

3 - 4 LUGLIO 2024

# La frontiera delle tecniche di indagine: le smart surveys

**Monica Perez**

Istat- Direzione centrale per la raccolta dati

**Claudia De Vitiis, Francesca Inglese**

Istat- Direzione centrale per la metodologia e il disegno dei processi statistici



Conferenza Nazionale di **Statistica**

**La statistica ufficiale  
nel tempo  
dell'Intelligenza  
Artificiale**

**#CNStatistica15**

- Cosa sono le smart surveys
- Potenzialità e problematiche connesse
- Sfide metodologiche, tecnologiche e organizzative
- Progetto ESSNet Smart Surveys, obiettivi e attività
- Conclusioni

# Cosa serve per realizzare una smart survey

1. un'indagine statistica, via web
2. un elemento 'smart'  
sensore per raccogliere dati, come immagini, posizione, spostamenti, etc.
3. l'integrazione tra l'indagine e l'elemento smart
  - durante la raccolta del dato, effettuata con **app** su **smartphone/tablet**
    - i sensori aiutano a svolgere i compiti richiesti al rispondente in modo più semplice
    - i dati del sensore sono processati e forniti tramite smartphone
    - i rispondenti possono interagire con i dati del sensore
  - dopo la raccolta del dato: e.g. Tracker di attività (Fitbits), con questionario dove sono riportate le info sulle attività (es.passi, attività cardiaca)



# Cosa sono le smart surveys

- ❑ Le smart surveys rappresentano una **forma ibrida di raccolta dati** poichè **combinano i dati offerti da sensori** (modalità passiva) **con l'attività del rispondente** (modalità attiva) **che agisce via web tramite un'app installata sul proprio smartphone/tablet**
  - ✓ mix tra tecnica (fonte) tradizionale e dati non-tradizionali
  - ✓ richiedono interazione -continua e di bassa intensità- dell'individuo con i propri dispositivi smart
  - ✓ il rispondente ha un ruolo centrale, deve consentire al sensore di acquisire i suoi dati personali
  - ✓ devono guadagnarsi la fiducia dell'intervistato, anche sul piano delle misure adottate a tutela dei dati personali
  - ✓ richiedono il coinvolgimento di soggetti terzi (cittadini - aziende private - enti pubblici) tra i quali si deve instaurare un rapporto di fiducia
  
- ❑ Nella statistica ufficiale, al momento le smart surveys trovano applicazione principalmente nelle indagini che prevedono la compilazione di diari come **l'indagine sulle Spese delle famiglie** (HBS) e **l'indagine Uso del tempo** (TUS), particolarmente onerose e svolte ancora oggi con questionari cartacei

# Quali sensori per le smart surveys

## ○ Tipologie di sensori per indagini smart:

- **Fotocamera** per scattare immagini (ad es. di scontrini per l'indagine sulle Spese delle famiglie)
- **Posizione - GPS** (ad es. per indagine sull' Uso del tempo)
- **Accelerazione** - accelerometro, giroscopio (tracker di attività per indagini sulla salute)
- Sensore (esterno) sulla **qualità dell'aria** (per la condizione abitativa)

○ I sensori offrono dati oggettivi per arricchire o sostituire le risposte fornite tramite questionario, possono ridurre l'errore umano introducendo però altre forme di errore, spostando per esempio i pregiudizi nella risposta (desiderabilità sociale) verso l'errore che può derivare dall'uso della tecnologia

○ Punto chiave è **identificare** i dati del sensore in grado di fornire la **misurazione più accurata** del fenomeno da indagare (dal dato grezzo alla variabile statistica)



5



- ❑ Per **elaborare i dati forniti dagli smart devices** - immagini, segnali gps, audio- e **trasformarli in informazioni statistiche**, corrispondenti a specifiche esigenze di classificazione in conformità agli standard del fenomeno osservato, si ricorre a tecniche di **Machine Learning (ML)**
- ❑ I dati acquisiti dal sensore sono **processati attraverso algoritmi di ML** in un microservizio (indipendente dalla piattaforma) e possono essere **restituiti al rispondente come suggerimenti** delle informazioni statistiche che il rispondente è chiamato a fornire per l'indagine

## ESEMPI

- **Indagine HBS**: le immagini degli scontrini fornite dal rispondente e acquisite dalla app vengono lette tramite tecnologia OCR (Optical Character Recognition) e tramite algoritmi di classificazione i prodotti dichiarati o acquisiti dalle immagini vengono ricondotti alle voci di classificazione COICOP (beni e servizi di spesa delle famiglie) → Al rispondente viene restituita l'informazione della spesa sostenuta per tipologia di prodotto che può confermare o correggere nella app che ha sullo smartphone
- **Indagine TUS**: gli algoritmi di ML possono essere utilizzati per supportare l'intervistato nella compilazione del diario delle attività quotidiane fornendo, attraverso la app usata per la compilazione, suggerimenti basati sulle attività svolte in base alla posizione (dati provenienti dal GPS abbinati a informazioni contestuali provenienti da servizi di mappe), oltre che informazioni sui luoghi visitati e sui mezzi di trasporto utilizzati per gli spostamenti

