

# NOTA SULLA TUTELA STATISTICA DELLA RISERVATEZZA PER I *FILE STANDARD* TRATTI DAL 14° CENSIMENTO DELLA POPOLAZIONE E DELLE ABITAZIONI - 2001<sup>1</sup>

## 1. Introduzione

La presente Nota descrive le operazioni effettuate sui dati dei *file standard* (collezioni campionarie di dati individuali) tratti dal 14° Censimento della popolazione e delle abitazioni – 2001 al fine di tutelare la riservatezza delle unità statistiche come prescritto dalla legge<sup>2</sup>. Ricordiamo che si tratta di due diverse tipologie di file: un file individuale, in cui ciascun record corrisponde ad un singolo individuo indipendentemente dagli altri; un file familiare, in cui è possibile associare ciascun individuo alla famiglia di appartenenza e, a questa, informazioni sull'abitazione occupata.

Seguendo l'evoluzione delle tecniche e procedure adottate dall'Istituto per la costruzione dei *file standard*, anche nel caso del Censimento della popolazione si è adottata una misura del rischio di identificazione a livello individuale<sup>3</sup> (per *record*) che ha sostituito, come nella maggior parte dei *file standard* rilasciati dall'Istituto, la precedente misura di rischio globale<sup>4</sup> (per l'intero *file*). Un tale approccio, ha consentito di applicare misure statistiche di protezione dei dati (che necessariamente comportano una riduzione del contenuto informativo del file) in maniera mirata solo a quei record effettivamente considerati a rischio piuttosto che intervenire sull'intero file. Ciò ha permesso di raggiungere, rispetto al passato, un maggior dettaglio di informazione rilasciata soprattutto per quanto riguarda (i) la dimensione dei file e (ii) gli aspetti gerarchici legati alla struttura familiare delle unità statistiche che nel file del precedente Censimento erano assenti.

La procedura di misura del rischio di identificazione è basata sulla combinazione di modalità che ogni record assume nell'insieme delle cosiddette variabili identificative indirette (o variabili chiave). Ricordiamo che definiamo come variabili identificative indirette quelle variabili che da sole, o in combinazione con altre, possono portare all'identificazione di uno o più record del file rilasciato. In particolare si è proceduto considerando "a rischio" quei record che presentavano una combinazione delle variabili chiave considerate particolarmente rare nella popolazione (ossia con frequenza inferiore a una soglia prefissata). Per il File Famiglie, invece, si è tenuto conto anche dell'appartenenza di ciascun individuo alla famiglia contemplando la possibilità di essere individuati 'indirettamente' tramite l'identificazione di un altro componente della famiglia<sup>5</sup>.

I metodi di protezione applicati sono la "ricodifica globale" (*global recoding*) e la "soppressione locale" (*local suppression*). Con la soppressione locale, i valori di una variabile per una particolare unità statistica vengono sostituiti con un valore mancante (il codice 9, eventualmente ripetuto per la lunghezza del campo). Questo tipo di intervento viene applicato ai soli record per cui la misura del rischio è superiore alla soglia fissata.

## 2. Protezione dei dati

La ricodifica delle variabili chiave coinvolte nella valutazione del rischio di violazione della riservatezza e la strategia di protezione che realizzi il miglior compromesso tra le esigenze degli utenti e il diritto alla tutela dei rispondenti è stata studiata tenendo conto dei domini di stima pianificati nel disegno campionario.

---

<sup>1</sup> Nota metodologica a cura di Giovanni Seri e Maurizio Lucarelli. Elaborazioni a cura di Maurizio Lucarelli.

<sup>2</sup> Decreto legislativo 30 giugno 2003, n. 196 - Codice in materia di protezione dei dati personali, cui è annesso il Codice di deontologia e di buona condotta per i trattamenti di dati personali a scopi statistici e di ricerca scientifica effettuati in ambito Sistan (in particolare art. 4 e art. 7); Decreto legislativo 6 settembre 1989, n. 322 (in particolare art. 9 e art. 10).

<sup>3</sup> Franconi, L. e Poletti, S. (2004) Individual risk estimation in  $\mu$ -Argus: a review. In J.Domingo-Ferrer and V.Torra eds. *Privacy in Statistical Databases*. Berlin: Springer, pp.262-272.

<sup>4</sup> Crescenzi, F. (1993). On estimating population uniques. *Methodological proposals and applications on Italian Census data*. Proceedings of International Seminar on Statistical Confidentiality, Dublin, Ireland, September 1992, pp. 247-260.

<sup>5</sup> L. Coppola e G. Seri (2005), "Confidentiality aspects of household panel survey: the case study of Italian sample from EU-SILC", *Monographs of official statistics – Proceedings della Work session on statistical data confidentiality*, Ginevra 9-11 novembre 2005, Eurostat.

