



## File di microdati per la ricerca

### **Rilevazione continua sulle Forze di Lavoro**

Periodo di riferimento: Trimestre 1 2021

### **Descrizione del file**

Marzo 2022

## INDICE

Introduzione .....	3
Le misure di protezione adottate.....	4
Riferimenti bibliografici.....	5

## Introduzione

I file di microdati per scopi di ricerca scientifica (MFR<sup>1</sup>) possono essere rilasciati su richiesta motivata e previa autorizzazione del Presidente dell'Istituto, purché siano resi anonimi e privi di ogni riferimento che ne permetta il collegamento con le unità intervistate.

Si segnala che relativamente a questa stessa indagine è disponibile anche il file mlcro.STAT, scaricabile via web (<http://www.istat.it/it/archivio/microdati-ad-uso-pubblico>). Il contenuto informativo di quest'ultimo è ridotto rispetto al file MFR a causa della maggiore protezione necessaria in mancanza di vincoli contrattuali.

Per la predisposizione del file MFR relativo all'indagine continua sulle Forze di Lavoro a partire dal primo trimestre 2021, è stata adottata una opportuna metodologia, allo scopo di limitare il rischio di violazione della riservatezza. Lo scenario di intrusione considerato è quello relativo all'identificazione spontanea, ossia conseguente a conoscenze a priori dell'utente che potrebbero permettere di attribuire correttamente i dati rilasciati alle unità della popolazione rilevata.

Le variabili coinvolte nel processo di protezione sono quelle che possono consentire l'associazione tra le informazioni e i rispondenti, ossia:

- individuano in maniera univoca le unità statistiche di rilevazione/analisi (come ad esempio indirizzo e codice fiscale);
- permettono di circoscrivere la popolazione alla quale appartengono i rispondenti e, da sole o in combinazione con altre, possono portare alla re-identificazione di uno o più record.

Mentre le prime sono cancellate dal file, le seconde vengono trattate statisticamente riducendone il contenuto informativo. Per il significato delle variabili e delle rispettive modalità si rinvia al questionario di indagine contenuto nella cartella dei metadati.

Si ricorda che gli utenti del file di microdati per scopi di ricerca scientifica sono obbligati ad adottare tutte le misure per la tutela della riservatezza cui si sono già impegnati all'atto della sottoscrizione del modulo di richiesta [http://www.istat.it/files/2010/10/modulo\\_ricerca.pdf](http://www.istat.it/files/2010/10/modulo_ricerca.pdf)

---

<sup>1</sup> MFR è l'acronimo per il file per la ricerca (Microdata File for Research). La documentazione metodologica di tali file è consultabile sul sito Istat (<http://www.istat.it/it/archivio/file-per-la-ricerca>).

## Le misure di protezione adottate

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 di seguito. Per significati e articolazioni delle restanti variabili, rimaste immutate, si rinvia al questionario di indagine e al relativo tracciato record.

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.

Per la predisposizione del file per la ricerca scientifica per l'indagine sulle Forze di Lavoro a partire dal primo trimestre 2021, sono state quindi adottate le misure di protezione di seguito specificate.

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

Codice Quartina

Codice Famiglia

## 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 1.

Tabella 1. Variabili recanti modificazioni delle modalità

NOME VARIABILE	PROVENIENZA	TIPO	DESCRIZIONE VARIABILE	VARIAZIONE
ETAM	r	NUM	età in anni compiuti	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
MFRFAM	r	CHAR	codice famiglia	ricodifica con metodo casuale
MFRIND	r	CHAR	codice individuo	ricodifica con metodo casuale
TFM	r	CHAR	tipologia familiare	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
STACIM	r	CHAR	stato civile	aggregazione delle modalità 3, 4 e 5 nella 3 (separati e divorziati) e ricodifica della 5 nella 4 (vedovi)
PROFM, PROFSM, PROFEM	r	CHAR	posizione professionale, posizione professionale secondaria, posizione	limitazione al terzo digit per le posizioni di legislatori e dirigenti

Ai fini di una più immediata comparabilità con i dati omologhi rilasciati da Eurostat sono state incluse le variabili elencate nella tabella 2:

Tabella 2: variabili inserite per comparabilità con i dati Eurostat

NOME VARIABILE	PROVENIENZA	TIPO	DESCRIZIONE VARIABILE
ISCO3D	r	CHAR	codifica ISCO delle professioni, omologa della CP2001 a tre digit
HATLEV_D	r	CHAR	Livello di istruzione secondo la Classificazione internazionale ISCED2011
HATLEV3MOD	r	CHAR	Livello di istruzione secondo la Classificazione internazionale ISCED2011 (3 modalità)
HATFIELD_D	r	CHAR	Area disciplinare del titolo di studio
HATWORK_D	r	CHAR	Esperienza formativa su un luogo di lavoro parte del programma di istruzione
EDUCFED4	r	CHAR	Partecipazione all'istruzione formale nelle 4 settimane precedenti
EDUCNFE4	r	CHAR	Partecipazione ad attività formative non-formali nelle 4 settimane precedenti

## Riferimenti bibliografici

- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K. E de Wolf, P.-P. (2012). Statistical Disclosure Control. Wiley.
- Willenborg, L. e de Waal, T. (1996). Statistical Disclosure Control in Practice. Lecture Notes in Statistics, 111, New York: Springer-Verlag.
- Willenborg, L. e de Waal, T. (2000). Elements of statistical disclosure control. Lecture Notes in Statistics, 115, New York: Springer-Verlag.
- AA.VV. La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione. Roma: Istat, 2006 (Metodi e Norme n. 32)