



**Rilevazione continua sulle Forze di Lavoro.**  
dal primo trimestre 2008

## **Descrizione del file**

## INDICE

1. Introduzione .....	3
2. Descrizione delle variabili .....	3
3. Metodologia statistica per la tutela della riservatezza .....	3
4. Riferimenti bibliografici.....	6
5. Contatti .....	7

# 1. Introduzione

La necessità di tutela della riservatezza discende da vincoli di legge e da obblighi assunti verso i rispondenti. Seguendo l'Handbook on Statistical Disclosure Control (AA.VV., 2009), l'intrusione in informazioni non pubbliche si concretizza quando tramite i dati rilasciati sia possibile l'identificazione del rispondente. Limitando l'attenzione al rilascio di microdati, l'intrusione avviene quando il singolo record è correttamente assegnato ad una unità nella popolazione. Le unità statistiche oggetto dell'analisi sono gli individui e, condizionatamente alla stratificazione di interesse, sono considerate a rischio quelle che, uniche nel campione, siano tali anche nella popolazione. Le variabili specificate nel questionario di indagine hanno suggerito due possibili ambiti di riferimento; il primo attiene al contributo di singole variabili ad identificazioni spontanee o accidentali; nel secondo, si assume venga perseguito intenzionalmente l'abbinamento tra i dati di indagine e una lista di microdati disponibile secondo un prefissato livello di dettaglio.

.

## 2. Descrizione delle variabili

In riferimento all'insieme di informazioni ottenuto secondo il questionario di indagine, la riduzione del rischio di violazione della riservatezza è stata perseguita combinando soppressioni e modifiche nel livello di aggregazione delle variabili. Tali interventi, nonché quelli volti ad assicurare la confrontabilità rispetto ad alcune mutabili rese disponibili da Eurostat (AA.VV., 2008), sono esposti nel paragrafo 3.2. Per significati e articolazioni delle restanti variabili, rimaste immutate, si rinvia al questionario di indagine e al relativo tracciato record.

## 3. Metodologia statistica per la tutela della riservatezza

### 3.1 Apprezzamento del Rischio

In ordine al primo ambito di riferimento, riferito alle cosiddette intrusioni accidentali, l'identificazione origina "spontaneamente" dall'osservazione di un ristretto numero di variabili, alcune delle quali aventi un livello di dettaglio sufficiente ad isolare poche unità statistiche in base alla sola distribuzione marginale di frequenze. Per ottenere una plausibile approssimazione degli ordini di grandezza coinvolti, in riferimento alle mutabili maggiormente suscettibili di criticità, è stato conteggiato il numero di modalità che in un trimestre siano state rappresentate da una sola unità statistica, espungendo quelle

presenti in tre trimestri consecutivi. Poiché “l’indagine sulle forze di lavoro segue uno schema di rotazione trimestrale in cui le famiglie vengono intervistate per due trimestri consecutivi, escluse per due trimestri e successivamente intervistate per altri due trimestri” (AA.VV., 2006), uno strato presente in tre trimestri consecutivi è verosimilmente ascrivibile alla presenza, nella popolazione, di almeno due unità statistiche che ne condividano le modalità. Con riferimento al secondo ambito, le unità statistiche a rischio di violazione della riservatezza sono state identificate ipotizzando uno scenario di abbinamento tra dati di indagine e una lista di microdati disponibile ad un terzo).

Un possibile approccio alla valutazione del rischio è descritto in Skinner e Shlomo (2008). A causa degli oneri computazionali derivanti dall’elevata disaggregazione delle variabili è stata implementata una strategia di classificazione alternativa, fondata su esercizi di simulazione per processi aleatori.

### **3.2. Modifiche e Integrazioni delle variabili**

#### **3.2.1. Soppressione delle variabili**

Diverse variabili, rilevate secondo il questionario di indagine, sono state rese non disponibili nel file. Tra di esse è significativo ricordare:

**Tabella 2: variabili rese non disponibili**

NOME VARIABILE	TIPO	PROVENIENZA (*)	DESCRIZIONE VARIABILE	SDC STATUS
CODFAM	CHAR	r	codice famiglia	soppressa
CODQUA	CHAR	r	codice quartina	soppressa

(\*) Le variabili rilevata nel questionario e ricodificate nella colonna sono indicate con r.

#### **Ricodifica globale**

La ricodifica globale è tra i più semplici metodi di riduzione del rischio di violazione della riservatezza nei dati individuali. Per caratteri raggruppati in classi, essa è definita dall’accorpamento di alcune modalità con ovvi cambiamenti per significati e frequenze. Le scelte effettuate a questo proposito sono state evidenziate nella tabella 2.

