

n. 8/2007

**Sviluppo del processo di acquisizione e trattamento
informatico degli archivi relativi al modello
di dichiarazione 770. Anni 2004 - 2005**

L. Pizzicannella

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT, del SISTAN, o da studiosi esterni.

La *Rivista di Statistica Ufficiale* accoglie lavori che hanno come oggetto la misurazione dei fenomeni economici, sociali, demografici e ambientali, la costruzione di sistemi informativi e di indicatori, le questioni di natura metodologica, tecnologica o istituzionale connesse al funzionamento dei sistemi statistici e al perseguimento dei fini della statistica ufficiale.

I lavori pubblicati in *Contributi ISTAT* sono diffusi allo scopo di stimolare il dibattito intorno ai risultati preliminari di ricerca in corso.

I *Documenti ISTAT* forniscono indicazioni su linee, progressi e miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei *Contributi ISTAT* e nei *Documenti ISTAT* sono valutati preventivamente dai dirigenti dell'Istituto, mentre i lavori pubblicati nella *Rivista di Statistica Ufficiale* sono subordinati al giudizio di referee esterni.

Direttore responsabile della Rivista di Statistica Ufficiale: Patrizia Cacioli

Comitato di Redazione delle Collane Scientifiche dell'Istituto Nazionale di Statistica

Coordinatore: Giulio Barcaroli

Membri:	Corrado C. Abbate	Rossana Balestrino	Giovanni A. Barbieri
	Giovanna Bellitti	Riccardo Carbini	Giuliana Coccia
	Fabio Crescenzi	Carla De Angelis	Carlo M. De Gregorio
	Gaetano Fazio	Saverio Gazzelloni	Antonio Lollobrigida
	Susanna Mantegazza	Luisa Picozzi	Valerio Terra Abrami
	Roberto Tomei	Leonello Tronti	Nereo Zamaro

Segreteria: Gabriella Centi, Carlo Deli e Antonio Trobia

Responsabili organizzativi per la *Rivista di Statistica Ufficiale*: Giovanni Seri e Carlo Deli

Responsabili organizzativi per i *Contributi ISTAT* e i *Documenti ISTAT*: Giovanni Seri e Antonio Trobia

DOCUMENTI ISTAT

n. 8/2007

**Sviluppo del processo di acquisizione e trattamento
informatico degli archivi relativi al modello
di dichiarazione 770. Anni 2004 - 2005**

L. Pizzicannella()*

(*) ISTAT - Servizio Istituzioni pubbliche e private

Contributi e Documenti Istat 2007

Istituto Nazionale di Statistica
Servizio Produzione Editoriale

Produzione libraria e centro stampa:
Carla Pecorario
Via Tuscolana, 1788 - 00173 Roma

Indice

- 1 Premessa.
- 2 Gli archivi iniziali.
- 3 Le basi dati.
 - 3.1 Frontespizio.
 - 3.2 Lavoro dipendente.
 - 3.3 Lavoro autonomo.
 - 3.4 Anagrafica.
 - 3.5 Vincoli di integrità.
- 4 Il processo di trattamento degli archivi.
 - 4.1 Conversione degli archivi.
 - 4.2 Trattamento dei record G e H.
 - 4.2.1 Verticalizzazione.
 - 4.2.2 Ricostruzione delle sequenze.
 - 4.2.2.1 Variabili con più di un valore.
 - 4.2.2.2 Variabili concatenate.
 - 4.3 Trattamento dei record B.
 - 4.4 Trattamento dei dati anagrafici.
 - 4.5 Eliminazione dei doppi e caricamento in base dati.
- 5 Conclusioni.

Bibliografia.

1 Premessa.

Il presente documento contiene la descrizione del lavoro che è stato svolto per effettuare il trattamento preliminare degli archivi elettronici contenenti i modelli M-770 semplificati relativi agli anni di dichiarazione 2004 e 2005. Precisamente vengono descritti i risultati relativi alla prima parte del processo che trasforma gli archivi amministrativi acquisiti da Sogei in basi dati utili per la produzione di informazioni statistiche; l'attività si inserisce nel progetto "Trattamenti monetari non pensionistici (TMNP)" in corso di realizzazione nel servizio "Istituzioni pubbliche e private" (DCSS / SIP).

Le basi dati prodotte in questa prima fase sono un'immagine fedele degli archivi di origine e rappresentano le informazioni ad un livello di "semilavorato"; perché diventino utili per la produzione di informazioni statistiche, debbono essere effettuati ulteriori passi di controllo e correzione statistico - metodologica dei dati.

Nel grafico riportato a pagina 2 è rappresentato il processo a cui devono essere sottoposti gli archivi M-770 provenienti da Sogei per poter essere utilizzati per elaborazioni statistiche.

Il processo si compone di due step: il primo "Procedura di acquisizione e trattamento informatico dei dati" è stato effettuato ed è descritto in questo documento; il secondo "Procedura di controllo e correzione statistica" è in corso ed è stato affidato al sottogruppo "Gestione informatica dell'archivio" del gruppo di lavoro "Istat - Ministero dell'Economia e delle Finanze / Dipartimento delle politiche fiscali per l'utilizzo a fini statistici dell'archivio dei modelli 770", definito nell'ambito del protocollo d'intesa tra i due enti, stipulato in data 11 giugno 2004.

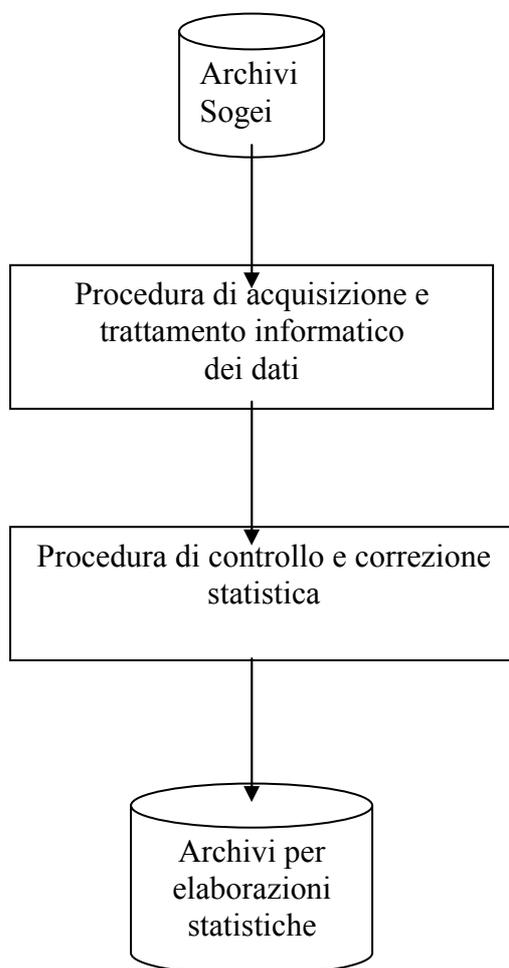
I modelli di dichiarazione 770 sono annuali ed è necessario distinguere tra l'anno di dichiarazione e l'anno di riferimento: il primo indica l'anno in cui il sostituto d'imposta compila e presenta la dichiarazione; il secondo indica l'anno al quale si riferiscono i dati contenuti. L'anno di dichiarazione e l'anno di riferimento differiscono di uno: pertanto, le dichiarazioni effettuate nell'anno 2004 contengono i dati relativi all'anno 2003 e le dichiarazioni dell'anno 2005 contengono i dati relativi all'anno 2004.

Nel presente documento si farà sempre riferimento all'anno di dichiarazione.

Prima dell'inizio di questa attività, nel servizio DCSS / SIP esisteva già un'ampia esperienza sull'argomento: erano stati trattati gli archivi contenenti i modelli 770 semplificati relativi agli anni di dichiarazione 2001, 2002 e 2003 e per ogni anno era stato rilasciato un data base relazionale composto da una tabella Frontespizio, una tabella Lavoro Dipendente e una tabella Anagrafica Lavoratori Dipendenti.

- La tabella Frontespizio contiene le informazioni di testata delle dichiarazioni e può essere considerata l'anagrafica dei sostituti di imposta per l'anno di riferimento.

Processo di trattamento archivi M770



- La tabella Lavoro Dipendente è strutturata per gruppi di variabili e contiene i dati identificativi, i dati fiscali, i dati previdenziali ed assistenziali, e i dati relativi all'assistenza fiscale dei lavoratori dipendenti ed assimilati.
- La tabella Anagrafica contiene i dati anagrafici dei lavoratori dipendenti percettori delle somme riportate nei modelli di dichiarazione.

La metodologia utilizzata e le lavorazioni effettuate sono descritte in dettaglio nel documento "Processo di Acquisizione e trattamento Informatico degli Archivi relativi al Modello di Dichiarazione 770" a cura del Dr. Vincenzo Spinelli (DCSS / SIP / A).

Il lavoro effettuato per gli anni 2004 e 2005 ha prodotto un data base relazionale composto da una tabella Frontespizio, una tabella Lavoro Dipendente, una tabella Lavoro Autonomo, una tabella Anagrafica Lavoratori Dipendenti e una tabella Anagrafica Lavoratori Autonomi; rispetto agli anni precedenti, sono state aggiunte la tabella Lavoro Autonomo e la tabella Anagrafica Lavoratori Autonomi.

