crociere meno evidenti rispetto al 2023, con un *linkage* più puntuale che considera contemporaneamente *porto di approdo – Imo – data di arrivo della nave*

Navi da crociera: prime 20 navi con più toccate secondo il questionario Istat – Anno 2023



Uso dei dati AIS in produzione: coerenza sul porto di origine

Individuare il porto di origine della nave quando il dato delle altre fonti è contraddittorio

DIREZIONE	MESE	IMO	PORTO DICHIARANTE	PORTO ORIGINE DESTINAZIONE TRAMAR	PORTO ORIGINE DESTINAZIONE PMIS	PORTO ORIGINE DESTINAZIONE AIS	NOTE
A	11	11111111	ITCVV	ESBCN	ITPTO	ESBCN	a. AIS coincide con Tramar
A	10	2222222	ITSPE	ITGIT	ITTRS	ITTRS	b. AIS coincide con Pmis
A	12	3333333	ITSPE	ESVLC	FRFOS	FRFOS	b. AIS coincide con Pmis
A	10	4444444	ITRAN	TRGEM	EGEDK	EGEDK	b. AIS coincide con Pmis
A	12	5555555	ITGOA	MACAS	FRMRS	FRMRS	b. AIS coincide con Pmis
A	10	7777777	ITAUG	ITCAG	ITPFX	ITPFX	b. AIS coincide con Pmis
A	12	8888888	ITRAN	EGALY	ITVCE	ITPMA	b. stesso porto Pmis (ITPMA diventa ITVCE)

La fonte Big Data (AIS)

L’AIS è un sistema di tracciamento automatico a bordo delle navi, utilizzato a scopo di monitoraggio e sicurezza in mare.

Il singolo record AIS fornisce informazioni relative alle nave e alla sua posizione in un dato momento.

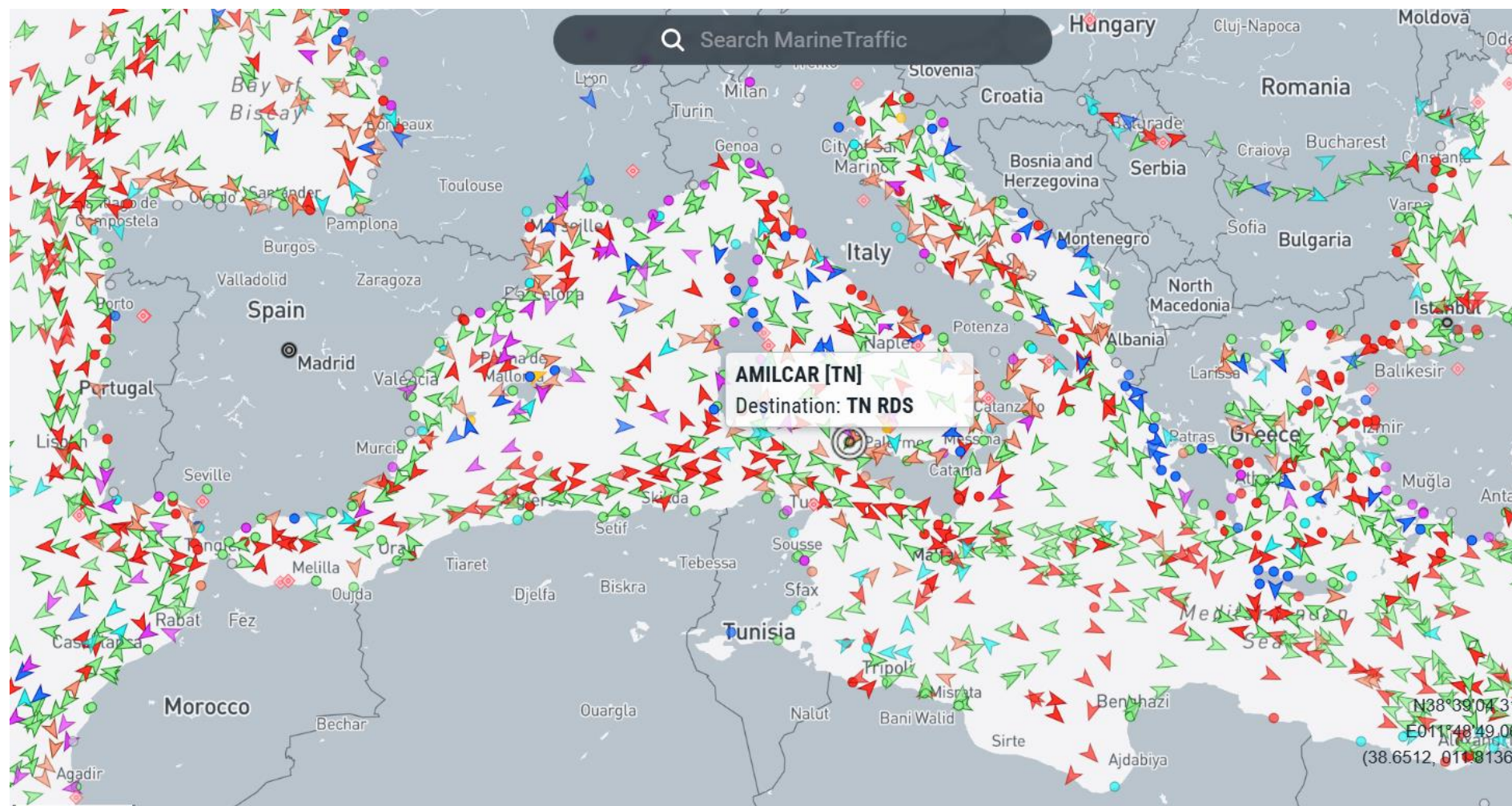
Più in dettaglio le variabili si dividono in dati statici (relativi alla nave), dati dinamici (posizione, velocità e direzione) e dati relativi al viaggio (inseriti dall’utente).

