

I sensi della ricerca: interdisciplinarietà e temi di frontiera negli studi di popolazione

Sessione “fonti e tecniche per l’interdisciplinarietà”

IL PASSAGGIO DALLA STATISTICA PER INDAGINI ALLA STATISTICA PER FONTI: IMPLICAZIONI DI ORDINE METODOLOGICO E CONSEGUENZE DI TIPO PRATICO

Alessandro Valentini*

Abstract

Fino a pochi anni or sono, al nascere di una esigenza conoscitiva non soddisfatta dalle statistiche esistenti corrispondeva molto spesso la realizzazione di una nuova indagine. L’accrescersi delle richieste, sempre più raffinate sia dal punto di vista tematico che per quanto concerne il riferimento territoriale, l’emergere di situazioni di fastidio statistico verso i rispondenti e la riduzione delle risorse economiche a disposizione hanno indotto i produttori dei dati statistici alla ricerca di nuove modalità di realizzazione delle indagini.

In questo scritto, in particolare, si intendono approfondire alcuni aspetti metodologici delle statistiche per fonti: dall’impatto sulla qualità della rilevazione alle modalità di espletamento delle diverse fasi. Vengono poi presentati alcuni casi tratti dal Programma Statistico Nazionale di integrazione tra la statistica per fonti e la statistica per indagini. Infine sono discussi due esempi concreti di indagini per fonti: il primo relativo alla stima delle principali misure fisiche per i cittadini residenti nel Comune di Lucca tramite l’elaborazione dei dati registrati al momento dell’emissione (o del rinnovo) della Carta di Identità; il secondo contenente alcune proposte per l’aggiornamento in via amministrativa di un Sistema Informativo Territoriale impiantato utilizzando le risultanze elementari dei Censimenti Generali degli anni 2000.

1. Introduzione

L’accrescersi, presso gli operatori istituzionali e più in generale presso la collettività, delle esigenze conoscitive relative a molteplici fenomeni sociali, demografici ed economici ha comportato, nel corso degli anni, una fortissima espansione nel numero e

* Ricercatore Istat, Istituto Nazionale di Statistica

nell'estensione territoriale delle indagini statistiche. Questo ha sicuramente garantito una migliore conoscenza di tali fenomeni non soltanto nelle aree di più grandi dimensioni, ma anche in contesti geografici sempre più fini, fino a pochi anni or sono nemmeno presi in considerazione dai tradizionali piani di diffusione.

A fronte di questi innegabili vantaggi vi è stata, tuttavia, un'espansione notevole sia nel numero di potenziali rispondenti che nella frequenza con la quale ad ogni singola famiglia o impresa è stato sottoposto un questionario. Il contemporaneo ridursi delle risorse economiche a disposizione delle Pubbliche Amministrazioni per lo svolgimento di molteplici attività, e di quella statistica in particolare, ha fatto sì che anche nell'ambito della Statistica Ufficiale si iniziassero a sperimentare nuovi modi di fare indagini. Tra questi una particolare attenzione è stata riservata alle c.d. indagini *per fonti* (o amministrative), ovvero a quelle analisi che utilizzano i giacimenti informativi disponibili presso le pubbliche Amministrazioni.

Nel presente lavoro, dopo un breve richiamo ad alcuni aspetti metodologici delle statistiche per fonti (Par. 2), si esamina l'impatto del passaggio dalle statistiche per indagine alle statistiche per fonti all'interno delle diverse fasi in cui si struttura un'indagine statistica (Par. 3). Si presenta poi lo stato dell'arte delle rilevazioni amministrative nell'ambito delle esperienze del Sistema Statistico Nazionale, con particolare riguardo alle fonti di tipo demografico (Par. 4). A supporto delle argomentazioni teoriche si presentano poi due diversi esempi di indagini per fonti proposti dall'Istat – Ufficio Regionale per la Toscana. Il primo (Par. 5) riguarda l'utilizzazione delle registrazioni effettuate dall'Ufficio di Anagrafe del Comune di Lucca per tutte le persone che hanno rinnovato la Carta di Identità, con la finalità di ottenere una stima delle principali misure fisiche dei cittadini residenti (altezza, colore degli occhi e dei capelli). Il secondo esempio (Par. 6) concerne una serie di studi di fattibilità volti a dimostrare le possibili modalità di aggiornamento in via amministrativa di un sistema informativo territoriale impiantato in una realtà di piccole dimensioni utilizzando le risultanze elementari dei Censimenti Generali degli anni 2000.

2. Aspetti metodologici delle statistiche per fonti

Si definiscono dati amministrativi le informazioni raccolte e conservate da un operatore (ente pubblico o società) per finalità istituzionali dell'operatore stesso. E' questo il caso, per esempio, dei dati sui trasferimenti di residenza o sulle licenze edilizie raccolti

presso un Comune; degli elenchi dei clienti di una certa ditta tenuti con la finalità della susseguente emissione della fattura, ecc... In tutte queste situazioni la raccolta e la successiva elaborazione delle informazioni sono pianificate con l'obiettivo di raggiungere specifici scopi amministrativi, relativi alle *single* unità oggetto di rilevazione. Anche in un'indagine statistica i dati vengono raccolti presso le singole unità di rilevazione, ma non per essere trattati singolarmente. Soltanto per estrarre da essi alcuni indicatori sintetici. In altri termini l'elemento identificativo dell'unità (se presente) è soltanto funzionale agli scopi dell'indagine statistica. Per esempio serve per ricontattare successivamente il rispondente nel caso di indagini panel, di indagini di copertura o di controllo, o per costruire archivi di unità. Il dato raccolto tramite un'indagine statistica non potrà mai essere utilizzato per identificare nuovamente l'unità¹.

Gli stessi dati individuali di un'indagine amministrativa possono essere oggetto di un ulteriore trattamento, che prescinde in questo caso singoli individui ma che si concreta invece nell'ottenimento di indicatori sintetici (distribuzioni di frequenza, medie, rapporti, totali, ecc...) rappresentativi dell'intera collettività oggetto di rilevazione. Quest'ultimo tipo di trattamento procede in parallelo rispetto alle finalità proprie per le quali i dati sono stati rilevati. Ad esso è possibile associare la denominazione di indagine per fonti (o amministrativa).