- La tabella Lavoro Autonomo contiene i dati relativi alle somme percepite dai lavoratori autonomi.

- La tabella Anagrafica Lavoratori Autonomi contiene i dati anagrafici dei lavoratori autonomi percettori delle somme riportate nei modelli di dichiarazione.

Il lavoro svolto ha tenuto conto, ovviamente, sull'esperienza precedente. Sono state acquisite la metodologia e la strumentazione utilizzata per il trattamento dei dati relativi all'anno 2003 e sono state apportate ad esse le implementazioni che si sono rese necessarie per l'implementazione delle nuove tabelle e per tener conto delle variazioni riscontrate negli archivi.

2. Gli archivi iniziali.

Gli archivi iniziali sono stati forniti a Istat da Sogei e si trovano in uno stato grezzo, perché provengono dagli archivi di "raccolta telematica delle dichiarazioni" della Sogei stessa, e non hanno subito alcun trattamento preliminare.

I dati sono su supporto magnetico (DVD); la fornitura è divisa in invii; ogni invio corrisponde a una directory; ogni directory contiene un certo numero di archivi. In tabella 1 sono riportati il numero di DVD, il totale degli invii e il numero complessivo degli archivi che costituiscono la fornitura per gli anni 2004 e 2005.

Tabella 1. Riepilogo acquisizioni archivi 770

Anno di dichiarazione	Totale DVD	Totale invii	Totale archivi
2004	5	5	261
2005	2	5	234

In tabella 2 sono riportate le dimensioni degli archivi espresse in Megabyte e la numerosità complessiva dei record.

Tabella 2. Dimensioni degli archivi acquisiti

Anno di dichiarazione	Dimensione dati (MB)	Numero record
2004	130.828	64.694.126
2005	140.400	69.589.153

Gli archivi informatici derivano dal modello cartaceo che si compone di una serie di quadri, a ciascuno dei quali corrisponde un determinato tipo record. Ogni anno il modello di dichiarazione è soggetto a modifiche per aggiunta, eliminazione e / o spostamento di campi: ciò determina una variazione nella composizione dell'archivio e, di conseguenza, un impatto sulle procedure di trattamento informatico dei dati, che, pertanto, occorre verificare e adeguare di anno in anno.

Per gli anni 2004 e 2005 sono stati acquisiti i quadri ai quali corrispondono i tipi record A, B, G, H, Z. La novità è costituita dal tipo record H che negli anni precedenti non era stato preso in considerazione.

- Record di tipo A. E' il record di testa della fornitura e contiene i dati identificativi della fornitura e del soggetto responsabile dell'invio telematico; viene scartato durante le elaborazioni successive.
- Record di tipo B. Contiene i dati relativi al sostituto d'imposta e al rappresentante firmatario della dichiarazione. Questo tipo di record corrisponde al Frontespizio della dichiarazione e può essere considerato l'anagrafica dell'azienda.
- Record di tipo G. Contiene i dati relativi al quadro "Comunicazione dati certificazioni lavoro dipendente, assimilati ed assistenza fiscale" (dati anagrafici, dati fiscali, dati previdenziali ed assistenziali, dati relativi all'assistenza fiscale).

- Record di tipo H. Contiene i dati relativi al quadro “Comunicazione dati certificazioni lavoro autonomo, provvigioni e redditi diversi” (dati anagrafici e dati relativi al percettore delle somme). Ha una struttura molto simile a quella del record G.
- Record di tipo Z. E’ il record di coda della fornitura e contiene alcuni dati riepilogativi della fornitura stessa; non viene preso in considerazione nelle elaborazioni successive.

Gli archivi acquisiti da Sogei contengono soltanto i campi valorizzati, ma tutti i record hanno lunghezza fissa; questo comporta la presenza di lunghe sequenze di caratteri “spazio” tra la fine delle informazioni significative e la fine fisica del record.

La lunghezza dei record dell’archivio può variare di anno in anno e dipende dalle specifiche tecniche della fornitura.

Lo schema logico dei dati è riportato nel grafico di pagina 5.

L’entità rappresentata negli archivi M-770 è la dichiarazione, identificabile tramite il campo Identificativo Telematico (IdTel), un campo alfanumerico di 25 caratteri che viene attribuito in base all’ordine di ricezione della dichiarazione stessa da parte del sistema informatico dell’Agenzia delle Entrate e che permette di identificare univocamente tutti i record di una stessa dichiarazione.

Questo valore è fondamentale per distinguere le dichiarazioni effettuate da un’azienda nel corso dello stesso anno: ogni IdTel è attribuibile a una sola dichiarazione di una sola azienda; le dichiarazioni multiple, che si verificano nei casi in cui un’azienda presenta più dichiarazioni nello stesso anno, hanno valori di IdTel differenti.

In ogni dichiarazione deve essere presente un solo Frontespizio, mentre possono essere presenti più comunicazioni relative al lavoro dipendente o al lavoro autonomo.

Questo significa che a fronte di un record tipo B ci possono essere più record tipo G e / o H; tutti hanno in comune lo stesso valore di IdTel.

Il tipo record B è a tracciato record fisso; i tipi record G e H che, come ho già detto, hanno struttura molto simile hanno tracciato fisso – variabile.

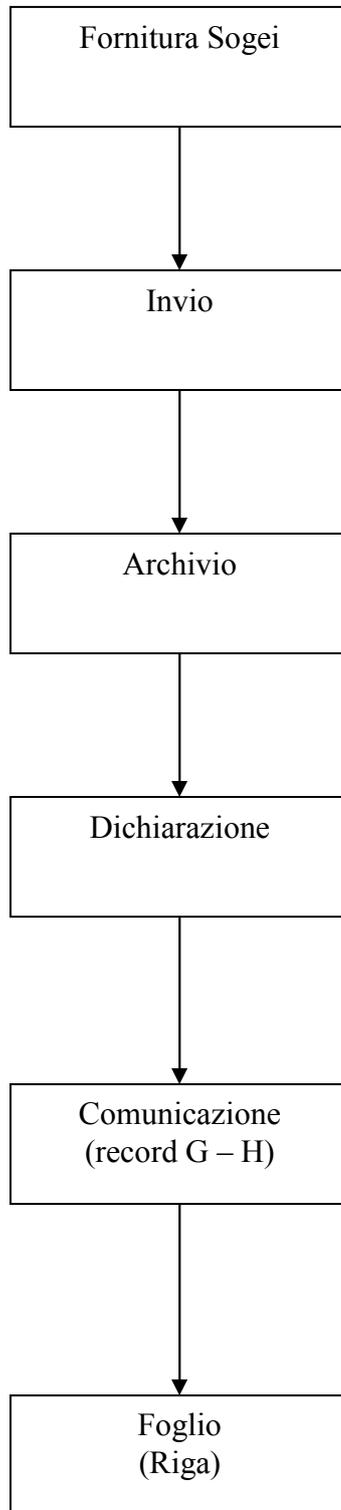
Questi tre tipi record hanno una parte comune fissa che viene utilizzata per creare il legame tra i record presenti nella fornitura e comprende i seguenti campi:

- Identificativo di fornitura. Ha una lunghezza di otto caratteri e si compone di tre valori costanti: “IST”, “7” e “anno”. Il primo indica l’inizio di ogni riga, il secondo indica che si tratta di modelli 770, il terzo riporta l’anno di dichiarazione (nel nostro caso 2004 o 2005).
- Progressivo di fornitura. E’ il numero progressivo della fornitura; questa informazione non è significativa ai nostri fini e, pertanto, viene trascurata.
- Identificativo telematico.
- Tipo record. E’ lungo un carattere e permette di individuare il record da elaborare nel nostro caso B, G o H.

Il caso del record B a tracciato fisso è il più semplice, perché di ogni variabile sono noti a priori posizione iniziale, lunghezza e formato e tutti i record hanno identica struttura.

Il caso dei record G e H a tracciato fisso – variabile è più complicato, perché, oltre a una parte fissa, i record presentano una struttura di campi “a serrare”; questo implica che una variabile è presente solo se ha un valore significativo. Sogei ha scelto questo tipo di organizzazione, perché le variabili valorizzate nei quadri del modello corrispondenti ai record G e H sono una percentuale minima rispetto a tutte quelle presenti nel modello stesso. Questa impostazione implica che la sequenza delle variabili presenti in ogni record non è nota a priori e che per individuare una variabile occorre introdurre una serie di “coordinate” che la identifichino univocamente.

Schema logico dei dati acquisiti da Sogei



Ogni variabile è formata dai campi sezione, numero di riga, numero di colonna, valore che hanno il seguente significato:

- Sezione. Il quadro “Comunicazione dati certificazioni lavoro dipendente, assimilati ed assistenza fiscale” relativo al lavoro dipendente è diviso in quattro sezioni: anagrafica, dati fiscali, dati previdenziali ed assistenziali, dati relativi all’assistenza fiscale, identificati rispettivamente dalle sequenze “DA”, “DB”, “DC”, “DD”. Il quadro “Comunicazione dati certificazioni lavoro autonomo, provvigioni e redditi diversi” relativo al lavoro autonomo contiene una sola sezione (dati anagrafici e dati relativi al percettore delle somme) identificata dalla sequenza “AU”.
- Numero di riga. Le informazioni relative ad un lavoratore dipendente o autonomo possono aver bisogno di più di un foglio per essere descritte (è questo il caso di un lavoratore dipendente che cambia qualifica nel corso dell’anno); per distinguere i vari fogli di una comunicazione relativa allo stesso lavoratore dipendente viene utilizzato il numero di riga.
- Numero di colonna. E’ Il numero che nel modello cartaceo compare a fianco di ciascuna variabile nei quadri relativi al lavoro dipendente e al lavoro autonomo.
- Valore. E’ il valore effettivo assunto dalla variabile. Alcune variabili, a parità di coordinate, hanno più di un valore; ad esempio, con riferimento all’anno di dichiarazione 2004, la variabile 105 della sezione “DC” (data inizio) ha due valori: giorno e mese. La lunghezza massima prevista per il campo valore è 16 caratteri.

