

# **Documento di Progetto “Sperimentazione dell’utilizzo a fini statistici di fonti di telefonia mobile ai fini della stima dei flussi per tipologie utenti nella matrice origine/destinazione nell’ambito del progetto Persons&Places” (Scheda PSN IST-02589 “Uso a fini statistici dei Big Data”)**

19 novembre 2015

## **1. Obiettivo del progetto**

La finalità del lavoro è quella di sperimentare l’uso di dati anonimizzati di telefonia mobile in un processo statistico di stima di flussi di popolazione intercomunale: il test sarà condotto nell’ambito del progetto *Persons and Places* dell’Istat – IST.2661 sulla *Popolazione insistente e city users*, utilizzando i cosiddetti *Call Data Record* forniti dalla compagnia telefonica Wind, riferiti ad un territorio prefissato.

L’ipotesi allo studio è che l’integrazione informativa realizzata considerando le due fonti - dati da sistema informativo statistico e dati da traffico telefonico – possa, in generale, migliorare la qualità delle stime dei flussi intercomunali di popolazione e, in particolare, contribuire alla stima di altri dati quali l’incidenza del pendolarismo giornaliero sul totale dei flussi intercomunali.

Il progetto è incardinato nella Scheda del Programma statistico nazionale 2014-2016 - Aggiornamento 2015-2016 (codice IST-02589) e si inserisce nel filone dei lavori aventi come finalità lo sviluppo di nuove metodologie di supporto alla produzione nell’ambito della statistica ufficiale.

## **2. I dati utilizzati**

I dati di telefonia mobile, acquisiti presso WIND, sono i CALL DETAIL RECORD (CDR) relativi agli eventi *chiamate* effettuate da utenti WIND in un territorio predefinito e in un arco di tempo stabilito. I dati sono relativi al seguente insieme di variabili:

- Progressivo corrispondente all’utente che effettua le chiamate
- Codice comune all’interno del quale si trova la cella di effettuazione della chiamata
- Data e ora della chiamata

Sulla base delle definizioni di dato personale - art. 4, comma 1, lett. b del d.lgs. n. 196/2003 - e di dato anonimo - art. 4, comma 1, lett. n del d.lgs. n. 196/2003 -, si può affermare che i dati richiesti a WIND sono anonimi e non personali. Infatti, essi non permettono di individuare il singolo utente di telefonia mobile né direttamente (il progressivo è totalmente indipendente dall’identificativo dell’utente), né indirettamente (non sono disponibili informazioni che, opportunamente linkate, possano consentire un’identificazione indiretta del singolo utente).

In particolare, viene esplicitamente richiesto al gestore WIND di assegnare un codice ad ogni CDR e successivamente provvedere a cancellare la tabella di corrispondenza tra codici assegnati e identificativi dei CDR, prima di trasmetterli ad Istat.

Per meglio garantire l’impossibilità di identificazione indiretta, è stato chiesto a WIND di fornire, anziché le coordinate della cella di chiamata, il solo codice del comune cui la cella appartiene.

Per la realizzazione del progetto saranno, inoltre, utilizzati microdati provenienti dal Sistema informativo statistico dell’Istat denominato *Persons and Places* relativo alla *Popolazione insistente e city users*. In particolare, i dati relativi ai luoghi di residenza, di lavoro e di studio permetteranno di calcolare gli aggregati utili per valutare la qualità di quelli corrispondenti ottenuti dai dati di telefonia mobile. Ovviamente, il confronto sarà possibile solo a livello aggregato, non essendo disponibile per i dati di telefonia mobile alcuna informazione che permetta di stabilire corrispondenze a livello micro.

## **3. Acquisizione dei dati**

Per l’acquisizione dei CDR sarà inviata una formale richiesta a WIND nella quale il responsabile del trattamento, oltre ad illustrare le finalità del progetto in conformità con quanto indicato nel Programma statistico nazionale 2014-2016 – Aggiornamento 2015-2016, descriverà le misure di sicurezza adottate per garantire la riservatezza dei dati forniti.

#### 4. Elementi metodologici

I dati forniti da WIND verranno elaborati dall'Istat in primo luogo per procedere ad una verifica della metodologia già utilizzata nella ricerca condotta dal CNR-ISTI e dall'Università di Pisa, replicando le elaborazioni descritte nel contributo *“Use of mobile phone data to estimate mobility flows.”* pubblicato negli Atti del Convegno SIS 2014 di Cagliari allegato al presente documento come parte integrante dello stesso. Tale metodologia si basa sulla caratterizzazione dei singoli utenti di telefonia mobile mediante le indicazioni spazio-temporali (data, ora e comune) associate alle chiamate effettuate, in modo da stimare le componenti di interesse della popolazione “insistente” su un dato Comune (nella sperimentazione già effettuata da CNR-ISTI e Università di Pisa, per l'appunto il Comune di Pisa). Per tale motivo sono richiesti i dati relativi a quattro settimane di traffico telefonico WIND in provincia di Pisa nel mese di ottobre 2011 (dati già in possesso di CNR-ISTI e Università di Pisa). Una volta replicata la sperimentazione di cui sopra, i dati verranno utilizzati per valutare la possibilità di elaborazioni alternative. Come indicato anche nel parere del Garante per la protezione dei dati personali del 18/9/2014, *“... potrebbe rendersi necessario avere la disponibilità di altri set di microdati, opportunamente anonimizzati ... riguardanti altre zone e differenti periodi di riferimento”*.