In termini più formali (Fortini, 2002) definiremo le indagini amministrative come quelle indagini che utilizzano dati amministrativi con finalità di tipo statistico. Se il *prodotto* di un'indagine amministrativa è lo stesso di quello di un'indagine statistica, le differenze tra le due modalità di fare indagine possono essere osservate lungo il *processo* di produzione dei dati, ovvero nelle diverse fasi in cui si struttura l'indagine statistica. Tra i vari elementi oggetto di osservazione (come vedremo anche nel Paragrafo successivo) una particolare attenzione dovrà essere posta al diverso impatto delle indagini statistiche e di quelle amministrative sulle varie dimensioni della qualità (Colombo, 1983; Istat, 1989f): la rilevanza, l'accuratezza, la tempestività e puntualità, l'accessibilità e chiarezza (trasparenza), la confrontabilità, la coerenza e la completezza.

Si intende per *rilevanza* la capacità dell'informazione di soddisfare le esigenze conoscitive degli utenti. Per *accuratezza* il grado di corrispondenza fra la stima ottenuta

¹ Come sancito dall'Art. 6 bis del D.lsg 322/89. L'unica eccezione a questa regola generale riguarda i risultati del Censimento della Popolazione, che sulla base di quanto indicato dal Regolamento Anagrafico (Art. 46 DPR 223/89) servono per revisionare le anagrafi sulla base delle modalità tecniche stabilite dall'Istat (per il Censimento del 2001 si veda la Circolare Istat n. 15/2001) . Per una completa raccolta della normativa relativa alla statistica ufficiale si veda il Codice della Statistica Ufficiale (Istat, 2000 e 2002)

dall'indagine e il vero (incognito) valore della caratteristica in oggetto nella popolazione obiettivo. La *tempestività e puntualità* fanno riferimento all'intervallo di tempo intercorrente fra il momento della diffusione dell'informazione prodotta e l'epoca di riferimento della stessa. La *trasparenza* consiste nella semplicità per l'utente di reperire, acquisire e comprendere l'informazione disponibile in relazione alle proprie finalità. La *confrontabilità* è la possibilità di effettuare paragoni nel tempo e nello spazio. La *coerenza* la possibilità di combinare inferenze semplici in induzioni più complessi (nel caso di fonte unica) e l'utilizzo di definizioni, classificazioni e standard metodologici comuni (nel caso di più fonti). La *completezza* infine consiste nella capacità dei vari processi di integrarsi per fornire un quadro informativo soddisfacente del dominio di interesse.

Qualora emerga il fabbisogno di dati quantitativi non soddisfatto dalle statistiche esistenti si pone il problema dell'identificazione della modalità più idonea per soddisfare tale esigenza. In particolare si dovrà valutare l'opportunità di impiantare una vera e propria indagine contro quella di utilizzare archivi o fonti amministrative già esistenti (ammesso che l'obiettivo conoscitivo ci consenta di farlo). Al momento dell'attivazione del processo decisionario sarà necessario tenere conto di vari aspetti degli archivi amministrativi, che possono essere raggruppati in due categorie: il grado di sostituzione rispetto all'indagine statistica e le caratteristiche intrinseche dell'archivio.

In merito al *grado di sostituzione* è necessario valutare di quanto il campo di osservazione dell'indagine statistica si differenzia rispetto all'indagine da realizzare, quali sono le differenze nelle unità di osservazione e di rilevazione, nelle classificazioni adottate, negli standard di riferimento, ecc... In termini più generali (Istat, 1998a) le informazioni contenute negli archivi amministrativi possono essere utilizzate a fini statistici solo dopo aver completato una lunga e complessa sequenza di operazioni, tese a trasformare tali informazioni amministrative in informazioni statistiche, modificando i concetti e le definizioni proprie delle singole amministrazioni (o aziende) in concetti e definizioni statistiche coerenti con gli obiettivi dell'indagine e con quelle utilizzate a livello nazionale e internazionale.

Per ciò che concerne le *caratteristiche intrinseche* dell'archivio, bisogna considerare la capacità di acquisizione dei dati in funzione delle finalità dell'indagine. A tal riguardo le principali variabili da valutare sono: il grado di centralizzazione degli archivi amministrativi, le caratteristiche tecnologiche del sistema amministrativo, le modalità di acquisizione delle informazioni. Sul grado di centralizzazione è necessario distinguere il caso in cui esiste un unico archivio organizzato e gestito presso un operatore

centrale dalla situazione in cui vi sono più archivi sul territorio, ciascuno relativo alle unità di competenza geografica. Esempi del primo tipo riguardano i dati dell'INPS, della Telecom, del Ministero delle Finanze. Esempi del secondo tipo sono i dati anagrafici tenuti presso i singoli Comuni.

Le caratteristiche tecnologiche del sistema amministrativo impattano sull'indagine nel senso che le problematiche da affrontare sono enormemente differenti a seconda del livello di organizzazione delle pratiche amministrative. E' infatti ben diverso il caso in cui i dati sono già tutti registrati su supporto informatico rispetto al caso in cui si tratti di documentazione cartacea: nel primo caso si salta completamente la fase di registrazione delle informazioni, con evidenti benefici in termini di tempo e di risorse impiegate, nel secondo è necessario – al pari di una qualsiasi indagine statistica – predisporre gli strumenti e il personale per la registrazione dei dati.

In merito alle modalità di acquisizione delle informazioni bisogna distinguere la situazione in cui si raccolgono informazioni relative ad ogni unità statistica da quella in cui sono invece raccolti dati di sommario su insiemi di unità. Con riguardo alla rilevazione sui decessi, un esempio del primo tipo riguarda la registrazione del modello relativo ai singoli individui, un esempio del secondo concerne invece la compilazione, da parte di ogni Comune, di una scheda riepilogativa in cui è riportato il numero mensile di decessi per sesso.