Concludendo sezione, numero di riga e numero di colonna sono le coordinate che servono ad identificare la variabile, mentre valore rappresenta il valore assunto dalla variabile nel record considerato. Un esempio: la sequenza DC001003D: identifica la variabile della sezione DC con numero di riga 1, numero di colonna 3; il suo valore è D.

3. Le basi dati.

Il data base relativo agli anni 2001, 2002, 2003 è composto da una tabella Frontespizio, una tabella per il Lavoro Dipendente e una tabella Anagrafica Lavoratori Dipendenti.

A partire dall’anno 2004 è stato elaborato anche il tipo record H: ciò ha portato al rilascio delle nuove tabelle Lavoro Autonomo e Anagrafica Lavoratori Autonomi.

Di seguito è riportata una breve descrizione delle tabelle che compongono il data base relativo agli anni 2004 e 2005.

3.1 Frontespizio.

La tabella Frontespizio contiene le informazioni di testata delle dichiarazioni e può essere considerata, in prima approssimazione, l’anagrafica dei sostituti d’imposta per l’anno al quale si riferisce.

Questa tabella è strutturata in una parte fissa e una parte variabile.

La parte fissa contiene i dati identificativi della dichiarazione e si mantiene abbastanza stabile nel tempo; la parte variabile contiene i dati di maggior dettaglio e cambia di anno in anno, in seguito alle modifiche apportate al modello cartaceo.

Nella parte fissa sono presenti anno di dichiarazione, tipo di modello, invio, identificativo telematico, codice fiscale dell’azienda, flag di validazione del codice fiscale, divisa, flag di identificazione della dichiarazione.

Il campo anno nel nostro caso vale 2004 o 2005; il campo modello indica che si tratta di 770.

Il campo invio, presente dal 2005, indica la directory contenente la dichiarazione ed è stato aggiunto per creare un legame tra la base dati e gli archivi originari forniti da Sogei.

Il flag di validazione del codice fiscale, aggiunto a partire dall'anno 2004, è una misura dell'affidabilità di questo campo; la divisa indica la valuta utilizzata nella dichiarazione che, per gli anni 2004 e 2005, vale sempre euro.

I flag di identificazione indicano se la dichiarazione è correttiva nei termini, integrativa, o effettuata in presenza di eventi eccezionali. Dei primi due parleremo più diffusamente nel seguito del documento; il terzo è impostato quando un sostituto d'imposta, al verificarsi di eventi particolari, si avvale della sospensione del termine di presentazione della dichiarazione.

Nella parte variabile i nomi delle colonne sono codificati coerentemente con il tracciato fornito da Sogei e possono variare da un anno all'altro: ogni variabile è individuata dal prefisso B e da un numero di quattro cifre che rappresenta la colonna di inizio del campo nel tracciato record Sogei. Un esempio: la variabile B0462 è presente nell'anno 2004 e rappresenta il campo che inizia a colonna 462 (comune di residenza anagrafica o di domicilio fiscale del dichiarante); nell'anno 2005 la variabile B0462 non esiste e la variabile con lo stesso significato è codificata B0453.

3.2 Lavoro dipendente.

La tabella Lavoro Dipendente contiene le comunicazioni relative ai lavoratori dipendenti ed assimilati ed è strutturata in una parte fissa e in una parte variabile.

La parte fissa contiene i dati identificativi della comunicazione ed è abbastanza stabile nel tempo; la parte variabile contiene i dati dichiarati e muta di anno in seguito alle modifiche apportate al modello cartaceo.

I campi presenti nella parte fissa sono: anno di dichiarazione, identificativo telematico, codice fiscale dell'azienda, codice fiscale del lavoratore, progressivo di comunicazione, tipo di operazione e numero di riga.

Il codice fiscale dell'azienda è stato aggiunto a partire dall'anno 2004.

Il progressivo di comunicazione, aggiunto con l'anno 2005, è il numero riportato sulla parte del modello M-770 cartaceo relativa ai dati delle certificazioni di lavoro dipendente e raggruppa tutti i dati relativi a un singolo lavoratore.

Il campo tipo di operazione è stato aggiunto con l'anno 2004 ed è impostato in presenza di una comunicazione correttiva o integrativa "parziale"; di esso parleremo nel seguito del documento.

Il numero di riga permette di distinguere i vari fogli di una comunicazione.

La parte variabile è costituita da quattro gruppi di variabili DA (anagrafica dipendente), DB (dati fiscali), DC (dati previdenziali ed assistenziali), DD (assistenza fiscale): questi gruppi seguono lo schema del quadro "Comunicazione dati certificazioni lavoro dipendente, assimilati ed assistenza fiscale" del modello M-770. Ogni variabile è individuata dall'identificativo del raggruppamento e dal progressivo riportato nel modello di dichiarazione.

I gruppi di variabili sono stabili nel tempo, ma, da un anno all'altro, la loro composizione è diversa; un esempio: il codice DC055, che rappresenta il 55-esimo campo del quadro C, per l'anno di dichiarazione 2005 identifica la variabile "Bonus L. 243/2004", per l'anno 2004 la variabile "Importo non trattenuto".

3.3 Lavoro autonomo.

La tabella Lavoro Autonomo è stata aggiunta con l'anno 2004 e contiene le comunicazioni relative ai lavoratori autonomi. E' strutturata in una parte fissa e in una parte variabile ed ha una struttura simile, ma più semplice, di quella del lavoro dipendente.

La parte fissa contiene i dati identificativi della comunicazione ed è abbastanza stabile nel tempo; la parte variabile contiene i dati dichiarati e muta di anno in seguito alle modifiche apportate al modello cartaceo.

La parte fissa contiene gli stessi campi della tabella Lavoro Dipendente: anno di dichiarazione, identificativo telematico, codice fiscale dell'azienda, codice fiscale del lavoratore, progressivo di comunicazione, tipo di operazione e numero di riga; il significato è, ovviamente, lo stesso.

La parte variabile comprende un solo gruppo di variabili. Il gruppo segue lo schema del quadro "Dati certificazioni lavoro autonomo, provvigioni e redditi diversi" del modello M-770; le variabili sono identificate dalla sigla del raggruppamento (AU) e dal progressivo riportato sul modello di dichiarazione: ad esempio per l'anno 2004 AU016 indica la variabile numero 16 del raggruppamento AU (la causale dell'erogazione delle somme).

Come nel caso del lavoro dipendente, il gruppo può essere considerato stabile nel tempo; da un anno all'altro, però, la sua composizione può essere diversa. A titolo di esempio, il quadro "Dati certificazioni lavoro autonomo, provvigioni e redditi diversi" del modello M-770 relativo all'anno di dichiarazione 2005 contiene due variabili in più dell'analogo quadro relativo all'anno 2004 (AU031 - Contributi previdenziali a carico del datore di lavoro e AU032 - Contributi previdenziali a carico del lavoratore).

3.4 Anagrafica.

La tabella Anagrafica contiene i dati identificativi dei percettori delle somme ed è stata ristrutturata con la lavorazione relativa all'anno di dichiarazione 2004, con l'aggiunta dei dati relativi ai lavoratori autonomi.

E' logicamente divisa in due parti, una relativa alle persone fisiche, l'altra relativa agli altri soggetti. L'anno di dichiarazione, il codice fiscale e il flag di validazione del codice fiscale sono comuni ad entrambe.

Oltre ai dati comuni, per le persone fisiche sono presenti: cognome, nome, sesso, data di nascita, denominazione del comune e sigla provincia di nascita, indirizzo di residenza composto da denominazione del comune, sigla provincia, via/piazza con relativa denominazione, numero civico e cap.

Per gli altri soggetti sono presenti denominazione, forma giuridica, indirizzo del domicilio fiscale composto da denominazione del comune, sigla provincia, via/piazza con relativa denominazione, numero civico e cap.

3.5 Vincoli di integrità.

Le tabelle Frontespizio, lavoro Dipendente, Lavoro Autonomo, Anagrafica non sono legate tra loro da vincoli di integrità referenziale; questo perché si è scelto che esse siano l'immagine più fedele possibile degli archivi d'origine, lasciando ogni controllo di merito alla procedura di controllo e correzione che deve essere effettuata affinché il data base diventi utilizzabile per la produzione statistica.