**Tabella 2: variabili recanti modificazioni delle modalità**

NOME VARIABILE	TIPO	PROVENIENZA (*)	DESCRIZIONE VARIABILE	SDC STATUS	VARIAZIONE
GRACOM	CHAR	r	codice comune	variato	In ciascuna provincia, codifica binaria in grande comune e comuni residuali
MFRFAM	CHAR	r	codice famiglia	variato	ricodifica con metodo casuale
MFRIND	CHAR	r	Identificativo dell'individuo all'interno della famiglia	variato	ricodifica con metodo casuale
ETAM	NUM	r	età in anni compiuti	variato	anni compiuti, fatti salvi i raggruppamenti: fino a 2 anni, da 3 a 5 anni, da 6 a 10 anni, da 11 a 14 anni, 75 anni e oltre.
STACIM	CHAR	r	stato civile	variato	aggregazione delle modalità 3, 4 e 5 nella 3 (separati e divorziati) e ricodifica della 5 nella 4 (vedovi)
TFM	CHAR	r	tipologia familiare	variato	aggregazioni effettuate per coerenza con quelle relative allo stato civile: le modalità 12 e 13 sono accomunate nella 12, le 17 e 18 nella 18, 26 e 27 nella 26, 31 e 32 nella 31
RETRIC	NUM	r	retribuzione in unità di euro del mese precedente escluse 13 <sup>me</sup> , 14 <sup>me</sup> ed emolumenti non percepiti tutti i mesi	variato	raggruppamento in classi delle intensità fino a € 250 e oltre € 3000 e per le rimanenti arrotondamento ai 10 euro
PROFM, PROFM, PROFEM	CHAR	r	posizione professionale, posizione professionale secondaria, posizione professionale precedente, espresse a 4 digit	variate	limitazione al terzo digit per le posizioni di legislatori e dirigenti

### Ricodifica a fini di integrazione

Ai fini di una più immediata comparabilità con i dati omologhi rilasciati da Eurostat sono state incluse le variabili di seguito elencate:

**Tabella 3: variabili inserite per comparabilità con i dati Eurostat**

NOME VARIABILE	TIPO	PROVENIENZA (*)	DESCRIZIONE VARIABILE
INCDEC	NUM	r	decile di appartenenza per la retribuzione mensile percepita
ISCO3D	CHAR	r	codifica ISCO delle professioni, omologa della CP2001 a tre digit
HATLEV	CHAR	r	codifica, omologa della ISCED, relativa a titolo di studio più elevato e corsi di formazione della sezione H: 00 Nessun titolo o titolo inferiore a ISCED 1 11 ISCED 1 21 ISCED 2 22 ISCED 3c (meno di due anni) 31 ISCED 3c (almeno due anni) 32 ISCED 3 a, b 30 ISCED 3 (senza distinzione a, b, c con almeno 2 anni) 41 ISCED 4 a, b 42 ISCED 4c 43 ISCED 4 (senza distinzione a, b o c) 51 ISCED 5b 52 ISCED 5a 60 ISCED 6 99 Non applicabile (persone con meno di 15 anni)

## 4. Riferimenti bibliografici

- AA.VV. *EU Labour Force Survey database User Guide*. European Commission, Eurostat, Directorate F, Unit F-2. [http://circa.europa.eu/irc/dsis/employment/info/data/eu\\_lfs/lfs\\_main/LFSuserguide/EULFS\\_Database\\_UserGuide\\_2008.pdf](http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/lfs_main/LFSuserguide/EULFS_Database_UserGuide_2008.pdf) . 04/09/2009.
- AA.VV. *Handbook on Statistical Disclosure Control*. CENTre of EXcellence for Statistical Disclosure Control, 2007. <http://neon.vb.cbs.nl/casc/handbook.htm>. 04/09/2009.
- AA.VV. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*. Roma: Istat, 2006 (Metodi e Norme n. 32).
- Skinner, Chris, e Natalie Shlomo. "Assessing Identification Risk in Survey Microdata Using Log-Linear Models". *Journal of the American Statistical Association* 103, 483 (2008): 989-1001.

## 5. Contatti

Per i dati dell'Indagine FOL:	infolav@istat.it	(DPTS/DCCV/FOL/B)
Per la tutela della riservatezza:	rilascio.microdati@istat.it	(DPTS/DCMT/MSS/C)

## Curatori dei paragrafi

Paragrafi 1, 2, 3.1: Flavio Foschi (DPTS/DCMT/MSS/C)

Paragrafo 3.2 : Luisa Franconi (DPTS/DCMT/MSS/C)<sup>1</sup>

---

<sup>1</sup> La tavola di raccordo tra codifiche concernenti il livello apicale di istruzione, esposta nella tabella 3, è stata elaborata e resa disponibile dalla DPTS/DCCV/FOL/B.