Sulla base di quanto osservato fino ad ora sembrerebbe meglio soddisfare qualsiasi esigenza conoscitiva (fatte salve situazioni del tutto particolari) per mezzo di un'indagine statistica invece che dover accettare perdite di qualità in termini di rilevanza, accuratezza, confrontabilità e coerenza accettando di reperire i dati da un archivio amministrativo. Invece il problema è più complesso di quanto sembri: intanto bisogna considerare che un'indagine amministrativa consente di ottenere i risultati sicuramente in maniera più tempestiva (i dati sono già disponibili, senza la necessità di perdere tempo per doverli raccogliere). Poi il livello di trasparenza e il requisito di completezza di un'indagine amministrativa non necessariamente risultano peggiori di quelli di una rilevazione statistica. Vi è oltretutto da considerare che non necessariamente l'accuratezza di un'indagine amministrativa risulta peggiore di quella di un'indagine statistica, specie quando ai potenziali rispondenti vengono propinati in tempi ristretti più questionari. In effetti bisogna sempre tenere in considerazione il rischio di indurre la sensazione di fastidio statistico, fattore che comporta la drastica riduzione nella voglia di collaborare e/o nel grado di veridicità delle risposte fornite. In ultimo, specie in questi anni in cui le risorse

economiche a disposizione delle varie Amministrazioni sono sempre più scarse, è necessario non sottacere un problema di costi vivi da affrontare nel caso di indagine diretta (in particolare per la stampa dei modelli e per i compensi agli organi di rilevazione), costi sicuramente molto più alti rispetto a quelli da sostenere nel caso di utilizzo degli archivi amministrativi. Nelle indagini amministrative, infatti, gli oneri da sopportare sono nulli nel caso in cui l'archivio sia già disponibile in formato digitale; possono diventare significativi se l'archivio è soltanto cartaceo e di conseguenza è necessario provvedere in ogni caso alla fase di registrazione.

La scelta tra indagine statistica e rilevazione amministrativa dovrà allora essere compiuta caso per caso, cercando ogni volta di realizzare una sorta di compromesso tra tutti gli elementi sopra citati. In particolare sarà opportuno stabilire un *trade off* tra economicità e livello qualitativo dei risultati.

3. Differenze tra le indagini statistiche e quelle amministrative nelle diverse fasi del processo produttivo del dato

Una volta deciso di realizzare un'indagine (sia essa statistica o amministrativa) si pone il problema di come implementarla. A tale riguardo l'indagine stessa può essere assimilata (Istat, 1989f) ad un processo produttivo manifatturiero in quanto – come questo – è composta da un insieme di fasi ed operazioni interrelate: la materia prima è costituita dai dati disponibili in relazione alle varie unità di analisi. Il prodotto finito dall'informazione statistica. Il flusso di produzione dell'indagine può essere scomposto nelle fasi di progettazione, rilevazione, registrazione su supporto informatico, revisione e codifica del materiale grezzo, elaborazione, validazione e diffusione dei risultati. Le specifiche delle varie fasi – come ovvio – sono molto diverse nel caso di indagine amministrativa rispetto alla classica indagine statistica.

La fase della *progettazione* (Istat, 1989a , 1989b), in particolare, è il momento in cui si decide quali sono gli obiettivi, il campo di osservazione, le unità di rilevazione e di analisi, quali sono le variabili da rilevare. Nell'ambito della progettazione si pianificano poi tutte le altre fasi dell'indagine, a partire dalla rilevazione dei dati, avvalendosi – possibilmente – degli strumenti di progettazione concettuale (Barcaroli e al., 1987). E' in questo segmento del processo produttivo che si devono attivare tutti i controlli di qualità e i piani di lavoro. Nel caso di una indagine statistica esiste la massima libertà di scelta delle

variabili di analisi, delle definizioni, delle classificazioni, ecc, in modo tale da poter soddisfare completamente le esigenze del ricercatore; in un'indagine amministrativa, invece, di fatto si subiscono le definizioni, le classificazioni e le variabili stabilite all'interno del procedimento amministrativo. Sarà soltanto possibile escludere le unità non di interesse dal campo di osservazione (effettuando un'operazione di filtraggio) ma ovviamente non risulterà possibile disporre di informazioni esterne al procedimento amministrativo. Questo a meno di non effettuare operazioni di *record linkage*, ovvero di aggancio dei microdati dell'archivio amministrativo con microdati disponibili da altri archivi. L'aggancio avverrà tramite il codice identificativo delle unità (generalmente il Codice Fiscale o la Partita Iva) utilizzando criteri di tipo probabilistico per quelle unità laddove la corrispondenza del codice non è perfetta. Per dettagli tecnici sui metodi statistici per il record linkage si veda Istat (2003).

Oltre che nella fase di progettazione, deviazioni evidenti tra indagine amministrativa e statistica si riscontrano anche nella fase di *rilevazione* dei dati. Se in un'indagine statistica, una volta decisi i piani di lavoro, è poi necessario raccogliere i dati, scegliendo anche le tecniche di rilevazione più idonee (Istat, 1989c), nelle indagini amministrative i dati (di solito) sono già disponibili. Anzi l'esame dei tracciati di registrazione, delle procedure e dei disciplinari seguiti è un elemento fondamentale da conoscere in fase di progettazione dell'indagine. Questo a meno che l'indagine statistica non sia stata progettata insieme a quella amministrativa, come è per esempio il caso della rilevazione Istat sui trasferimenti di residenza (in cui il modello utilizzato ha valenza sia statistica che amministrativa), o che l'iter amministrativo preveda anche la compilazione del modello statistico (come nel caso delle pratiche edilizie).

In altri termini la fase di rilevazione dei dati nelle indagini amministrative non può essere in alcun modo controllata dal processo produttivo del dato statistico, risultandone completamente esogena. Ciò rende impossibile l'intervento nel merito della qualità delle informazioni raccolte (in termini di accuratezza, completezza, tempestività, ecc...) se non in rare circostanze. Nelle indagini statistiche, invece, la fase di raccolta dei dati è generalmente "governata" dal responsabile dell'indagine, a meno che non sia completamente delegata ad organismi di rilevazione esterni.