Alcune variabili, tuttavia, permettono di legare tra loro gruppi di dati e possono rappresentare il punto di partenza per le elaborazioni successive:

- **Identificativo telematico.** Lega ogni riga della tabella Frontespizio a tutte le righe dettaglio presenti sulle tabelle Lavoro Dipendente e Lavoro Autonomo; in altre parole, permette di ricostruire una dichiarazione.
Questo stesso risultato può essere ottenuto utilizzando il codice azienda presente sulle tabelle Frontespizio, Lavoro Dipendente e Lavoro Autonomo.
- **Numero di comunicazione.** Permette di raggruppare tutti i fogli utilizzati per descrivere un dipendente nell'ambito di una dichiarazione.
- **Codice fiscale del lavoratore dipendente / autonomo.** Permette l'aggancio tra i dati presenti nella dichiarazione e i dati anagrafici del lavoratore.

4 Il processo di trattamento degli archivi.

In questo paragrafo sono descritte le operazioni che sono state effettuate per trattare gli archivi iniziali acquisiti da Sogei e caricarli nelle basi dati descritte nel paragrafo precedente.

Il processo si articola nelle seguenti fasi:

- 1) Conversione degli archivi.
- 2) Trattamento dei record G e H.
- 3) Trattamento dei record B.
- 4) Trattamento dei dati anagrafici.
- 5) Eliminazione dei doppioni e caricamento in base dati.

Tutte le elaborazioni sono state effettuate utilizzando il sistema operativo GNU-Linux; anche l'ambiente di sviluppo è GNU (gcc, gawk, bash scripting).

Come DBMS è stato scelto Oracle; le tabelle sono state caricate con il tool sql loader.

4.1 Conversione degli archivi.

E' la fase di controllo e correzione alla quale devono essere sottoposti gli archivi iniziali prima del processo di trattamento vero e proprio.

L'analisi preliminare degli archivi acquisiti da Sogei ha evidenziato le seguenti problematiche:

- Presenza di caratteri di fine riga e di fine file che in ambiente Unix / Linux non sono più necessari e possono creare problemi.
- Presenza di caratteri "non stampabili" (caratteri di fine file o di fine riga in posizione non corretta, caratteri dovuti ai meccanismi di conversione da una codifica all'altra, caratteri generati dall'errata conversione di caratteri con accenti).
- Presenza del "pipe": questo carattere deve essere eliminato, perché è utilizzato come separatore di campo in tutte le elaborazioni successive e, pertanto, la sua presenza nei dati di ingresso genera errori di interpretazione durante il processo di ricostruzione delle variabili.

Le azioni intraprese sono state nell'ordine le seguenti:

- 1) Utilizzo del comando di sistema dos2unix per eliminare i caratteri DOS di fine riga e di fine file.
- 2) Rimozione manuale dei caratteri di fine file in posizione anomala (non come ultimo carattere dell'ultimo record dell'archivio); senza questo intervento, tutti i record successivi sarebbero stati scartati.
- 3) Utilizzo del programma ascii, realizzato da DCSS / SIP /A che svolge le seguenti funzioni:
 - Compattazione delle righe. Partendo dall'ipotesi che tutti i record dell'archivio devono iniziare con la sequenza "IST", una riga che inizia con una sequenza differente non dà luogo a un nuovo record, ma viene accodata alla riga precedente.
 - Sostituzione dei caratteri "non stampabili" con spazi.
 - Sostituzione del carattere "pipe" con spazi.

Nella tabella 3 sono riportati i risultati di questa prima parte del processo di lavorazione.

Nei caratteri "sporchi" sono compresi i pipe e i caratteri "non stampabili", con esclusione dei caratteri anomali di fine riga e fine file che sono conteggiati a parte. L'alto numero di caratteri "sporchi" riscontrato negli archivi relativi all'anno 2005 è dovuto alla massiccia presenza di caratteri "non stampabili" utilizzati in alcuni casi tra la fine delle informazioni significative e la fine fisica di record o campi.

Non abbiamo, invece, quantificato esattamente l'informazione che si perde a causa dell'errata conversione dei caratteri con accenti: stando alle prove effettuate a campione, i casi sono relativamente pochi; inoltre, i caratteri accentati si trovano nei dati anagrafici (nomi, cognomi, denominazioni), il cui trattamento nel nostro ambito non è essenziale.

Tabella 3. Risultati del passo di conversione degli archivi.

Anno di dichiarazione	Righe accodate	Fine file anomali	Caratteri sporchi
2004	6	0	95.201
2005	7	3	1.077.514

Nella tabella 4 è riportata la numerosità dei tipi record acquisiti risultante alla fine della fase di conversione: Frontespizio (record "B"), Lavoro Dipendente (record "G"), Lavoro Autonomo (record "H"); la voce Altri comprende i record "A" e "Z" che, come è stato già detto, non vengono considerati dal processo di lavorazione.

Tabella 4. Numerosità dei tipi record acquisiti

Anno di dichiarazione	Frontespizio	Lavoro dipendente	Lavoro autonomo	Altri
2004	3.784.447	48.156.395	12.753.212	72
2005	4.146.466	51.755.492	13.687.187	8

4.2 Trattamento dei record G e H.

Nel paragrafo "Gli archivi iniziali" abbiamo già detto che i record G e H hanno entrambi un tracciato record costituito da una parte fissa e da una parte "a serrare", nella quale le variabili sono presenti soltanto se hanno un valore significativo e sono individuate dalla tre coordinate sezione, numero di riga, numero di colonna.

Pertanto, il software utilizzato per passare da questi tipi record al formato richiesto per il caricamento delle tabelle lavoro Dipendente e Lavoro Autonomo è logicamente simile ed è costituito dal passo di verticalizzazione e dal passo di ricostruzione delle sequenze.

Il trattamento del Lavoro Autonomo è iniziato con l'anno di dichiarazione 2004; di conseguenza, il software per l'elaborazione dei record "H" è stato sviluppato ex-novo.

4.2.1 Verticalizzazione.

Il passo di verticalizzazione permette di identificare e separare le singole variabili. Per ognuna delle variabili presenti nella parte a "serrare" di ogni record G e H dell'archivio acquisito da Sogei viene generata una riga contenente:

- I campi presenti nella parte fissa delle tabelle Lavoro Dipendente e Lavoro Autonomo (anno, identificativo telematico, codice fiscale dell'azienda, codice fiscale del lavoratore, progressivo di comunicazione, tipo di operazione).
- Le coordinate che permettono di individuare una variabile (sezione, numero di riga, numero di colonna).
- Il valore della variabile.

L'output prodotto costituisce l' "archivio verticalizzato" nel quale ogni record rappresenta una variabile; l'ordine dei record rispetta la sequenza delle variabili nell'archivio Sogei.

Durante questa fase vengono standardizzati i formati delle variabili; in particolare, i campi data sono convertiti in un unico formato che prevede anno (quattro cifre), mese, giorno.

Nessuna operazione, comunque, modifica il contenuto dei dati.

4.2.2 Ricostruzione delle sequenze.

Il passo di ricostruzione delle sequenze trasforma l' "archivio verticalizzato" in un file avente il tracciato richiesto dal software di caricamento delle tabelle Lavoro Dipendente o Lavoro Autonomo; in questo archivio sono presenti tutte le variabili: quelle con valore significativo conservano il valore iniziale, quelle non significative non sono valorizzate.

La posizione nel file finale di ogni variabile dell' "archivio verticalizzato" è determinata da codice sezione e numero di colonna. Se le coordinate non sono significative, se cioè sono presenti codici sezione, numeri di riga e / o di colonna non ammessi, (es. quadri non esistenti o numeri di colonna maggiori del massimo previsto nelle specifiche Sogei), il record viene scartato: a questo livello di lavorazione, non si può fare alcuna ipotesi sulle variabili e risulta praticamente impossibile recuperare questi casi, assegnandoli alla cella più probabile. Fortunatamente questo tipo di errori riguarda un numero molto ristretto di casi e tende a ridursi con il passare degli anni per il miglioramento delle procedure Sogei di acquisizione e controllo dei dati.

La scrittura dell'archivio di output viene effettuata a ogni rottura di uno dei seguenti campi: identificativo telematico, codice fiscale lavoratore, numero di comunicazione (a partire dall'anno 2005), numero di riga. Generalmente viene scritto un solo record; ne vengono scritti due nei casi in cui in una comunicazione della stessa azienda relativa allo stesso dipendente esistono una o più variabili con codice sezione, numero riga, numero colonna identici e valore differente. E' stato deciso così perché si vuole che il data base prodotto da questa fase della lavorazione sia l'immagine più fedele possibile degli archivi di origine, demandando ogni scelta di merito alla fase di controllo e validazione che deve essere effettuata prima che gli archivi possano essere utilizzati per la produzione di informazioni statistiche.

Durante la ricostruzione delle sequenze abbiamo dovuto affrontare altre due problematiche:

- Variabili con più di un valore.
- Variabili concatenate.

4.2.2.1 Variabili con più di un valore.

Abbiamo già detto che una variabile dell'archivio Sogei è individuata dalle tre coordinate codice sezione, numero di riga, numero di colonna.

Quasi tutte le variabili hanno un solo valore; ve ne sono tuttavia alcune che ne hanno più di uno; un tipico esempio è la durata temporale espressa in anno / mese.

In questi casi nell'archivio acquisito da Sogei sono presenti due o più variabili appartenenti alla stessa dichiarazione e relative allo stesso soggetto aventi stesso codice sezione e stesso numero riga. Il numero di colonna, che nel caso di variabili con un solo valore è un campo numerico di tre caratteri, ha il primo carattere alfabetico ("A" per la prima variabile, "B" per la seconda, "C" per la terza); il secondo e il terzo numerici e uguali.