Il database è composto dalle osservazioni di tutte le navi ad intervalli di tempo più o meno regolari (ogni 10 minuti circa).



Static data					Dinamic 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	04/06/2023 19:45	41.1323 16.8530	MOORED	0	258	Ravenna	null

Una rappresentazione dei dati AIS: snapshot in tempo reale



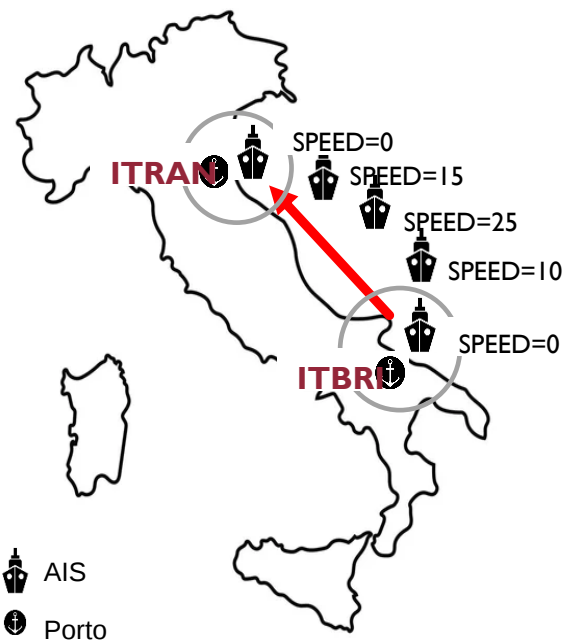
Trasformazione dei dati AIS in dati statistici

Le entità che abbiamo bisogno di ricostruire per utilizzare questi dati a fini statistici sono i viaggi di ogni singola nave.

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

Un viaggio è identificato da due eventi: la partenza e l'arrivo.
La trasformazione dei dati AIS in viaggi avviene attraverso le seguenti definizioni:

- Un evento di partenza è l'ultima osservazione della nave a velocità 0 in un'area portuale.
- Un evento di arrivo è la prima osservazione della nave a velocità 0 nell'area portuale successiva.