La fase a valle della rilevazione, ovvero la *registrazione* dei dati, costituisce invece il punto di aggancio tra il supporto cartaceo (questionario o modello di rilevazione) e quello informatico. In altri termini è la fase che rende elaborabili le informazioni. Nel caso

delle indagini statistiche la registrazione coincide con la raccolta dei dati se vengono utilizzate tecniche assistite dal computer di tipo CATI (Computer Assisted Telephone Interviewing) o CAPI (Computer Assisted Personal Interviewing). Negli altri casi è necessario predisporre un piano di registrazione con dei check che possano essere in grado di prevenire eventuali errori di posizione (registrazione dei dati in un campo sbagliato) e di valore (errata digitazione dell'informazione). Tra le indagini amministrative si distingue la situazione in cui i dati sono già disponibili su supporto magnetico da quella in cui si dispone soltanto delle schede cartacee. Il primo caso è certamente più favorevole sul piano della tempestività e dell'economicità, ma meno sul fronte dell'accuratezza del dato (in quanto le modalità di registrazione costituiscono una "scatola nera" per il produttore dell'informazione statistica). Il secondo caso è invece del tutto simile a quello delle indagini statistiche: visto che i dati devono ancora essere registrati si possono attivare tutti i controlli del caso.

La fase di *revisione* di un'indagine statistica si pone l'obiettivo di eliminare gli errori e le incongruenze presenti nel materiale di rilevazione, ovvero di:

- a) ricostruire il legame tra le unità nel caso in cui questo sia venuto meno (per esempio è questo il caso della relazione tra il foglio di famiglia e quello relativo ai singoli componenti)
- b) adottare piani di compatibilità e correzione per i singoli valori fuori range e per la relazione tra le variabili, eventualmente ricontattando le unità di rilevazione
- c) effettuare una più generale analisi di qualità delle diverse variabili rilevate

Nel caso di indagine amministrativa la revisione è – gioco forza – semplificata. Questo perché in primo luogo non dovrebbe verificarsi alcun tipo di incongruenza di cui al punto a). Poi – in particolare – per il fatto che eventuali valori relativi a singole variabili fuori range e/o casi di incompatibilità tra le variabili relative alla stessa unità di rilevazione non possono essere risolti ricontattando il rispondente. Questo implica – necessariamente – il ricorso alle procedure di revisione da parte di esperti (che – caso per caso – rimuovano le incoerenze) o, meglio, alle procedure automatiche (di tipo deterministico o da donatore). Per dettagli tecnici si vedano Masselli e Marchetti (1984) e Istat (1989f). Nessuna differenza tra indagine statistica e amministrativa si riscontra invece nelle procedure di analisi di qualità delle singole variabili.

Una volta registrati su supporto informatico e "corretti", i dati sono pronti per il successivo trattamento statistico. Ovvero per la costruzione di tavole, indicatori, ecc..

(*elaborazione*) e quindi per la *diffusione*. In questi aspetti non emerge alcuna differenza tra le due tipologie di indagine.

Una ultima difformità tra indagine statistica e amministrativa si riscontra invece nelle procedure di *validazione*, ovvero di valutazione se l'informazione statistica sia o meno sufficientemente valida per poter essere diffusa in funzione degli obiettivi di qualità programmati. In particolar modo nel caso di indagini amministrative sarà estremamente difficile la conduzione di studi di coerenza con fonti statistiche similari, a meno del caso di ripetizione nel tempo della stessa rilevazione. Risulterà poi – di fatto – preclusa l'attività di calcolo di indicatori diretti di qualità del processo in quanto (salvo casi particolari) non è ipotizzabile un ritorno presso un campione di unità di rilevazione (come è per esempio il caso delle indagini di copertura). Ci si dovrà accontentare di determinare alcuni indicatori indiretti di qualità del processo di produzione (Fortini, 1998).

4. Esempi di rilevazioni amministrative condotte nell'ambito del Programma Statistico Nazionale

Come è noto, il Programma Statistico Nazionale (PSN) raccoglie l'insieme delle rilevazioni statistiche di interesse pubblico che i soggetti del Sistema Statistico Nazionale si impegnano a realizzare nel corso di un triennio. Il PSN è predisposto dall'Istat e – dopo un lungo iter interno e esterno all'Istituto – approvato con DPCM previa deliberazione del CIPE. Il PSN è aggiornato ogni anno. Le linee guida ispiratrici del PSN (Istat, 2004) si orizzontano secondo tre grandi filoni: gli obiettivi di contesto (cioè quelli che tendono a migliorare l'ambiente di fruizione e produzione della statistica ufficiale), gli obiettivi di produzione, diffusione e utilizzazione dell'informazione statistica (ovvero il “core business”, l'attività centrale del PSN) e gli obiettivi strumentali, funzionali alla produzione e utilizzazione dell'informazione statistica.

Nell'ambito del PSN il ruolo rivestito dalle fonti amministrative è evidenziato e richiamato più volte tra gli obiettivi specifici. Sia in termini di valorizzazione statistica di fonti organizzate pubbliche e private: archivi registri, basi di dati (come secondo sub-obiettivo di produzione) che come normalizzazione della modulistica amministrativa e dell'integrazione a fini statistici dei sistemi informativi strutturali (secondo sub obiettivo strumentale). E' stabilito che ciò consentirà, tra l'altro, di ridurre la pressione statistica sui rispondenti, soprattutto le imprese.

Entrando nel merito della produzione statistica del PSN (per il triennio 2003-2005) rispetto alle 467 rilevazioni e alle 418 elaborazioni presentate, in 197 casi si fa ricorso all'acquisizione diretta da documenti amministrativi, in 379 si utilizzano fonti pubbliche organizzate (archivi e/o registri cartacei e/o informatizzati) e in 56 casi si procede all'acquisizione da fonti private organizzate).

Tra i diversi casi di indagini amministrative previste nel PSN limitiamo la nostra attenzione a quelle di tipo demografico, in particolar modo ai dati relative all'ammontare e alla struttura della popolazione. In questo settore, infatti, si è realizzata una coniugazione tra l'accresciuta domanda dell'utenza in termini di incremento dei contenuti informativi, di crescente dettaglio territoriale dell'informazione richiesta, di aumento della tempestività nel rilascio dei dati e l'effettiva disponibilità di fonti amministrative (le anagrafi) in genere ben aggiornate.