Di conseguenza, nell' "archivio verticalizzato" sono presenti due o più record, uno per ogni valore. Il codice sezione e il numero riga sono uguali; il numero di colonna è diviso in due sottocampi: un carattere alfabetico, diverso per ogni record, e la parte numerica, lunga due caratteri, identica per tutti.

Sull'archivio di output tutti i valori sono inseriti in un unico campo e sono separati da opportuni caratteri di inizio e fine sequenza che li rendano distinguibili.

4.2.2.2 Variabili concatenate.

Le variabili concatenate sono utilizzate nella parte “a serrare” dei record G e H, se la lunghezza del dato da inserire nel campo valore eccede la lunghezza massima prevista (16 caratteri). Il loro trattamento è iniziato con l’elaborazione relativa ai dati dell’anno 2004.

In questi casi nell’archivio acquisito da Sogei ci sono due, o anche più, variabili appartenenti alla stessa dichiarazione e relative allo stesso soggetto aventi codice sezione, numero di riga e numero di colonna identici; le variabili successive alla prima hanno il primo carattere del campo valore impostato a “+” per indicare che si tratta della continuazione del valore della variabile precedente e non di una nuova variabile.

Nell’ “archivio verticalizzato” è presente una riga per ogni variabile.

Nell’archivio di output figura una sola variabile, il cui valore è il risultato della concatenazione dei valori delle variabili componenti; il carattere “+” è stato sostituito dal carattere “ ”.

Questa situazione si verifica soltanto per alcuni campi con formato alfanumerico, come nei casi AXXXX013 (località di residenza estera) e AXXXX014 (via e numero civico); la sequenza XXX rappresenta il numero di riga.

4.3 Trattamento dei record B.

Come detto in precedenza, il record di tipo B è a tracciato fisso: ogni record contiene tutte le variabili e sono note posizione iniziale, lunghezza e tipologia di ognuna di esse; se una variabile non è valorizzata, la sua posizione è occupata da un numero di spazi pari alla sua lunghezza.

In questo caso il processo di elaborazione deve individuare i campi presenti e convertire il tracciato Sogei nel tracciato richiesto dal software di caricamento della tabella Frontespizio, avendo cura di mantenere la corretta sequenza delle variabili; non è necessaria la fase di verticalizzazione, né la fase di ricostruzione delle sequenze.

Per questo tipo record il tracciato relativo all’anno 2004 è diverso rispetto a quello del 2003 e, analogamente, il tracciato dell’anno 2005 è differente da quello del 2004; pertanto, per entrambi, è stato necessario modificare il software di elaborazione e il tracciato della tabella Frontespizio, in funzione di posizione iniziale, lunghezza e formato delle variabili nel nuovo tracciato.

4.4 Trattamento dei dati anagrafici.

Diversamente dal 2003, in cui erano stati forniti a parte, per gli anni 2004 e 2005 i dati anagrafici dei percettori delle somme sono parte integrante della fornitura e si trovano nella parte fissa del record G per i lavoratori dipendenti e del record H per i lavoratori autonomi.

Nel caso delle persone fisiche sono presenti le seguenti informazioni: cognome, nome, sesso, data, nome del comune e sigla provincia di nascita, indirizzo di residenza, completo di denominazione del comune, sigla provincia, via, numero civico e cap.

Nel caso dei soggetti differenti dalle persone fisiche le informazioni presenti sono: denominazione, forma giuridica e indirizzo del domicilio fiscale; anche in questo caso l’indirizzo comprende denominazione del comune, sigla provincia, via, numero civico e cap.

Il software di trattamento dei dati anagrafici è stato sviluppato ex-novo. Il problema che abbiamo dovuto affrontare è stato come stabilire se si fosse in presenza di dati relativi a una persona fisica o a un soggetto differente.

Dall’analisi degli archivi è risultato che non sempre i dati anagrafici presenti nella parte fissa del record G sono i dati tipici di una persona fisica, né che quelli presenti nel record H sono i dati tipici di un soggetto differente; in generale è così, ma ci sono delle eccezioni. Inoltre, non è presente nei tracciati record alcun flag che permette di distinguere tra i due casi, né può essere di aiuto la

lunghezza del codice fiscale, visto che, sugli archivi, a un codice fiscale lungo 16 caratteri possono corrispondere dati tipici di un soggetto differente da una persona fisica e, in modo analogo, a un codice fiscale lungo 11 caratteri possono corrispondere dati tipici di una persona fisica.

4.5 Eliminazione dei doppioni e caricamento in base dati.

Nella tabella 5 è riportata la numerosità dei record ricostruiti dopo il trattamento dei record B, G, H, compresi i dati anagrafici.

Tabella 5. Numerosità dei record ricostruiti.

Anno di dichiarazione	Lavoro dipendente	Lavoro autonomo	Frontespizio	Anagrafica persone fisiche	Anagrafica persone giuridiche
2004	53.138.122	12.848.219	3.784.447	58.856.259	2.008.716
2005	60.484.787	13.778.772	4.146.466	63.212.515	2.183.138

Prima del caricamento in base dati, gli archivi prodotti dai processi descritti in precedenza sono stati sottoposti al trattamento per l'eliminazione dei doppioni: di due o più record completamente uguali, uno solo viene mantenuto, gli altri sono scartati.

Quest'operazione elimina i duplicati "fisici", cioè i record aventi tutti i campi uguali; ma non i duplicati "logici", cioè i casi in un lavoratore (dipendente o autonomo) è presente in due o più dichiarazioni presentate dall'azienda in momenti diversi, o in due comunicazioni distinte facenti parte della stessa dichiarazione. Per risolvere questi casi è necessario implementare la procedura di controllo e correzione che renda la base dati adatta alla produzione di informazioni statistiche.

Nella tabella 6 è riportata la numerosità dei record distinti dopo l'eliminazione dei doppioni.

Tabella 6. Numerosità dei record distinti (dopo l'eliminazione dei doppioni)

Anno di dichiarazione	Lavoro dipendente	Lavoro autonomo	Frontespizio	Anagrafica persone fisiche	Anagrafica persone giuridiche
2004	53.113.167	12.815.035	3.784.447	34.830.224	258.424
2005	60.484.787	13.778.756	4.146.466	35.558.309	275.711

Nella tabella 7 è riportata la percentuale dei record distinti risultante alla fine del processo di trattamento.

Tabella 7. Percentuale dei record distinti

Anno di dichiarazione	Lavoro dipendente	Lavoro autonomo	Frontespizio	Anagrafica persone fisiche	Anagrafica persone giuridiche
2004	99,95	99,75	100,00	59,17	12,86
2005	100,00	99,99	100,00	56,25	12,63

5 Conclusioni.

L'entità rappresentata negli archivi iniziali è la dichiarazione formata da un frontespizio (record B), una o più comunicazioni relative al lavoro dipendente (record G) e / o una o più comunicazioni relative al lavoro autonomo (record H).

Nel processo di trattamento degli archivi, descritto nel paragrafo precedente, si è posto l'accento sulla ricostruzione delle sequenze delle variabili appartenenti allo stesso riquadro e sulla trasformazione dei record di input in tabelle aventi struttura omogenea. Nella fase di progettazione del processo si è scelto che il data base risultante fosse l'immagine più fedele possibile degli archivi di origine; pertanto, sono stati effettuati solo alcuni controlli: eliminazione di errori sintattici, ricostruzione della sequenza delle variabili, eliminazione dei record doppi, standardizzazione dei formati, come nel caso dei campi data che sono stati tutti convertiti nel formato AAAA-MM-GG. In definitiva, riprendendo quanto già detto nella premessa, la base dati implementata rappresenta le informazioni ad un livello di "semilavorato" e non può essere utilizzato per la produzione di informazioni statistiche.

Perché ciò sia possibile occorre attendere la conclusione della "Procedura di controllo e correzione statistica", che è attualmente in corso a cura del sottogruppo "Gestione informatica dell'archivio", come già detto nella Premessa di questo documento. Questa procedura ha il compito di risolvere, in particolare, le problematiche connesse alla presenza di dichiarazioni multiple per la stessa azienda.

A questo riguardo si possono fare le seguenti considerazioni:

I motivi per cui un'azienda è presente nell'archivio 770 con più di una dichiarazione possono essere molteplici; riportiamo quelli noti allo stato attuale della lavorazione:

- 1) Errori. Un sostituto invia la stessa dichiarazione in momenti diversi, senza alcun motivo valido; in questo caso le dichiarazioni sono identiche, a meno dell'identificativo telematico.
- 2) Spilt dimensionale. Alcuni sostituti d'imposta non riescono a inserire tutte le comunicazioni in un'unica dichiarazione; inviano, pertanto più dichiarazioni che differiscono tra loro non solo per l'identificativo telematico, ma anche per i dati contenuti. Un esempio: l'Inps nel 2005 ha inviato 1.471 dichiarazioni.
- 3) Dichiarazione correttiva nei termini. Siamo in questo caso quando un sostituto d'imposta, prima della scadenza del termine di presentazione, intende rettificare o integrare una dichiarazione già presentata.
In questo caso egli può compilare e presentare una nuova dichiarazione completa di tutte le sue parti (correttiva totale), oppure una nuova dichiarazione contenente le sole comunicazioni oggetto di integrazione o rettifica, complete in ogni loro parte, indicando gli stessi numeri progressivi delle comunicazioni originarie (correttiva parziale).
- 4) Dichiarazione integrativa. Siamo in questo caso quando un sostituto d'imposta, scaduti i termini di presentazione, intende rettificare o integrare una dichiarazione già presentata.
In questo caso, come nel caso precedente, il sostituto può compilare e presentare una nuova dichiarazione (integrativa totale) completa di tutte le sue parti, oppure una nuova dichiarazione contenente le sole comunicazioni oggetto di integrazione o rettifica, complete in ogni loro parte, indicando gli stessi numeri progressivi delle comunicazioni originarie (integrativa parziale).