Problemi tecnici e pratici nell'uso dei dati AIS

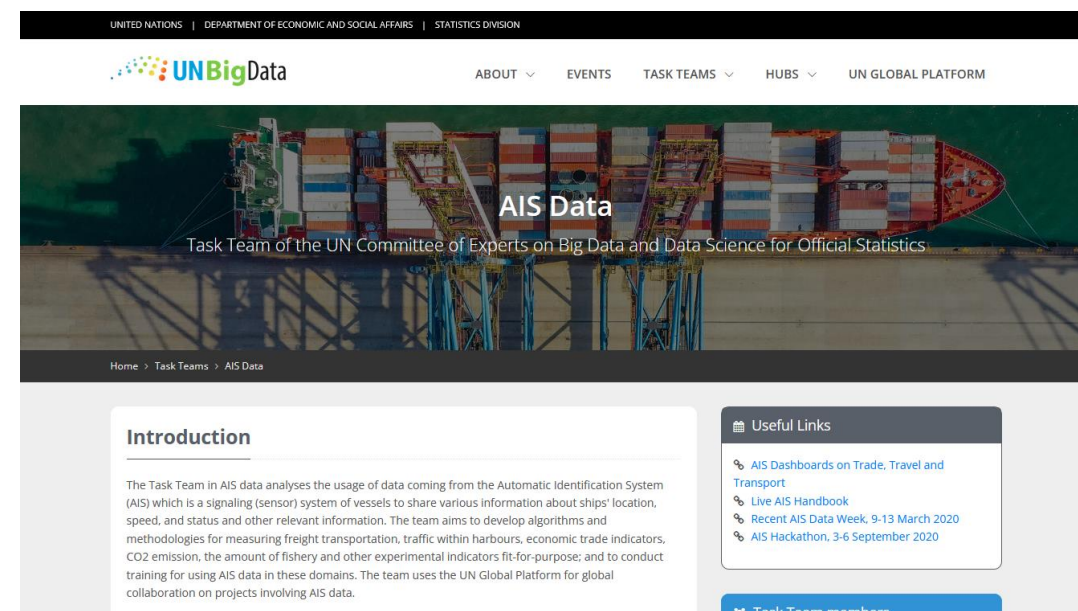
- **Disponibilità dei dati**

I dati AIS non sono disponibili a titolo gratuito, ma solo previo pagamento di una licenza.

Entrando a far parte dell'**AIS Task Team** delle **Nazioni Unite** per la prima volta Istat può accedere gratuitamente ad un data-set globale (UN-AIS) di dati AIS tramite la piattaforma **UNGP** (*United Nations Global Platform*).

- **Mole dei dati**

Il data-set AIS cresce alla velocità di oltre 1 milione di osservazioni l'ora (serie storica dal 2018). L'unico modo per trattare questi dati è usare piattaforme distribuite (*cloud*) e definire processi da eseguire in parallelo. La piattaforma UNGP si basa su SPARK.