## □ Potenzialità

- ✓ Riduzione del **carico statistico** sui rispondenti
- ✓ Maggiore **accuratezza** delle informazioni
- ✓ Possibilità di **integrare** e, potenzialmente, anche **sostituire** la compilazione dei questionari, disponendo di dati ad **alta frequenza e molto dettagliati**
- ✓ Riduzione della **mancata risposta** nelle indagini sociali, se sono combinate con altri metodi di raccolta dati (tradizionali)



## □ Aspetti connessi all'uso delle app e dei dati da sensore

- Acquisire il **consenso** del rispondente (fornitore di dati) e coinvolgerlo come **soggetto attivo**
- Soltanto una parte di popolazione è raggiungibile (dispone degli strumenti, ha adeguate abilità per svolgere azioni smart o è disponibile a farlo) e, di conseguenza, può essere osservata (**copertura**)
- **Ridisegnare il processo** d'indagine, ricadute su GSBPM (raccolta dati, sistema di motoraggio, trattamento dati ML)
- **Garantire i requisiti di sicurezza dei dati** con un adeguato sistema di trasmissione, archiviazione e trattamento dei dati
- Ricorso a **nuovi metodi** per trasformare i dati acquisiti dai sensori in **informazione statistica** (OCR, ML, HITL, etc)
- **Ricadute sulla qualità** dei dati, **nuove forme di errore**, le stime si ottengono combinando dati da smart survey e tecniche d'indagine tradizionali
- Dare certezza giuridica agli utenti e ai cittadini (GDPR), trasferire al rispondente informazioni sulle tutele e garanzie adottate per la realizzazione della smart survey per aumentare la **trasparenza** e la **fiducia** verso chi gestisce i dati



## ➤ Le sfide metodologiche legate all'uso di app e di dati da sensori

- selettività dei partecipanti a causa dell'attitudine all'uso di strumenti smart, disponibilità dei devices, abilità tecniche
- indisponibilità dei rispondenti a fornire i dati dei sensori o a svolgere attività aggiuntive
- qualità e utilità dei dati offerti durante la raccolta (efficienza del microservizio)

## ➤ Elaborazione dei dati e accuratezza degli algoritmi di ML

- Utilizzo di algoritmi di Machine Learning per trasformare i dati dei sensori (immagini, segnali, voce) in informazioni statistiche
- Utilizzo di strategie/metodi ML appropriati per migliorare l'accuratezza della previsione

## ➤ La qualità dei dati

- Nuove tipologie di errori: errore di copertura/rappresentazione - Errore di misura - Implicazioni sull'effetto modale
- Strategie per prevenire/affrontare gli errori:
  - ✓ paradata
  - ✓ dati di contesto

## ❑ Errori di rappresentazione/selezione dovuti all'utilizzo di dispositivi mobili

### ▪ Mancata risposta

- ✓ Indisponibilità a partecipare, a scaricare e installare un'app, a utilizzare l'app attivamente o passivamente
- ✓ Indisponibilità a condividere i dati dei sensori a causa dell'incapacità dello smartphone di effettuare le operazioni e/o per problemi di privacy
- ✓ Indisponibilità a fornire il consenso di acquisire i dati tramite sensore

## ❑ Errori di misura dei dati da sensore

- ✓ Imprecisione del sensore (mancanza di accuratezza della misurazione, divario temporale, divario dispositivo)
- ✓ Errori di misurazione sistematici e casuali dovuti alla qualità del sensore (eterogeneità dei dispositivi)
- ✓ Anomalie nelle misurazioni (valori anomali, rumore, dati mancanti, ecc.)
- ✓ Comportamento degli intervistati (errata inizializzazione della misurazione o errato utilizzo dei dispositivi, qualità delle immagini, ecc)