Caso 1. Se due o più dichiarazioni sono identiche a meno dell'identificativo telematico, basta considerarne una sola, eliminando le altre.

Caso 2. Se un sostituto d'imposta spezza una dichiarazione in più dichiarazioni diverse non soltanto per l'identificativo telematico, ma anche per i dati presenti in ciascuna di esse, possiamo correttamente pensare di essere in presenza di dichiarazioni differenti.

Nei casi 3 e 4 la stessa azienda ha inviato per gli stessi dipendenti più dichiarazioni differenti sia per il campo identificativo telematico, sia per i dati contenuti. Si tratta allora di stabilire se basta considerare valida una sola dichiarazione, scartando tutte le altre, oppure se occorre costruire una nuova dichiarazione "fittizia", combinando i dati parziali presenti nelle varie dichiarazioni.

Teoricamente queste situazioni possono essere risolte considerando i flag “dichiarazione correttiva” e “dichiarazione integrativa” presenti sulla tabella Frontespizio, che devono essere impostati nel caso di dichiarazioni correttive e integrative totali e parziali, e il flag “tipo operazione” presente sulle tabelle Lavoro dipendente e Lavoro autonomo, che deve essere impostato in caso di comunicazioni correttive e integrative parziali e indica se la comunicazione è nuova, elimina o sostituisce la precedente.

Ad oggi, però, Sogei non garantisce la qualità di questi flag e in Istat non è stato effettuato alcuno studio per avere una misura della loro affidabilità, né è stata implementata una procedura automatica per il trattamento delle dichiarazioni multiple.

Nella tabella 8 è rappresentata la numerosità dei flag dichiarazione, correttiva, integrativa presenti sulla tabella Frontespizio.

Tabella 8. Numerosità flag correttiva, integrativa sulla tabella Frontespizio

Anno di dichiarazione	Dichiarazioni correttive	Dichiarazioni integrative	Totale
2004	32.595	15.751	48.346
2005	33.995	16.539	50.534

Nella tabella 9 è rappresentata la numerosità del flag tipo operazione sulle tabelle Lavoro Dipendente e Lavoro Autonomo.

Tabella 9. Numerosità flag tipo operazione sulle tabelle Lavoro dipendente e autonomo

Anno di dichiarazione	Tipo operazione	
	Lavoro dipendente	Lavoro autonomo
2004	5.612	2.935
2005	34.074	2.451

Bibliografia.

- Consolini P. De Carli R.: “Trattamenti monetari non pensionistici: unità di analisi, fonti e rappresentazione statistica dei dati”. Collana Contributi Istat- 2002.
- Spinelli Vincenzo: “Processo di Acquisizione e trattamento Informatico degli Archivi relativi al Modello di Dichiarazione 770”. Collana Documenti Istat – 2007.
- www.agenziaentrate.it per modelli cartacei, istruzioni di compilazione, tracciati record e specifiche tecniche.
- www.gnu.org per l’ambiente di sviluppo GNU.
- www.linux.org per il sistema operativo Linux.

Contributi ISTAT(*)

- 1/2002 - Francesca Biancani, Andrea Carone, Rita Pistacchio e Giuseppina Ruocco - *Analisi delle imprese individuali*
- 2/2002 - Massimiliano Borgese - *Proposte metodologiche per un progetto d'indagine sul trasporto aereo alla luce della recente normativa comunitaria sul settore*
- 3/2002 - Nadia Di Veroli e Roberta Rizzi - *Proposta di classificazione dei rapporti di lavoro subordinato e delle attività di lavoro autonomo: analisi del quadro normativo*
- 4/2002 - Roberto Gismondi - *Uno stimatore ottimale in presenza di non risposte*
- 5/2002 - Maria Anna Pennucci - *Le strategie europee per l'occupazione dal Libro bianco di Delors al Consiglio Europeo di Cardiff*
- 1/2003 - Giovanni Maria Merola - *Safety Rules in Statistical Disclosure Control for Tabular Data*
- 2/2003 - Fabio Bacchini, Pietro Gennari e Roberto Iannaccone - *A new index of production for the construction sector based on input data*
- 3/2003 - Fulvia Ceroni e Enrica Morganti - *La metodologia e il potenziale informativo dell'archivio sui gruppi di impresa: primi risultati*
- 4/2003 - Sara Mastrovita e Isabella Siciliani - *Effetti dei trasferimenti sociali sulla distribuzione del reddito nei Paesi dell'Unione europea: un'analisi dal Panel europeo sulle famiglie*
- 5/2003 - Patrizia Cella, Giuseppe Garofalo, Adriano Paggiaro, Nicola Torelli e Caterina Viviano - *Demografia d'impresa: l'utilizzo di tecniche di abbinamento per l'analisi della continuità*
- 6/2003 - Enrico Grande e Orietta Luzi - *Metodologie per l'imputazione delle mancate risposte parziali: analisi critica e soluzioni disponibili in Istat*
- 7/2003 - Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino - *Indagine sperimentale sui posti di lavoro vacanti*
- 8/2003 - Mario Adua - *L'agricoltura di montagna: le aziende delle donne, caratteristiche agricole e socio-rurali*
- 9/2003 - Franco Mostacci e Roberto Sabbatini - *L'euro ha creato inflazione? Changeover e arrotondamenti dei prezzi al consumo in Italia nel 2002*
- 10/2003 - Leonello Tronti - *Problemi e prospettive di riforma del sistema pensionistico*
- 11/2003 - Roberto Gismondi - *Tecniche di stima e condizioni di coerenza per indagini infraannuali ripetute nel tempo*
- 12/2003 - Antonio Frenda - *Analisi delle legislazioni e delle prassi contabili relative ai gruppi di imprese nei paesi dell'Unione Europea*
- 1/2004 - Marcello D'Orazio, Marco Di Zio e Mauro Scanu - *Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints*
- 2/2004 - Giovanna Brancato - *Metodologie e stime dell'errore di risposta. Una sperimentazione di reintervista telefonica*
- 3/2004 - Franco Mostacci, Giuseppina Natale e Elisabetta Pugliese - *Gli indici dei prezzi al consumo per sub popolazioni*
- 4/2004 - Leonello Tronti - *Una proposta di metodo: osservazioni e raccomandazioni sulla definizione e la classificazione di alcune variabili attinenti al mercato del lavoro*
- 5/2004 - Ugo Guarnera - *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi: il software Quis*
- 6/2004 - Patrizia Giaquinto, Marco Landriscina e Daniela Pagliuca - *La nuova funzione di analisi dei modelli implementata in Genesees v. 3.0*
- 7/2004 - Roberto Di Giuseppe, Patrizia Giaquinto e Daniela Pagliuca - *MAUSS (Multivariate Allocation of Units in Sampling Surveys): un software generalizzato per risolvere il problema dell'allocazione campionaria nelle indagini Istat*
- 8/2004 - Ennio Fortunato e Liana Verzicco - *Problemi di rilevazione e integrazione della condizione professionale nelle indagini sociali dell'Istat*
- 9/2004 - Claudio Pauselli e Claudia Rinaldelli - *La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*
- 10/2004 - Eugenio Arcidiacono, Marina Briolini, Paolo Giuberti, Marco Ricci, Giovanni Sacchini e Giorgia Telloli - *Procedimenti giudiziari, reati, indagati e vittime in Emilia-Romagna nel 2002: un'analisi territoriale sulla base dei procedimenti iscritti nel sistema informativo Re.Ge.*
- 11/2004 - Enrico Grande e Orietta Luzi - *Regression trees in the context of imputation of item non-response: an experimental application on business data*
- 12/2004 - Luisa Frova e Marilena Pappagallo - *Procedura di now-cast dei dati di mortalità per causa*
- 13/2004 - Giorgio DellaRocca, Marco Di Zio, Orietta Luzi, Emanuela Scavalli e Giorgia Simeoni - *IDEA (Indices for Data Editing Assessment): sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI*
- 14/2004 - Monica Pace, Silvia Bruzzone, Luisa Frova e Marilena Pappagallo - *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe*
- 15/2004 - Elisa Berntsen - *Modello Unico di Dichiarazione ambientale: una fonte amministrativa per l'Archivio delle Unità Locali di Asia*
- 16/2004 - Salvatore F. Allegra e Alessandro La Rocca - *Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la definizione di un indice composto*
- 17/2004 - Francesca R. Pogelli - *Un'applicazione del modello "Country Product Dummy" per un'analisi territoriale dei prezzi*
- 18/2004 - Antonia Manzari - *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*
- 19/2004 - Claudio Pauselli - *Intensità di povertà relativa: stima dell'errore di campionamento e sua valutazione temporale*
- 20/2004 - Maria Dimitri, Ersilia Di Pietro, Alessandra Nuccitelli e Evelina Paluzzi - *Sperimentazione di una metodologia per il controllo della qualità di dati anagrafici*
- 21/2004 - Tiziana Pichiorri, Anna M. Sgamba e Valerio Papale - *Un modello di ottimizzazione per l'imputazione delle mancate risposte statistiche nell'indagine sui trasporti marittimi dell'Istat*