L'anagrafe della popolazione residente, tenuta presso ciascuno degli 8.101 Comuni d'Italia ai sensi del DPR 223/89, è infatti la raccolta sistematica dell'insieme di posizioni relative alle singole persone, alle famiglie e alle convivenze che hanno fissato nel comune la loro residenza, nonché delle posizioni relative alle persone senza fissa dimora che hanno stabilito nel comune il proprio domicilio. Lo Stato Civile deve effettuare all'ufficio di Anagrafe le comunicazioni concernenti le nascite, le morti e le celebrazioni di matrimonio. Dall'estrazione dei dati anagrafici è allora possibile, senza svolgere alcuna indagine diretta, conoscere le informazioni relative alla dinamica (nascite, decessi, movimento migratorio, saldo) e struttura della popolazione (numerosità, età, sesso, stato civile e cittadinanza).

Fino all'inizio degli anni '90 le uniche informazioni strutturali sulla popolazione potevano essere acquisite soltanto una volta ogni 10 anni, in occasione dei Censimenti generali (le indagini più complesse per eccellenza, se non altro per il numero di unità di rilevazione coinvolte). Tuttavia a partire dal rilascio dei dati definitivi strutturali del Censimento 1991 (nel 1995) l'informazione su ammontare e struttura della popolazione (resa possibile dall'introduzione della rilevazione POSAS, popolazione per sesso, stato civile e singolo anno di nascita) ha progressivamente portato alla costruzione di un sistema di stime di popolazione che è all'avanguardia a livello europeo sia in termini di contenuti informativi che di dettaglio territoriale e di tempestività. Questo ha comportato una progressiva transizione da un sistema fortemente *census oriented* (come è nel caso del Regno Unito e degli Stati Uniti) ad un sistema più di tipo *register-oriented* (dove si cerca di valorizzare al massimo le fonti amministrative già esistenti). Visto l'elevato livello di copertura raggiunto dalla Posas, e considerate ulteriori esigenze informative emerse, la

rilevazione è stata integrata a partire dal 2002 dalla STRASA (stranieri per sesso e anno di nascita al 1° gennaio di ogni anno). Un sistema di statistiche strutturali di questo genere, aggiornate annualmente, era addirittura impensabile fino alla fine degli anni '80.

E' ovvio che in questo nuovo sistema di indagine i dati censuari non sono assolutamente accantonati, oltre al contenuto informativo intrinseco servono infatti per integrare e rettificare quelli desunti dall'anagrafe. I risultati strutturali del Censimento 2001 comporteranno infatti un ridisegno complessivo della popolazione comunale dal 1° gennaio 1991 al 1° gennaio 2001, e il conseguente ricalcolo di tutti i principali parametri e indicatori demografici stimati annualmente per mezzo del dato anagrafico.

Il processo di progressivo sganciamento dai risultati censuari e di maggiore ancoraggio alla fonte anagrafica è apparso una condizione necessaria per rispondere al sempre più ampio ed esigente fabbisogno conoscitivo dell'utenza, sia dal punto di vista della produzione statistica che per quanto concerne il dettaglio territoriale di analisi. Si tratta di una scelta strategica in relazione alla quale il sistema di produzione dell'informazione demografica (ma anche quello della statistica in generale e della statistica ufficiale in particolar modo) non può arretrare, pur dovendo scontare un elevato prezzo in termini di avanzamenti metodologici e di tempestività.

Come anticipato, tuttavia, dal dato anagrafico non scaturiscono soltanto le informazioni strutturali, ma anche quelle relative ai singoli eventi che condizionano il saldo della popolazione (nascite, decessi e movimento migratorio) sia in termini *aggregati* (con la sola finalità del calcolo del bilancio anagrafico mensile e annuale di ciascun comune) che *analitici* (relativi cioè alle caratteristiche delle nascite, dei decessi e delle migrazioni).

Relativamente ai dati aggregati, ovvero al bilancio demografico della popolazione (e al conteggio dei residenti a fine anno), si realizza ancora una volta un'interessante integrazione tra indagine statistica e fonte amministrativa. La base di partenza per il calcolo è infatti rappresentata dalla popolazione legale (l'ammontare delle persone censite). A questa si aggiungono algebricamente tutti gli eventi rilevati mediante le registrazioni anagrafiche che incidono sul numero di residenti (nascite, decessi, migrazioni, integrazioni e rettifiche d'ufficio per tener conto di errori censuari). Un conteggio separato riguarda il bilancio demografico degli stranieri e la struttura per sesso e cittadinanza dei cittadini stranieri residenti in anagrafe al 1° gennaio di ogni anno.

Per ciò che concerne l'esame analitico (strutturale) dei singoli eventi che hanno incidenza sul bilancio demografico ci si appoggia su tre distinti sistemi di rilevazione, in

gran parte di tipo amministrativo: quello relativo alle nascite, ai decessi e al movimento migratorio.

Per le finalità del presente scritto il sistema informativo più interessante appare quello relativo alle nascite. Anche in questo caso si realizza infatti un ottimo esempio di integrazione tra la statistica per fonti e la statistica per indagine. A seguito delle riforme sulla certificazione amministrativa (Leggi 59 e 127 del 1999) la denuncia di nascita non deve necessariamente essere effettuata presso il Comune di nascita, ma anche presso il Presidio Ospedaliero dove è avvenuto l'evento o presso il Comune di Residenza. Questo fatto ha comportato, da parte dell'Istat, l'abolizione della tradizionale rilevazione sulle nascite di fonte stato civile. Dal 1999 per coprire il vuoto informativo è stato istituito un nuovo sistema di indagini:

- una rilevazione amministrativa basata sulle procedure di iscrizione anagrafica in grado di identificare soltanto le caratteristiche anagrafiche del parto (sesso, età della madre, comune di residenza)
- una rilevazione amministrativa basata sull'elaborazione dei Certificati di Assistenza al Parto (Cedap) che consentirà – a regime – di acquisire le informazioni di carattere bio-sanitario relative al parto, con valenza sia demografica che sanitaria
- una indagine campionaria su nascite e madri, che ha l'obiettivo di coprire il vuoto informativo lasciato dalle indagini amministrative ed anche di incrementare il prodotto statistico agli elementi motivazionali della fecondità (tale indagine, di tipo Cati, si basa su una lista di unità estratta dalla rilevazione amministrativa sulle procedure di iscrizione anagrafica)

La rilevazione sui decessi invece è effettuata a cura degli Uffici di Stato Civile attraverso la scheda di morte ² che comprende sia le notizie relative al decesso, con particolare riferimento alle cause di morte (parte A della scheda) che le informazioni di carattere demografico e sociale (parte B). Queste ultime informazioni vengono trasmesse dallo Stato Civile del Comune di decesso all'anagrafe di residenza attraverso il modello anagrafico AP/3, che ha conseguenze sul calcolo della popolazione residente.

La rilevazione sui trasferimenti di residenza, infine, viene effettuata attraverso i modelli APR/4, anch'essi con una valenza amministrativa (in quanto il modello deve essere compilato dal comune di iscrizione, integrato da quello di cancellazione e ritornare

² Modello denominato D.4 e D.5 per decessi, rispettivamente, di maschi e femmine oltre il primo anno di vita; e denominato D.4bis e D.4bis per decessi di maschi e femmine nel primo anno di vita.

poi presso il comune di iscrizione stesso per il completamento della pratica) e statistica (in quanto incidono sul conteggio della popolazione residente).

La fonte anagrafica, per la vastità informativa che propone, può essere utilizzata anche per il soddisfacimento di esigenze informative diverse rispetto a quelle richiamate fino a questo momento. Un paio di esempi di un certo interesse sono richiamati nei Paragrafi che seguono.

Tra i molteplici altri casi di integrazione tra dati amministrativi e dati statistici nell'ambito della statistica ufficiale, questa volta in campo economico, si segnalano la realizzazione e l'aggiornamento dell'archivio ASIA (Archivio Statistico delle Imprese Attive) in occasione del Censimento Intermedio dell'Industria e dei Servizi (Istat, 1998a) e l'integrazione fra dati statistici e dati amministrativi per la contabilità nazionale (Istat, 1998b).

5. Un esempio di utilizzo della fonte anagrafica per rilevare le principali misure fisiche relative ai cittadini residenti nel Comune di Lucca

Presso l'Amministrazione del Comune di Lucca nel corso del 2003 si è manifestata l'esigenza di conoscere le principali misure biometriche dei cittadini residenti in città, con particolare attenzione al peso, alla statura, al colore degli occhi e quello dei capelli. L'Ufficio di Statistica, supportato per gli aspetti tecnici dall'Ufficio Regionale Istat per la Toscana (in virtù di un accordo di programma preesistente), ha cercato di identificare il sistema più rapido ed economico per soddisfare questa esigenza.

In primo luogo ha svolto una attenta opera di ricognizione delle fonti statistiche eventualmente già preesistenti su questi temi. Dall'esame preliminare è emersa una sostanziale carenza di dati, fatta eccezione per la statura rilevata ai maschi iscritti alla visita di leva (in età di circa 18 anni), per il peso e la statura registrati presso le AUSL ai diversi residenti in casi particolari (alla nascita, in occasione di visite sportive, visite mediche, ricoveri ospedalieri e simili). La frammentarietà dei dati a disposizione ha reso necessaria, per soddisfare le esigenze manifestate dall'Amministrazione, la progettazione di una vera e propria indagine, ovviamente di tipo campionario. Gli ulteriori requisiti di tempestività, riduzione del fastidio statistico dei cittadini e – non ultimo – di minimizzazione dei costi hanno di fatto impedito la realizzazione di un'indagine diretta, aprendo il campo ad una soluzione di tipo amministrativo. In questo l'unica fonte che potesse in qualche modo

contenere la maggior parte degli item desiderati è quella anagrafica. In effetti al momento del rinnovo della carta di identità l'Ufficiale di Anagrafe compila per ciascun individuo una scheda contenente, tra gli altri elementi, i dati relativi alla statura, al colore degli occhi e dei capelli.

Dato che il rinnovo della Carta di Identità interessa tutti i cittadini residenti nel Comune, la presentazione di ciascuno di questi presso lo sportello anagrafico può essere considerata di tipo casuale. Per questo siamo stati legittimati ad assimilare tale situazione a quella di un cittadino che – per caso – sia stato estratto dalle liste anagrafiche per rispondere ad un questionario statistico su tali variabili.

Abbiamo così deciso di considerare come campione statisticamente significativo per tutta la popolazione residente le persone che hanno ricevuto per la prima volta (o che hanno rinnovato) la carta di identità nel mese di Giugno 2003 (il dato più recente a disposizione al momento della realizzazione del lavoro), pari a 695 unità (337 maschi e 358 femmine), una numerosità compatibile con gli obiettivi di precisione richiesti dalle stime³ (Fabbris, 1989; Istat, 1989d e 1989e). Dal punto di vista operativo, sono stati registrati in pochi giorni (e quindi resi elaborabili) i dati relativi ai cartoncini tramite un semplice programma Access.

Prima di estendere i risultati campionari all'universo si è reso necessario valutare entro quali limiti fosse valida l'ipotesi di rappresentatività del campione per tutta la popolazione. Verificare la casualità nelle operazioni di rinnovo della carta di identità significa, nel nostro caso, studiare se esiste compatibilità nella struttura dei flussi di rinnovo per il mese preso come riferimento (Giugno 2003) rispetto alla struttura per sesso e per età della popolazione a quella data. A tal riguardo la Tavola 1 riporta la distribuzione percentuale per classi per età e sesso del campione e della popolazione residente con oltre 15 anni.