**Errore totale**

# Sfide organizzative

- ❑ **Strategie per una raccolta dati efficace**, in grado di reclutare e motivare le persone e aumentare la partecipazione alle smart surveys
  - ✓ **Supporto di intervistatori (operatori smart)** almeno per alcune categorie di popolazione (anziani, persone con scarse abilità digitali, persone diffidenti per motivi di privacy e sicurezza dei dati) nel ruolo di facilitatori e/o supporto tecnico
  - ✓ **Forme di incentivi**
    - Fornire feedback informativo ai rispondenti (es. livelli e composizione delle spese; tempo trascorso per tipologia di attività)
    - Incentivi monetari (in uso in alcuni Paesi europei)
  - ✓ **Materiali di ingaggio** realizzati con testi sintetici, usando un linguaggio semplice e efficace, non tecnico
    - Lettera di invito, opuscolo
    - Video, tutorial
  - ✓ **Strategie e campagne di comunicazione** orientate a favorire la partecipazione ma anche il cambiamento culturale e l'accettazione delle nuove tecniche di indagine (smart)
  - ✓ **Interfaccia utente** intuitiva e semplice da usare
    - Integre dati passivi e attivi nella stessa interfaccia
    - trade-off tra dati passivi e dati attivi



Nell'ambito dello European Statistical System (ESS), progetti finanziati da Eurostat

- ❑ **ESSNet on Smart Surveys (2020-2022)**, un lavoro preparatorio per la creazione di quadro metodologico e architetturale per condividere e riutilizzare soluzioni e componenti smart utili a supportare gli INS nella realizzazione di smart surveys
  
- ❑ **ESSNet on Smart Surveys Implementation (SSI) 2023-2025**, un nuovo progetto che mira a implementare il quadro definito nel precedente progetto nel dominio delle indagini sociali, principalmente per le indagini sull'uso del tempo e sulle spese delle famiglie, attraverso soluzioni di processo end-to-end da testare sul campo, che tengano conto di:
  - Coinvolgimento dei cittadini come contributori attivi
  - Acquisire, trattare e combinare dati raccolti con app su smartdevice e dati acquisiti con sensori
  - Contribuire all'affidabilità del prodotto finale garantendo una forte tutela della privacy
  
- ❑ NSI partner: Olanda, Belgio, Italia, Germania, Francia, Norvegia, Slovenia
  
- ❑ Università: Utrecht, Mannheim, Vrije Bruxelles

# Il coinvolgimento dell' Istat

- ❑ **Definizione di standard per modelli di machine learning** all'interno del processo di rilevamento intelligente (raccolta ed elaborazione dei dati) e **ricerca di soluzioni** per raggiungere i livelli di precisione e accuratezza richiesti dalla statistica ufficiale
- ❑ Per l'indagine TUS si stanno definendo le **procedure metodologiche e i requisiti per lo sviluppo di un microservizio**, basato sui dati di geolocalizzazione, in grado di prevedere le attività del rispondente in relazione al tracciamento dei luoghi visitati (selezione di caratteristiche dai dati GPS e OSM per dedurre la posizione e un modello di ML per prevedere le attività dalla posizione)
- ❑ Due indagini sperimentali:
  - **Nuove modalità di raccolta dati**, indagine cross-nazionale sulla percezione e le attitudini della popolazione verso le smart surveys (svolta in Italia, Olanda e Slovenia) in cui ai rispondenti è stato chiesto di **svolgere azioni con i sensori** presenti sul proprio smartphone per la compilazione del questionario
  - **Indagine pilota Smart sull'Uso del tempo**, con app per valutare l'**effect-mode**
- ❑ **Aspetti legali del trattamento dei dati personali** nell'ambito delle smart surveys e l'elaborazione di linee guida per la realizzazione di valutazioni di impatto del trattamento dati (DPIA)

- **Smartphone e tablet**, utili alla realizzazione di smart surveys, sono devices **ampiamente utilizzati dai rispondenti**
- Gli **italiani al di sotto dei 65 anni**, con **livello di istruzione elevato**, che vivono nelle **regioni del Nord**, in **comuni al di sotto dei 50 mila abitanti**, in **famiglie con almeno due componenti**, presentano una maggiore **propensione a partecipare a una smart online survey**
- Le **preoccupazioni relative alla sicurezza dei dati e alla privacy** influiscono negativamente sullo svolgimento di azioni smart
- La **propensione a svolgere azioni smart** aumenta al crescere della **disponibilità di diversi tipi di smart devices** e del **reddito familiare**
- Le azioni smart legate alla cattura di immagini, riconducibili a temi specifici come **abitudini di spesa** e **consumi energetici**, ricevono una **adesione più limitata** da parte dei rispondenti rispetto alla condivisione della posizione e al conteggio dei passi

## SMART TASKS

Tra i partecipanti alla smart online survey (pari al 25,3% delle unità eleggibili):

Il **22,8%** accetta di condividere la propria posizione



Il **49,7%** dispone di uno scontrino della spesa.  
Di questi, il **17,8%** ne condivide l'immagine

Il **37,3%** dispone di un contapassi.  
Di questi, il **90,1%** condivide il conteggio dei passi



Il **69%** è a casa durante la compilazione del questionario online.  
Di questi, l'**11,6%** condivide la fotografia del contatore