- 22/2004 – Diego Bellisai, Piero D. Falorsi, Annalisa Lucarelli, Maria A. Pennucci e Leonello G. Tronti – *Indagine pilota sulle retribuzioni di fatto nel pubblico impiego*
- 23/2004 – Lidia Brondi – *La riorganizzazione del sistema idrico: quadro normativo, delimitazione degli ambiti territoriali ottimali e analisi statistica delle loro caratteristiche strutturali*
- 24/2004 – Roberto Gismondi e Laura De Sandro – *Provisional Estimation of the Italian Monthly Retail Trade Index*
- 25/2004 – Annamaria Urbano, Claudia Brunini e Alessandra Chessa – *I minori in stato di abbandono: analisi del fenomeno e studio di una nuova prospettiva d'indagine*
- 26/2004 – Paola Anzini e Anna Ciammola – *La destagionalizzazione degli indici della produzione industriale: un confronto tra approccio diretto e indiretto*
- 27/2004 – Alessandro La Rocca – *Analisi della struttura settoriale dell'occupazione regionale: 8° Censimento dell'industria e dei servizi 2001 7° Censimento dell'industria e dei servizi 1991*
- 28/2004 – Vincenzo Spinelli e Massimiliano Tancioni – *I Trattamenti Monetari non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10*
- 29/2004 – Paolo Consolini – *L'indagine sperimentale sull'archivio fiscale modd.770 anno 1999: analisi della qualità del dato e stime campionarie*
- 1/2005 – Fabrizio M. Arosio – *La stampa periodica e l'informazione on-line: risultati dell'indagine pilota sui quotidiani on-line*
- 2/2005 – Marco Di Zio, Ugo Guarnera e Orietta Luzi – *Improving the effectiveness of a probabilistic editing strategy for business data*
- 3/2005 – Diego Moretti e Claudia Rinaldelli – *EU-SILC complex indicators: the implementation of variance estimation*
- 4/2005 – Fabio Bacchini, Roberto Iannaccone e Edoardo Otranto – *L'imputazione delle mancate risposte in presenza di dati longitudinali: un'applicazione ai permessi di costruzione*
- 5/2005 – Marco Broccoli – *Analisi della criminalità a livello comunale: metodologie innovative*
- 6/2005 – Claudia De Vitiis, Loredana Di Consiglio e Stefano Falorsi – *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*
- 7/2005 – Edoardo Otranto e Roberto Iannaccone – *Continuous Time Models to Extract a Signal in Presence of Irregular Surveys*
- 8/2005 – Cosima Mero e Adriano Pareto – *Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie*
- 9/2005 – Filippo Oropallo – *Enterprise microsimulation models and data challenges*
- 10/2005 – Marcello D' Orazio, Marco Di Zio e Mauro Scanu – *A comparison among different estimators of regression parameters on statistically matched files through an extensive simulation study*
- 11/2005 – Stefania Macchia, Manuela Murgia, Loredana Mazza, Giorgia Simeoni, Francesca Di Patrizio, Valentino Parisi, Roberto Petrillo e Paola Ungaro – *Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI*
- 12/2005 – Piero D. Falorsi, Monica Scannapieco, Antonia Boggia e Antonio Pavone – *Principi Guida per il Miglioramento della Qualità dei Dati Toponomastici nella Pubblica Amministrazione*
- 13/2005 – Ciro Baldi, Francesca Ceccato, Silvia Pacini e Donatella Tuzi – *La stima anticipata OROS sull'occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento*
- 14/2005 – Stefano De Francisci, Giuseppe Sindoni e Leonardo Tininini – *Da Winci/MD: un sistema per data warehouse statistici sul Web*
- 15/2005 – Gerardo Gallo e Evelina Palazzi – *I cittadini italiani naturalizzati: l'analisi dei dati censuari del 2001, con un confronto tra immigrati di prima e seconda generazione*
- 16/2005 – Saverio Gazzelloni, Mario Albisinni, Lorenzo Bagatta, Claudio Ceccarelli, Luciana Quattrociochi, Rita Ranaldi e Antonio Toma – *La nuova rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*
- 17/2005 – Maria Carla Congia – *Il lavoro degli extracomunitari nelle imprese italiane e la regolarizzazione del 2002. Prime evidenze empiriche dai dati INPS*
- 18/2005 – Giovanni Bottazzi, Patrizia Cella, Giuseppe Garofalo, Paolo Misso, Mariano Porcu e Marianna Tosi – *Indagine pilota sulla nuova imprenditorialità nella Regione Sardegna. Relazione Conclusiva*
- 19/2005 – Fabrizio Martire e Donatella Zindato – *Le famiglie straniere: analisi dei dati censuari del 2001 sui cittadini stranieri residenti*
- 20/2005 – Ennio Fortunato – *Il Sistema di Indicatori Territoriali: percorso di progetto, prospettive di sviluppo e integrazione con i processi di produzione statistica*
- 21/2005 – Antonella Baldassarini e Danilo Birardi – *I conti economici trimestrali: un approccio alla stima dell'input di lavoro*
- 22/2005 – Francesco Rizzo, Dario Camol e Laura Vignola – *Uso di XML e WEB Services per l'integrazione di sistemi informativi statistici attraverso lo standard SDMX*
- 1/2006 – Ennio Fortunato – *L'analisi integrata delle esigenze informative dell'utenza Istat: Il contributo del Sistema di Indicatori Territoriali*
- 2/2006 – Francesco Altarocca – *I design pattern nella progettazione di software per il supporto alla statistica ufficiale*
- 3/2006 – Roberta Palmieri – *Le migranti straniere: una lettura di genere dei dati dell'osservatorio interistituzionale sull'immigrazione in provincia di Macerata*
- 4/2006 – Raffaella Amato, Silvia Bruzzone, Valentina Delmonte e Lidia Fagiolo – *Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage*
- 5/2006 – Alessandro La Rocca – *Fuzzy clustering: la logica, i metodi*
- 6/2006 – Raffaella Cascioli – *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province*
- 7/2006 – Gianluca Brogi, Salvatore Cusimano, Giuseppina del Vicario, Giuseppe Garofalo e Orietta Patacchia – *La realizzazione di Asia Agricoltura tramite l'utilizzo di dati amministrativi: il contenuto delle fonti e i risultati del processo di integrazione*
- 8/2006 – Simonetta Cozzi – *La distribuzione commerciale in Italia: caratteristiche strutturali e tendenze evolutive*
- 9/2006 – Giovanni Seri – *A graphical framework to evaluate risk assessment and information loss at individual level*

- 10/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Feasibility studies for the coverage of public institutions in sections N and O*
- 11/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Quarterly labour cost index in public education*
- 12/2006 – Silvia Montagna, Patrizia Collesi, Florinda Damiani, Danila Fulgenzio, Maria Francesca Loporcaro e Giorgia Simeoni – *Nuove esperienze di rilevazione della Customer Satisfaction*
- 13/2006 – Lucia Coppola e Giovanni Seri – *Confidentiality aspects of household panel surveys: the case study of Italian sample from EU-SILC*
- 14/2006 – Lidia Brondi – *L'utilizzazione delle surveys per la stima del valore monetario del danno ambientale: il metodo della valutazione contingente*
- 15/2006 – Carlo Boselli – *Le piccole imprese leggere esportatrici e non esportatrici: differenze di struttura e di comportamento*
- 16/2006 – Carlo De Gregorio – *Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione*
- 1/2007 – Paolo Roberti, Maria Grazia Calza, Filippo Oropallo e Stefania Rossetti – *Knowledge Databases to Support Policy Impact Analysis: the EuroKy-PIA Project*
- 2/2007 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, e Marina Sorrentino – *Production of job vacancy statistics: coverage*
- 3/2007 – Carlo Lucarelli e Giampiero Ricci – *Working times and working schedules: the framework emerging from the new Italian lfs in a gender perspective*
- 4/2007 – Monica Scannapieco, Diego Zardetto e Giulio Barcaroli – *La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESSEES/SAS*
- 5/2007 – Giulio Barcaroli, Tiziana Pellicciotti – *Strumenti per la documentazione e diffusione dei microdati d'indagine: il Microdata Management Toolkit*

Documenti ISTAT(*)