Tavola.1 Distribuzione percentuale per classi di età e per sesso

Classi per età	% nel campione		% nella popolazione	
	M	F	M	F
15 – 39	41,84	39,94	38,65	33,57
40 – 64	39,47	38,27	40,03	37,56
65 e oltre	18,69	21,79	21,32	28,87
Totale	100,00	100,00	100,00	100,00

³ ± 1 cm per quanto concerne la statura; ± 0,05 punti per quanto riguarda l'incidenza delle diverse modalità di colore degli occhi e dei capelli

Considerando che l'età è nota per 337 maschi e 358 femmine del campione, questo risultato è compatibile con l'ipotesi di estrazione del campione dalla popolazione anagrafica ad un livello di significatività dell' 1% sulla base di un test φ^2 con 4 gradi di libertà. Il risultato ci ha confortato circa l'esistenza di una certa *casualità* nelle operazioni di rinnovo della carta di identità per le persone con 15 anni e più. A questo punto, utilizzando le classiche tecniche dell'inferenza statistica siamo stati in grado di procedere con il calcolo dei parametri rappresentativi per tutte le variabili di interesse (statura media, incidenza delle varie modalità del colore degli occhi e del colore dei capelli).

Seguendo la strada amministrativa si è dovuto abbandonare l'obiettivo conoscitivo relativo al peso dei cittadini, variabile non contemplata nel questionario di rilevazione. Per recuperare anche quest'ultima variabile avremmo dovuto chiedere a tutti i cittadini che hanno rinnovato la carta di identità nel periodo identificato per la definizione del campione anche un'informazione in più – esogena rispetto a quelle del procedimento amministrativo: un'operazione che è apparsa eccessivamente onerosa sia in termini di fastidio statistico verso il rispondente che come impianto di tipo organizzativo necessario rispetto all'effettivo miglioramento del patrimonio conoscitivo che ne sarebbe derivato.

Anche la qualità della rilevazione ha risentito del fatto che non è stato possibile porre alcun tipo di controllo sull'input dei dati. In particolare in merito all'altezza è stato sufficiente escludere i cittadini al di sotto di 20 anni (che potenzialmente possono variare la statura per effetto dello sviluppo). Per ciò che concerne il colore degli occhi e dei capelli, invece, la variabile rilevata è quella che emerge alla vista dell'ufficiale di anagrafe, senza entrare nel merito che una persona possa essersi messa le lenti a contatto di un colore diverso da quello naturale o che possa essersi tinta i capelli. Del resto per conoscere anche questi ulteriori aspetti sarebbe stato necessario imbastire una indagine motivazionale sui singoli cittadini, elemento pure in questo caso escluso a priori in virtù dell'eccessiva onerosità di un'iniziativa del genere rispetto all'effettivo fabbisogno conoscitivo. Per dettagli sulla metodologia utilizzata e sui risultati raggiunti si vedano Valentini e Particelli (2004).

6. Proposte operative per l'aggiornamento in via amministrativa di un sistema informativo statistico comunale realizzato utilizzando le risultanze elementari dei Censimenti Generali degli anni 2000

Durante le operazioni di rilevazione censuaria dei primi anni 2000 è emersa, da parte di un elevato numero di Enti Sistan presenti sul territorio regionale della Toscana, l'esigenza di utilizzare i dati raccolti – una volta validati e restituiti da parte dell'Istat – per la costituzione di un Sistema Informativo Territoriale. Questo in virtù del fatto che, per la prima volta, i dati dovevano essere diffusi (sulla base del progetto Census 2000) utilizzando le stesse griglie territoriali di tipo sub-comunale. Le esigenze espresse dal territorio sono state riprese dall'Ufficio Regionale Istat per la Toscana⁴ che ha sviluppato una riflessione consistente in uno studio progettuale per la realizzazione tramite i dati censuari e il successivo aggiornamento mediante fonti amministrative di un Sistema Informativo Territoriale (SIT) a carattere comunale o sovra-comunale.

Al di là degli aspetti tecnici connessi con l'effettiva realizzazione del SIT, si intende in questa sede proporre tale esempio come un ulteriore caso concreto di integrazione tra fonti statistiche e fonti amministrative. Dal punto di vista logico il SIT si struttura secondo tre fasi:

1. identificazione delle aree sub-comunali di specifico interesse per l'Amministrazione (ciascuna area dovrà, possibilmente, essere separata dalle altre da limiti fisici evidenti)
2. associazione a ciascuna area dei risultati delle rilevazioni censuarie derivanti dall'elaborazione dei dati per sezione di censimento
3. attribuzione alle varie aree dei risultati di tutte le indagini amministrative di un certo interesse

L'aggancio dei risultati di ciascuna rilevazione sul territorio avviene tramite la sezione di censimento nel caso di dati censuari, mediante l'indirizzo stradale (via e numero civico) per quanto concerne le rilevazioni amministrative. La sovrapposizione sulle stesse griglie di riferimento territoriale delle due diverse tipologie di fonti è garantita da uno specifico

⁴ Si veda VALENTINI A. (2004), *Proposte operative per la costituzione di un Sistema Informativo Territoriale a livello comunale*, documento interno dell'Istat – Ufficio Regionale per la Toscana redatto con la supervisione del dirigente dell'Ufficio Dr. Franco Corea, ed anche grazie contatti intrattenuti con il Dipartimento di Statistica dell'Università di Pisa (ed in particolare il Prof. A. Bonaguidi).

modello compilato in occasione del Censimento della Popolazione (denominato CP.5), che associa a ciascuna Sezione di Censimento i rispettivi archi di strada.

In relazione al tipo di indagine amministrativa da utilizzare per l'aggiornamento periodico del sistema - il punto più interessante in questa sede - si tratta di valorizzare opportunamente tutti i giacimenti informativi in possesso delle Amministrazioni comunali secondo una duplice logica. Da una parte si intendono integrare le carenze dei dati censuari nei settori da questi non trattati (per esempio relativamente al reddito prodotto o consumato, al movimento turistico, al traffico merci, alle strutture logistiche), dall'altra si tratta di "rinfrescare" periodicamente i dati censuari stessi (che, come si ricorderà, sono riferiti al 21 Ottobre 2001 per il Censimento della Popolazione, al 22 Ottobre dello stesso anno per il Censimento dell'Industria e al 20 Ottobre 2000 per quello dell'Agricoltura). Per il raggiungimento di questi obiettivi è necessario che ogni Amministrazione censisca gli archivi in suo possesso, e che selezioni quelli rilevanti ai fini dell'aggiornamento del SIT tra la grande quantità a disposizione.