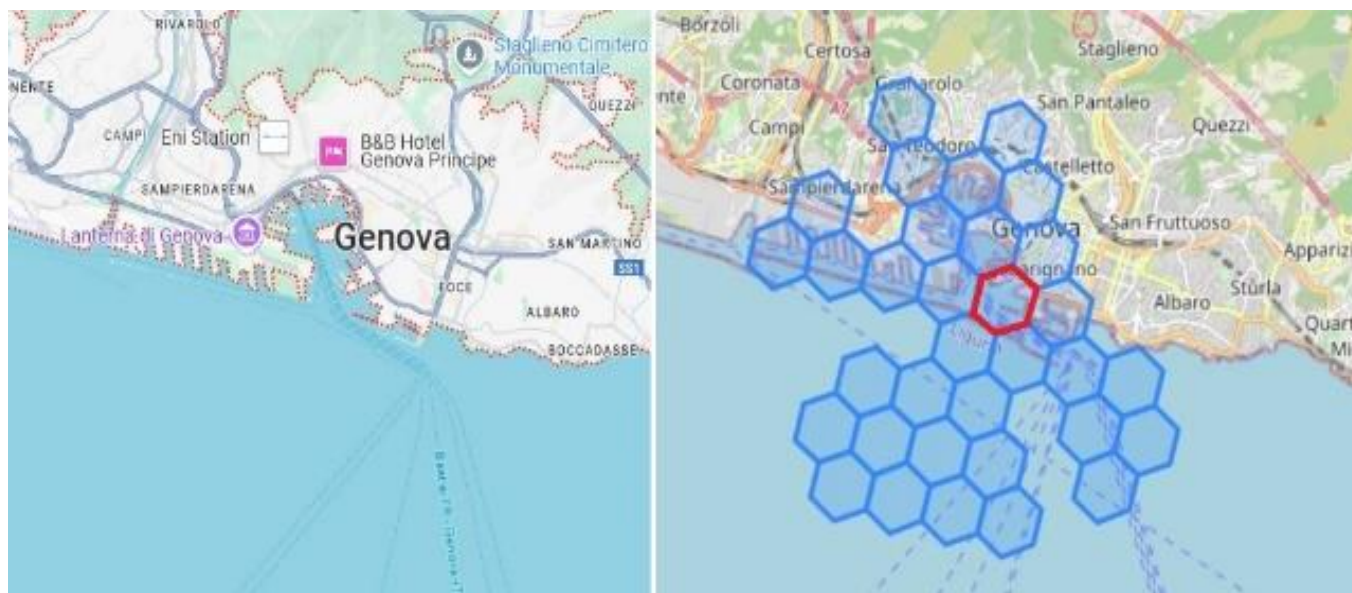


Definire le aree portuali

Per individuare una nave in un porto questo deve essere definito come un poligono geo-referenziato (in particolare usiamo il sistema H3 introdotto da Uber).

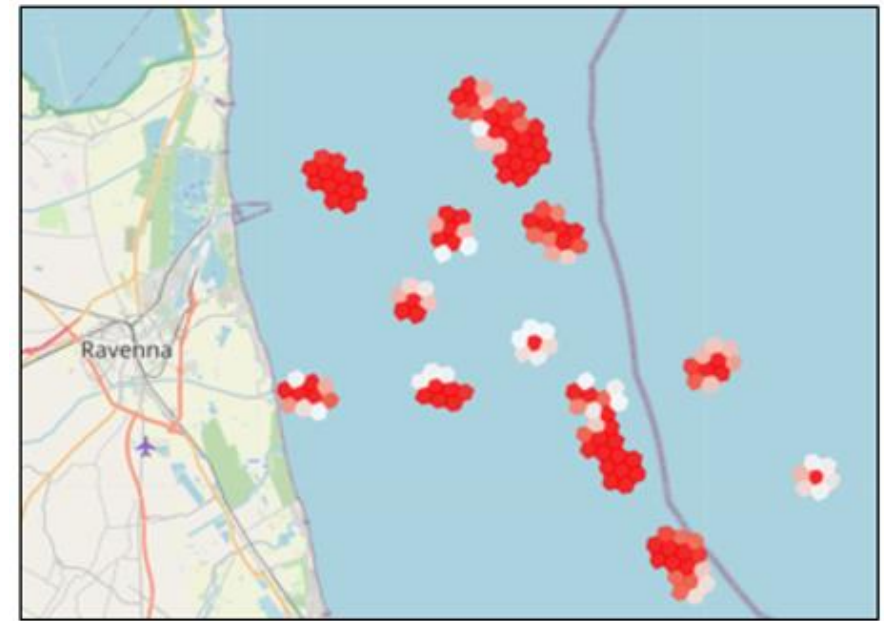
A partire dal punto centrale del porto (fonte DCAT per i porti italiani) si costruisce l'area portuale considerando tutti gli esagoni adiacenti in cui ricadono le coordinate delle navi stazionarie, secondo i dati AIS, in un periodo di riferimento (6 mesi).

Esempio:
il porto di Genova



Le piattaforme off-shore

- Con questa metodologia abbiamo individuato anche delle aree di stazionamento delle navi isolate in mare.
- Si tratta delle piattaforme off-shore che sono comunque previste e classificate da Eurostat con il codice generico IT88P.