# Considerazioni finali e prospettive future

- Le smart surveys hanno un grande potenziale e rappresentano una **grande sfida per la statistica ufficiale** a livello metodologico, tecnologico, organizzativo e comunicativo che deve essere affrontata attraverso un **approccio multidisciplinare** e attraverso un **patto di fiducia** tra i principali attori coinvolti: cittadini- Uffici di statistica- soggetti privati detentori di dati
- Le smart surveys **riducono l'onere statistico e i costi finanziari nella produzione statistica**, ma richiedono uno **sforzo iniziale maggiore** delle indagini tradizionali e la **diffusione di una cultura orientata all'uso di strumenti smart** anche per scopi statistici, che deve superare i limiti e le resistenze legati principalmente a questioni di **privacy e sicurezza dei dati**
- L'accettazione delle smart surveys va sostenuta anche attraverso adeguate **azioni e campagne di comunicazione** atte a (i) informare i rispondenti sugli **aspetti della sicurezza dei dati e della privacy** in modo semplice e concreto (ii) spiegare **lo scopo e l'utilità** di queste nuove modalità di raccolta dei dati, in modo da sviluppare **consapevolezza e fiducia**
- Prevedere forme di **supporto mirato** in fase di raccolta dati per (i) **favorire la partecipazione di segmenti di popolazione** più difficili da coinvolgere (anziani, individui con basso livello di istruzione, ecc.) (ii) **facilitare la disponibilità** dei rispondenti a fornire **informazioni in formato non tradizionale**

# Un po' di bibliografia

- Bähr, S., Haas, Georg-C., Keusch, F., Kreuter, F., & Trappmann M., 2020. "Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data". *Social Science Computer Review*. DOI: 10.1177/0894439320944118
- Benedikt, L., Joshi, C., Nolan, L., de Wolf, Nick. & Schouten, B., 2020. Optical Character Recognition and Machine Learning Classification of Shopping Receipts - @HBS>An app-assisted approach for the Household Budget Survey
- Biemer, P. P., de Leeuw, E., Eckman S., Edwards B., Kreuter T., Lyberg L. E., Tucker N. C. 2017. *Total Survey Error in Practice*. West B. T. (Editors). John Wiley & Sons, Inc., Hoboken, New Jersey
- ESSnet on Smart Surveys (2020-2021). <https://ec.europa.eu/eurostat/cros/content/essnet-smart>
- Immonen A., Pääkkönen P., and Ovaska E. 2015. "Evaluating the Quality of Social Media Data in Big Data Architecture". *Digital Object Identifier* 10.1109/ACCESS.2015.2490723
- Mussmann, Ole; Schouten, Barry (2019): Final methodological report discussing the use of mobile device sensors in ESS surveys. MIXED MODE DESIGNS FOR SOCIAL SURVEYS - MIMOD, WP5 deliverable. Edited by EUROSTAT.
- Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P.; Kreuter, F. 2019. "Willingness to Participate in Passive Mobile Data Collection". In *Public Opinion Quarterly* 83, pp. 210–235.
- Perez, M. Cipparrone, S., D'Amen, B., De Cicco, F., Lore', B., Nuccitelli, A., Porciani, L., 2024, " Primi passi verso le smart surveys nella statistica ufficiale". 15° Conferenza nazionale di statistica. <https://www.istat.it/evento/quindicesima-conferenza-nazionale-di-statistica/>
- Ricciato, F., Wirthmann, A., Giannakouris, K., Reis, F., Skaliotis, M. 2019a. "Trusted smart statistics: motivations and principles". *Statistical Journal of the IAOS*, 35. <https://ec.europa.eu/eurostat/cros/system/files/sji190584.pdf> 2)
- Ricciato, F., Giannakouris, K., Wirthmann, A., Hahn, M. 2020. "Trusted Smart Surveys: a possible application of Privacy Enhancing Technologies in Official Statistics". *SIS Conference*. [https://it.pearson.com/content/dam/region-core/italy/pearson\\_italy/pdf/Docenti/Universit%C3%A0/Pearson-SIS-2020-atti-convegno.pdf](https://it.pearson.com/content/dam/region-core/italy/pearson_italy/pdf/Docenti/Universit%C3%A0/Pearson-SIS-2020-atti-convegno.pdf)
- Struminskaya, B., Lugtig, P., Keusch, F., Hohne, J.K. 2020. "Augmenting Surveys with Data from Sensors and Apps; Opportunities and Challenges". *Social Science Computer Review*, 089443932097995

3 - 4 LUGLIO 2024

grazie  
per l'attenzione

**MONICA PEREZ**

Dirigente di ricerca| Istat



Conferenza Nazionale di **Statistica**

**La statistica ufficiale  
nel tempo  
dell'Intelligenza  
Artificiale**

**#CNStatistica15**