Censire gli archivi significa non tanto conoscere di quante basi dati si dispone ma capire il contenuto informativo di ogni archivio, il livello di informatizzazione, il grado di accesso all'archivio stesso, la periodicità di aggiornamento, ecc... Il modello di rilevazione che abbiamo consigliato ad alcune Amministrazioni per censire gli archivi (o comunque le basi dati esistenti) risponde all'obiettivo di documentazione omogenea dei patrimoni informativi settoriali. Si è trattato, cioè, di trovare un compromesso tra l'esigenza di documentare in modo standard i vari archivi e le caratteristiche specifiche che questi assumono in alcuni settori. In concreto per ogni archivio si è identificato il nome, la struttura di appartenenza, un referente, le modalità alla base della tenuta (per il soddisfacimento di obblighi di legge, di esigenze conoscitive dell'Ente o per altre motivazioni), le eventuali procedure di accesso da parte del pubblico, la tempistica di acquisizione delle informazioni, la tempistica di registrazione, le modalità di registrazione delle informazioni e le eventuali modalità di diffusione dei dati disaggregati o aggregati.

Un interessantissimo lavoro relativo alla realizzazione del repertorio del basamento informativo esistente è stato svolto da parte della Regione Toscana e ad esso si rimanda per eventuali approfondimenti su questo tema (Regione Toscana, 1999).

Una volta terminato il censimento degli archivi, per l'aggiornamento del SIT dovranno essere scelte le basi di dati:

- che coprono specifiche esigenze conoscitive dell'Amministrazione, lasciate scoperte dalle rilevazioni censuarie. Per esempio, per un comune marino con un

- porto, assumeranno primario interesse i dati sul movimento portuale. Per un comune con un aeroporto, i traffici di passeggeri e merci per via aerea, e così via.
- che siano assimilabili alle principali variabili rilevate in occasione dei Censimenti generali, consentendo di effettuare un aggiornamento periodico di questi aspetti informativi (anagrafe, notizie sull'attività edilizia, registri delle imprese attive, ecc...)
 - che siano facilmente reperibili, a cadenze di tempo prefissate o in maniera continuativa, con dati possibilmente già disponibili a cadenze prefissate
 - che siano possibilmente già disponibili su supporto informatico
 - che si “agganciano” senza problemi sul territorio, ovvero che presentano un indirizzo stradale (area di circolazione e numero civico) affidabile.

Realizzati questi passi, l'Ente locale con una spesa minima sarà in grado di disporre di uno strumento statistico molto utile grazie all'utilizzo con un'ottica nuova dei giacimenti informativi già in suo possesso.

7. Bibliografia essenziale

- BARCAROLI e al. (1987), *Manuale per la progettazione concettuale dei dati statistici*, Istat
- COLOMBO B. (1983), *La qualità dei dati statistici* in Atti del Convegno 1983 della S.I.S, Trieste
- FABBRIS L. (1989), *L'indagine campionaria: metodi, disegni e tecniche di campionamento*, La Nuova Italia Scientifica, Roma
- FORTINI M. (1998), *Gli indicatori standard di qualità nel sistema informativo di documentazione delle indagini*, Contributi Istat n. 7/1998
- FORTINI M. (2000), *Linee guida metodologiche per le rilevazioni statistiche*, Istat, Roma
- ISTAT (1989a), *Manuale di tecniche di indagine: 1 – Pianificazione della produzione dei dati*, Note e relazioni n. 1, Istat
- ISTAT (1989b), *Manuale di tecniche di indagine: 2 –Il questionario: progettazione, redazione, verifica*, Note e relazioni n. 1, Istat
- ISTAT (1989c), *Manuale di tecniche di indagine: 3 – Tecniche di somministrazione del questionario*, Note e relazioni n. 1, Istat

ISTAT (1989d), *Manuale di tecniche di indagine: 4 – Tecniche di campionamento, teoria e pratica*, Note e relazioni n. 1, Istat

ISTAT (1989e), *Manuale di tecniche di indagine: 5 – Tecniche di stima della varianza campionaria*, Note e relazioni n. 1, Istat

ISTAT (1989f), *Manuale di tecniche di indagine: 6 – Il sistema di controllo della qualità dei dati*, Note e relazioni n. 1, Istat

ISTAT (1998a), *L'impianto normativo, metodologico e organizzativo*, Censimento Intermedio dell'Industria e dei Servizi (31/12/1998), Istat

ISTAT (1998b), *Vecchie e nuove fonti: l'integrazione fra dati statistici e dati amministrativi per la contabilità nazionale*, Metodi e Norme – n. 2, Istat

ISTAT (2000), *Codice della Statistica Ufficiale - I, L'organizzazione, Seconda edizione*, Sistema Statistico Nazionale, Istat

ISTAT (2002), *Codice della Statistica Ufficiale- I, L'organizzazione, Appendice di aggiornamento alla Seconda Edizione*, Sistema Statistico Nazionale, Istat

ISTAT (2003), *Metodi statistici per il record linkage*, Metodi e Norme – n. 16, Istat

ISTAT (2004), *Programma statistico nazionale, triennio 2004/2006 [Parte prima – obiettivi e progetti]*, Sistema Statistico Nazionale, Istat

MASSELLI M., MARCHETTI E. (1984), *I piani di compatibilità ed il controllo dell'attendibilità del dato*, Atti della XXXII Riunione Scientifica della S.I.S., Sorrento

REGIONE TOSCANA (1999), *Il repertorio del basamento informativo*, Informazioni Statistiche n. 12 (Anno X), Novembre 1999

VALENTINI A., PARTICELLI S. (2004), *Aspetti statistici inediti relativi al Comune di Lucca*, Monografia edita dal Comune di Lucca

ZARKOVICH (1967), *Sampling methods and censuses – Volume II – Quality of statistica data*, F.A.O.