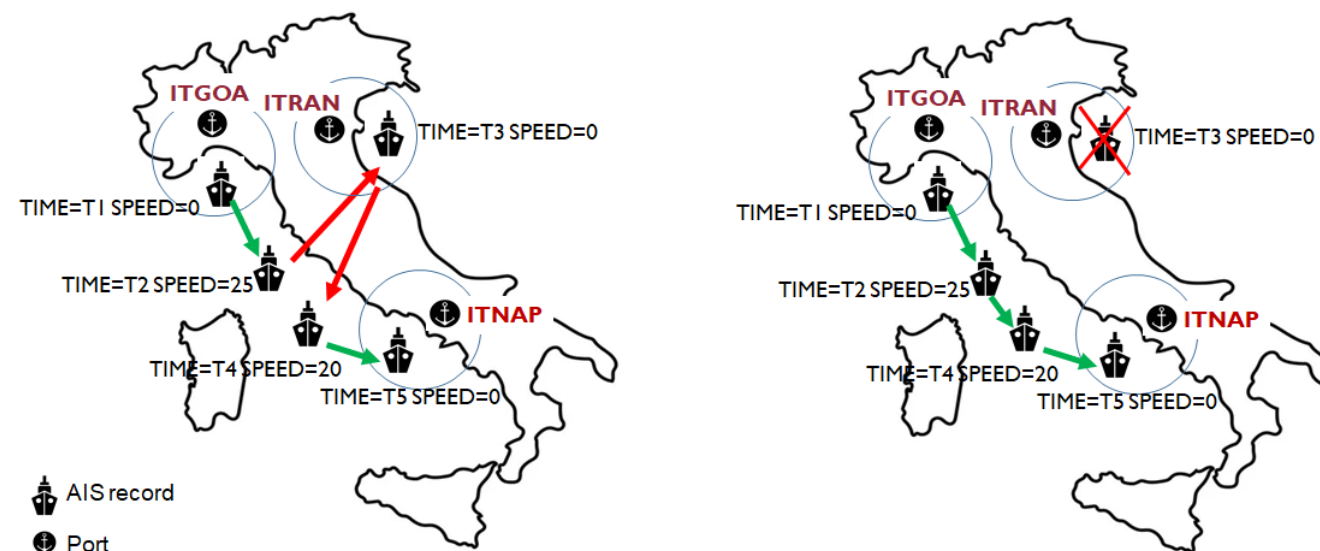
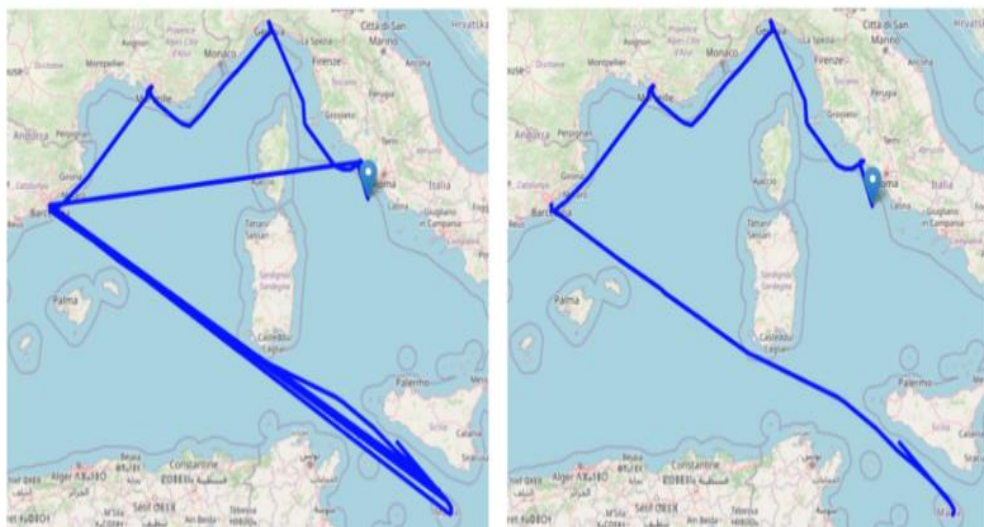


- Con AIS siamo stati in grado di mappare le singole piattaforme come normali porti e quindi ricostruire in dettaglio il traffico navale che le riguarda.