Sono previsti report intermedi e un report finale sui risultati delle varie sperimentazioni. In ogni caso, particolare attenzione verrà posta a che in nessun caso vengano prodotte e diffuse stime di flussi aventi frequenze inferiori a tre unità.

#### 5. Gli attori coinvolti

**Titolare del trattamento dei dati** è l'Istat e **responsabile del trattamento** è il Direttore del Dipartimento Integrazione Qualità e Reti (DIQR).

**Incaricati del trattamento dei dati** verranno individuati all'interno del Dipartimento Censimenti e Archivi (DICA) e del Dipartimento Integrazione, Qualità e Reti, e definiti con apposita delibera.

**Compartecipante** al progetto è il CNR-ISTI. La compartecipazione si realizza tramite confronto sulle modalità della sperimentazione e sui risultati che verranno conseguiti. In particolare, verranno richieste informazioni sugli applicativi software utilizzati e la loro eventuale disponibilità per l'Istat. I soggetti compartecipanti non avranno accesso ai dati indicati al paragrafo 2. Ai compartecipanti saranno forniti esclusivamente i report prodotti dall'Istat che descrivono le attività di sperimentazione e i relativi risultati, nel rispetto dei vincoli posti dalla disciplina del segreto statistico.

#### 6. Misure fisiche per la sicurezza dei dati

I dati WIND oggetto di sperimentazione saranno memorizzati e custoditi su una base di dati Oracle attestata su server Istat. Tale server è ubicato nel CED dell'Istituto a via Balbo 16, ed è soggetto alle policy di sicurezza descritte nel [Documento annuale della sicurezza](#), richiamate brevemente a seguire.

Nella sede di via Balbo, che ospiterà i dati utilizzati per il progetto, al piano interrato dello stabile è individuata un'area controllata e riservata (Sala CED), situata in locali chiusi muniti di serratura meccanica, cui accede il solo personale interno munito di autorizzazione personale, che si autentica, ai varchi di accesso, attraverso il proprio *badge* identificativo. L'utilizzo delle uscite di sicurezza delle aree controllate e delle aree riservate è consentito solo in caso di effettivo pericolo.

Tutte le autorizzazioni all'accesso alle aree controllate e riservate sono rilasciate dal Dirigente il Servizio ETS della DCIT o da persona da lui delegata.

Non è consentita l'uscita di apparecchiature informatiche, accessori o parti di esse dalle aree interne se non espressamente autorizzata.

Tale ambiente è controllato nelle ore di chiusura dello stesso da un impianto di videosorveglianza; le immagini registrate sono raccolte su appositi dispositivi di registrazione digitale presenti in via Balbo nella centrale operativa al piano terra e nelle altre sedi di Roma nei locali di guardiania.

Tali immagini vengono distrutte entro le ventiquattro ore successive alla loro registrazione e, nei fine settimana e in caso di giorni festivi, entro le ventiquattro ore del primo giorno lavorativo successivo alle festività.

## **7. Misure logiche per la sicurezza dei dati**

I dati relativi ai CDR dei periodi e delle zone richiesti a WIND verranno trasmessi in modalità sicura (ossia con protocollo di trasferimento cifrato) e memorizzati in ambiente di database relazionale Oracle.

Per garantire la sicurezza di tali dati nell'ambiente Istat, si provvederà sia a crittografarli, sia a limitarne l'accesso solo alle persone formalmente incaricate del loro trattamento, mediante la definizione di opportune *userid* e *password* nel rispetto delle disposizioni dettate dal d.lgs. n. 196/2003.

Ogni azione sui dati da parte del personale sopra indicato sarà tracciata. Qualora sia necessario copiare i dati dall'ambiente relazionale Oracle ad altre piattaforme (esempio: postazioni di lavoro individuali), in ragione della possibilità di utilizzo di specifici sistemi di elaborazione, gli stessi saranno completamente rimossi dopo le lavorazioni.

## **8. Misure organizzative per la sicurezza dei dati**

Potrà accedere ai dati solo il personale dell'Istat individuato nel presente documento (par. 5 "*Gli attori coinvolti*"), espressamente incaricato dal responsabile del trattamento dei dati.

## **9. Misure informatiche a supporto della sicurezza dei dati**

Il database Oracle e la cifratura dei dati di cui alle misure logiche saranno opportunamente predisposti dal personale informatico della Direzione centrale per le tecnologie dell'informazione e della comunicazione (DCIT), garantendo il massimo livello di sicurezza e limitando al minimo i rischi di perdita accidentale di informazioni.

## **10. Durata del progetto**

Il progetto descritto ha durata di 36 mesi a partire da gennaio 2016 e si concluderà a dicembre 2018. Le misure descritte saranno naturalmente applicate per tutta la durata del progetto