## 2.1 File Individui

Le classificazioni derivanti dalle operazioni di ricodifica effettuate sul File Individui sono riportate nella Tabella 1. Le variabili che non hanno subito modifiche non sono riportate. Inoltre, per le combinazioni delle variabili: Luogo di residenza, Sesso, Età e Stato civile, il numero di record a rischio è risultato pari a 4764. Per questi record è stato necessario sopprimere l'informazione relativa allo Stato civile (considerata la variabile meno informativa e sulla quale sono state concentrate le soppressioni).

**Tabella 1. Codifiche applicate alle variabili del File Individui**

Descrizione Variabile	Classificazione applicata
Codice della provincia di rilevazione (istat)	Oltre alle provincie sono identificabili i comuni capoluogo di provincia con più di 150000 abitanti dal resto della provincia (se anche questo aggregato territoriale conta più di 150000 abitanti)
Tipologia familiare	Famiglia senza nuclei unipersonale - Non in coabitazione Famiglia senza nuclei unipersonale - In coabitazione Altra famiglia senza nuclei unipersonale Famiglia con un solo nucleo senza altre persone residenti: Coppia senza figli Famiglia con un solo nucleo senza altre persone residenti: Coppia con figli Famiglia con un solo nucleo senza altre persone residenti: Madre con figli Famiglia con un solo nucleo senza altre persone residenti: Padre con figli Famiglia con un solo nucleo con altre persone residenti: Coppia senza figli Famiglia con un solo nucleo con altre persone residenti: Coppia con figli Famiglia con un solo nucleo con altre persone residenti: Madre con figli Famiglia con un solo nucleo con altre persone residenti: Padre con figli Famiglia con due o più nuclei
Anno di nascita	Sostituito con Età in 20 classi: 0-2 3-5 6-10 11-13 14-17 18-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85+
Luogo di nascita	Nello stesso Comune di residenza In altro Comune della stessa Provincia In altra Provincia della stessa Regione In altre Regioni Estero
Codice dello stato estero di nascita	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania

Codice dello stato estero di cittadinanza	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania Apolide
Stato civile	Celibe/nubile Coniugato/a Separato/a di fatto; Separato/a legalmente; Divorziato/a Vedovo/a
Codice della provincia di dimora abituale al 21 ottobre 2000	Nello stesso Comune di residenza In altro Comune della stessa Provincia In altra Provincia della stessa Regione In altre Regioni Estero
Codice dello stato estero di dimora abituale al 21 ottobre 2000	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania
Titolo di studio più elevato	Analfabeti Alfabeti privi di titolo di studio Licenza elementare Licenza media Diploma scolastico di qualifica (corso scolastico di 2-3 anni) Diploma di maturità (corso scolastico di 4-5 anni) Diploma terziario di tipo non universitario Diploma universitario Diploma di laurea
Luogo abituale di studio o di lavoro	Nello stesso Comune di residenza In altro Comune della stessa Provincia In altra Provincia della stessa Regione In altre Regioni Estero
Codice dello stato estero di studio o di lavoro	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania

## 2.2 File Famiglie

Le classificazioni derivanti dalle operazioni di ricodifica effettuate sul File Famiglie sono riportate nella Tabella 2. Le variabili che non hanno subito modifiche non sono riportate. Inoltre, per le combinazioni delle variabili: Sesso, Età, Stato civile, Relazione di parentela, Numero componenti e Luogo di residenza, il numero di record a rischio è risultato pari a 25179. Per questi record è stato necessario procedere sopprimendo una o più informazioni come riportato nelle tabelle 3 e 4. Il numero di record protetti (che subiscono almeno una soppressione locale) è pari a 25401 (4,16% sul totale).

Al fine di compensare la mancanza di informazioni sull'età degli individui dovuta alle soppressioni locali, è stata introdotta una variabile Età in 4 classi (0-17, 18-34, 35-64, 64 e oltre) non sottoposta ad alcuna soppressione.