Problemi dovuti a dati errati

I dati errati sono osservazioni AIS che presentano un valore anomalo, ad esempio una posizione errata.

In questi casi il processo di trasformazione può generare dei viaggi mai avvenuti nella realtà.



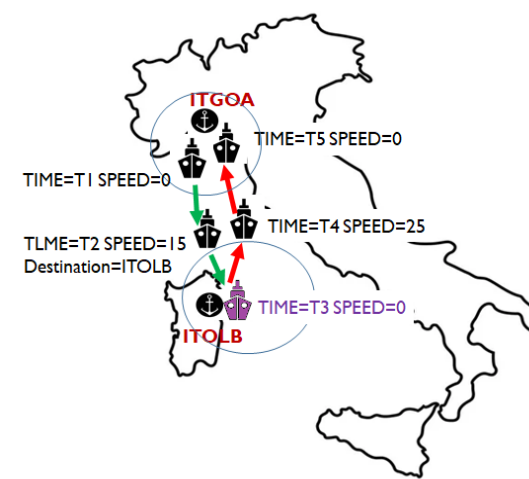
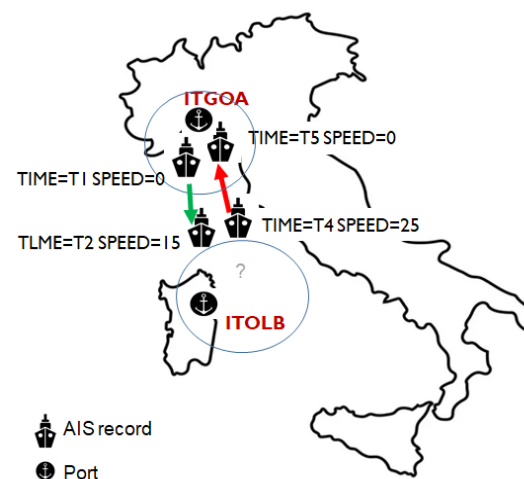
E' possibile individuare la maggior parte di questi errori confrontando ogni record col record successivo. Se la distanza percorsa non è compatibile con il tempo intercorso uno dei due record è errato e dunque entrambi vengono rimossi dalla rotta.

Problemi dovuti a dati mancanti

Un altro problema che si verifica con una certa frequenza è la presenza di «buchi» nei dati, ovvero periodi più o meno lunghi di tempo in cui mancano le osservazioni di una nave. Se durante uno di questi «buchi» la nave arriva in un porto perdiamo un evento di arrivo e quindi un viaggio.

Questo problema è affrontato come segue:

- analizziamo i «buchi», se supponiamo che ci sia stato un arrivo etichettiamo la rotta come «incompleta».
- Usiamo metodi di ML per imputare il porto di arrivo per le rotte incomplete.



Conclusioni

- L'uso delle fonti innovative, come i dati AIS, è un vantaggio in un'indagine complessa come il trasporto marittimo e serve sia per produrre stime più tempestive (per i dati provvisori trimestrali) sia come fonte di confronto aggiuntiva (per i dati definitivi).
- Dall'altro lato, nell'ambito della produzione delle statistiche ufficiali, non possiamo permetterci di sostituire completamente le fonti tradizionali: **la strada è dunque quella di usare la fonte Big Data a supporto delle fonti tradizionali.**
- Già oggi i dati prodotti con l'AIS sono utilizzati nella pipeline di produzione dell'indagine e hanno permesso di individuare anomalie e duplicazioni, e pertanto di migliorare la qualità delle statistiche prodotte.
- In prospettiva futura AIS può essere utile per:
 - individuare nuovi viaggi e comportamenti non descritti, o descritti solo parzialmente, nelle fonti tradizionali (caso del traffico da e per le piattaforme off-shore).
 - costruire modelli di rappresentazione del traffico marittimo di merci e passeggeri, basato sull'intero viaggio della nave e non solo sui singoli eventi di arrivo o partenza nei porti italiani.

Grazie

NORINA SALAMONE | norina.salamone@istat.it

LUCA VALENTINO | luca.valentino@istat.it

ANGELA PAPPAGALLO | angela.pappagallo@istat.it