**Tabella 2. Codifiche applicate alle variabili del File famiglie**

Descrizione Variabile	Classificazione applicata	
Regione di residenza	19 regioni: Piemonte e Val d'Aosta accorpate	
Tipologia familiare	Famiglia senza nuclei unipersonale - Non in coabitazione Famiglia senza nuclei unipersonale - In coabitazione Altra famiglia senza nuclei unipersonale Famiglia con un solo nucleo senza altre persone residenti: Coppia senza figli Famiglia con un solo nucleo senza altre persone residenti: Coppia con figli Famiglia con un solo nucleo senza altre persone residenti: Madre con figli Famiglia con un solo nucleo senza altre persone residenti: Padre con figli Famiglia con un solo nucleo con altre persone residenti: Coppia senza figli Famiglia con un solo nucleo con altre persone residenti: Coppia con figli Famiglia con un solo nucleo con altre persone residenti: Madre con figli Famiglia con un solo nucleo con altre persone residenti: Padre con figli Famiglia con due o più nuclei	
Anno di nascita	Sostituito con Età in 20 classi: 0-2 3-5 6-10 11-13 14-17 18-19 20-24 25-29 30-34 35-39 40-44 45-49 50-54 55-59 60-64 65-69 70-74 75-79 80-84 85+	e con Età in 4 classi: 0-17 18-34 35-64 65+
Luogo di nascita	Nello stesso Comune di residenza In altro Comune della stessa Provincia In altra Provincia della stessa Regione In altre Regioni Estero	
Codice dello stato estero di nascita	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania	

Codice dello stato estero di cittadinanza	Ricodificato come segue: Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania Apolide
Codice della provincia di dimora abituale al 21 ottobre 2000	Nello stesso Comune di residenza In altro Comune della stessa Provincia In altra Provincia della stessa Regione In altre Regioni Esteri
Codice dello stato estero di dimora abituale al 21 ottobre 2000	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania
Titolo di studio più elevato	Ricodificato come segue: Analfabeti Alfabeti privi di titolo di studio Licenza elementare Licenza media Diploma scolastico di qualifica (corso scolastico di 2-3 anni) Diploma di maturità (corso scolastico di 4-5 anni) Diploma terziario di tipo non universitario Diploma universitario Diploma di laurea
Luogo abituale di studio o di lavoro	Ricodificato come segue: Nello stesso Comune di residenza In altro Comune della stessa Provincia In altra Provincia della stessa Regione In altre Regioni Esteri
Codice della provincia di studio o di lavoro	19 regioni: Piemonte e Val d'Aosta accorpate
Codice dello stato estero di studio o di lavoro	Unione Europea 15 Paesi di nuova adesione all'Unione Europea Europa centro-orientale Altri paesi europei Africa settentrionale Africa occidentale Africa orientale Africa centro-meridionale Asia occidentale Asia centro-meridionale Asia orientale America settentrionale America centro-meridionale Oceania

**Tabella 3. Soppressioni locali File Famiglie per variabile e dimensione familiare**

		Ampiezza familiare (numero componenti)											Soppress. per var chiave (% su 610198 individui)
		1	2	3	4	5	6	7	8	9	10	11	
		Numero di individui nel campione per ampiezza familiare											
		50881	107904	132813	163084	89030	46332	13237	4152	1782	730	253	
Variabili chiave	Regione (19 regioni)	0	0	0	0	0	0	0	0	0	0	0	0
	Sesso	0	0	0	0	0	0	0	0	1782	730	253	2765 (0.45%)
	Stato Civile (4 modalità)	132	1118	1442	1652	1920	2334	1575	806	1782	730	253	13744 (2.25%)
	Età (20 classi)	13	2	0	3	35	463	2146	2238	1782	730	253	7665 (1.26%)
	Relazione di parentela (6 modalità)	0	595	940	1165	1664	3171	2694	1237	1782	730	253	14231 (2.33%)
Soppressioni per ampiezza familiare		145	1715	2382	2820	3619	5968	6415	4281	7128	2920	1012	38405

**Tabella 4. Distribuzione per ampiezza familiare delle unità statistiche protette**

Ampiezza familiare	N. di famiglie protette	% sul numero di famiglie	N. di individui protetti	% sul numero di individui/record
1	145	0.28	145	0.28
2	1574	2.92	1691	1.57
3	1558	3.52	2282	1.72
4	1446	3.55	2527	1.55
5	1622	9.11	3085	3.47
6	1912	24.76	4801	10.36
7	1110	58.70	5049	38.14
8	459	88.44	3056	73.60
9	198	100	1782	100
10	73	100	730	100
11	23	100	253	100
Totale	10120	4.64	25401	4.16

## 2.3 File Alloggi

Per quanto riguarda gli alloggi le misure di protezione si sono limitate ad alcune operazioni di top coding relativamente alle variabili:

- numero totale stanze (classificata in 1, 2, 3, + di 3);
- numero stanze ad uso professionale (classificata in 1, 2, 3, + di 3);
- numero bagni (classificata in 1, 2, 3, + di 3);
- numero impianti doccia (classificata in 1, 2, 3, + di 3);

e di arrotondamento alle decine e *top coding* per la variabile:

- superficie dell'abitazione in mq (classificata in 10, 20, ..., 140, 150 o +).