- 1/2002 – Paolo Consolini e Rita De Carli - *Le prestazioni sociali monetarie non pensionistiche: unità di analisi, fonti e rappresentazione statistica dei dati*
- 2/2002 – Stefania Macchia - *Sperimentazione, implementazione e gestione dell'ambiente di codifica automatica della classificazione delle Attività economiche*
- 3/2002 – Maria De Lucia - *Applicabilità della disciplina in materia di festività nel pubblico impiego*
- 4/2002 – Roberto Gismondi, Massimo Marciani e Mauro Giorgetti - *The italian contribution towards the implementation of an european transport information system: main results of the MESUDEMO project*
- 5/2002 – Olimpio Cianfarani e Sauro Angeletti - *Misure di risultato e indicatori di processo: l'esperienza progettuale dell'Istat*
- 6/2002 – Riccardo Carbinì e Valerio De Santis – *Programma statistico nazionale: specifiche e note metodologiche per la compilazione delle schede identificative dei progetti*
- 7/2002 – Maria De Lucia – *Il CCNL del personale dirigente dell'area 1 e la valutazione delle prestazioni dei dirigenti*
- 8/2002 – Giuseppe Garofalo e Enrica Morganti – *Gruppo di lavoro per la progettazione di un archivio statistico sui gruppi d'impresa*
- 1/2003 – Francesca Ceccato, Massimiliano Tancioni e Donatella Tuzi – *MODSIM-P: Il nuovo modello dinamico di previsione della spesa pensionistica*
- 2/2003 – Anna Pia Mirto – *Definizioni e classificazioni delle strutture ricettive nelle rilevazioni statistiche ufficiali sull'offerta turistica*
- 3/2003 – Simona Spirito – *Le prestazioni assistenziali monetarie non pensionistiche*
- 4/2003 – Maria De Lucia – *Approfondimenti di alcune tematiche inerenti la gestione del personale*
- 5/2003 – Rosalia Coniglio, Marialuisa Cugno, Maria Filmeno e Alberto Vitalini – *Mappatura della criminalità nel distretto di Milano*
- 6/2003 – Maria Letizia D'Autilia – *I provvedimenti di riforma della pubblica amministrazione per l'identificazione delle "Amministrazioni pubbliche" secondo il Sec95: analisi istituzionale e organizzativa per l'anno 2000*
- 7/2003 – Francesca Gallo, Pierpaolo Massoli, Sara Mastrovita, Roberto Merluzzi, Claudio Pauselli, Isabella Siciliani e Alessandra Sorrentino – *La procedura di controllo e correzione dei dati Panel Europeo sulle famiglie*
- 8/2003 – Cinzia Castagnaro, Martina Lo Conte, Stefania Macchia e Manuela Murgia – *Una soluzione in-house per le indagini CATI: il caso della Indagine Campionaria sulle Nascite*
- 9/2003 – Anna Pia Maria Mirto e Norina Salamone – *La classificazione delle strutture ricettive turistiche nella normativa delle regioni italiane*
- 10/2003 – Roberto Gismondi e Anna Pia Maria Mirto – *Le fonti statistiche per l'analisi della congiuntura turistica: il mosaico italiano*
- 11/2003 – Loredana Di Consiglio e Stefano Falorsi – *Alcuni aspetti metodologici relativi al disegno dell'indagine di copertura del Censimento Generale della Popolazione 2001*
- 12/2003 – Roberto Gismondi e Anna Rita Giorgi – *Struttura e dinamica evolutiva del comparto commerciale al dettaglio: le tendenze recenti e gli effetti della riforma "Bersani"*
- 13/2003 – Donatella Cangialosi e Rosario Milazzo – *Fabbisogni formativi degli Uffici comunali di statistica: indagine rapida in Sicilia*
- 14/2003 – Agostino Buratti e Giovanni Salzano – *Il sistema automatizzato integrato per la gestione delle rilevazioni dei documenti di bilancio degli enti locali*
- 1/2004 – Giovanna Brancato e Giorgia Simeoni – *Tesori del Sistema Informativo di Documentazione delle Indagini (SIDI)*
- 2/2004 – Corrado Peperoni – *Indagine sui bilanci consuntivi degli Enti previdenziali: rilevazione, gestione e procedure di controllo dei dati*
- 3/2004 – Marzia Angelucci, Giovanna Brancato, Dario Camol, Alessio Cardacino, Sandra Maresca e Concetta Pellegrini – *Il sistema ASIMET per la gestione delle Note Metodologiche dell'Annuario Statistico Italiano*
- 4/2004 – Francesca Gallo, Sara Mastrovita, Isabella Siciliani e Giovanni Battista Arcieri – *Il processo di produzione dell'Indagine ECHP*
- 5/2004 – Natale Renato Fazio e Carmela Pascucci – *Gli operatori non identificati nelle statistiche del commercio con l'estero: metodologia di identificazione nelle spedizioni "groupage" e miglioramento nella qualità dei dati*
- 6/2004 – Diego Moretti e Claudia Rinaldelli – *Una valutazione dettagliata dell'errore campionario della spesa media mensile familiare*
- 7/2004 – Franco Mostacci – *Aspetti Teorico-pratici per la Costruzione di Indici dei Prezzi al Consumo*
- 8/2004 – Maria Frustaci – *Glossario economico-statistico multilingua*
- 9/2004 – Giovanni Seri e Maurizio Lucarelli – *"Il Laboratorio per l'analisi dei dati elementari (ADELE): monitoraggio dell'attività dal 1999 al 2004"*
- 10/2004 – Alessandra Nuccitelli, Francesco Bosio e Luciano Fioriti – *L'applicazione RECLINK per il record linkage: metodologia implementata e linee guida per la sua utilizzazione*
- 1/2005 – Francesco Cuccia, Simone De Angelis, Antonio Laureti Palma, Stefania Macchia, Simona Mastroluca e Domenico Perrone – *La codifica delle variabili testuali nel 14° Censimento Generale della Popolazione*
- 2/2005 – Marina Peci – *La statistica per i Comuni: sviluppo e prospettive del progetto Sisco.T (Servizio Informativo Statistico Comunale. Tavole)*
- 3/2005 – Massimiliano Renzetti e Annamaria Urbano – *Sistema Informativo sulla Giustizia: strumenti di gestione e manutenzione*
- 4/2005 – Marco Broccoli, Roberto Di Giuseppe e Daniela Pagliuca – *Progettazione di una procedura informatica generalizzata per la sperimentazione del metodo Microstrat di coordinamento della selezione delle imprese soggette a rilevazioni nella realtà Istat*
- 5/2005 – Mauro Albani e Francesca Pagliara – *La ristrutturazione della rilevazione Istat sulla criminalità minorile*
- 6/2005 – Francesco Altarocca e Gaetano Sberno – *Progettazione e sviluppo di un "Catalogo dei File Grezzi con meta-dati di base" (CFG) in tecnologia Web*

- 7/2005 – Salvatore F. Allegra e Barbara Baldazzi – *Data editing and quality of daily diaries in the Italian Time Use Survey*
- 8/2005 – Alessandra Capobianchi – *Alcune esperienze in ambito internazionale per l'accesso ai dati elementari*
- 9/2005 – Francesco Rizzo, Laura Vignola, Dario Camol e Mauro Bianchi – *Il progetto "banca dati della diffusione congiunturale"*
- 10/2005 – Ennio Fortunato e Nadia Mignolli – *I sistemi informativi Istat per la diffusione via web*
- 11/2005 – Ennio Fortunato e Nadia Mignolli – *Sistemi di indicatori per l'attività di governo: l'offerta informativa dell'Istat*
- 12/2005 – Carlo De Gregorio e Stefania Fatello – *L'indice dei prezzi al consumo dei testi scolastici nel 2004*
- 13/2005 – Francesco Rizzo e Laura Vignola – *RSS: uno standard per diffondere informazioni*
- 14/2005 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, Annalisa Lucarelli e Marina Sorrentino – *Launching and implementing the job vacancy statistics*
- 15/2005 – Stefano De Francisci, Massimiliano Renzetti, Giuseppe Sindoni e Leonardo Tininini – *La modellazione dei processi nel Sistema Informativo Generalizzato di Diffusione dell'ISTAT*
- 16/2005 – Ennio Fortunato e Nadia Mignolli – *Verso il Sistema di Indicatori Territoriali: rilevazione e analisi della produzione Istat*
- 17/2005 – Raffaella Cianchetta e Daniela Pagliuca – *Soluzioni Open Source per il software generalizzato in Istat: il caso di PHPSurveyor*
- 18/2005 – Gianluca Giuliani e Barbara Boschetto – *Gli indicatori di qualità dell'Indagine continua sulle Forze di Lavoro dell'Istat*
- 19/2005 – Rossana Balestrino, Franco Garritano, Carlo Cipriano e Luciano Fanfoni – *Metodi e aspetti tecnologici di raccolta dei dati sulle imprese*
- 1/2006 – Roberta Roncati – www.istat.it (versione 3.0) *Il nuovo piano di navigazione*
- 2/2006 – Maura Seri e Annamaria Urbano – *Sistema Informativo Territoriale sulla Giustizia: la sezione sui confronti internazionali*
- 3/2006 – Giovanna Brancato, Riccardo Carbini e Concetta Pellegrini – *SIQual: il sistema informativo sulla qualità per gli utenti esterni*
- 4/2006 – Concetta Pellegrini – *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza SIDI*
- 5/2006 – Maurizio Lucarelli – *Una valutazione critica dei modelli di accesso remoto nella comunicazione di informazione statistica*
- 6/2006 – Natale Renato Fazio – *La ricostruzione storica delle statistiche del commercio con l'estero per gli anni 1970-1990*
- 7/2006 – Emilia D'Acunto – *L'evoluzione delle statistiche ufficiali sugli indici dei prezzi al consumo*
- 8/2006 – Ugo Guarnera, Orietta Luzi e Stefano Salvi – *Indagine struttura e produzioni delle aziende agricole: la nuova procedura di controllo e correzione automatica per le variabili su superfici aziendali e consistenza degli allevamenti*
- 9/2006 – Maurizio Lucarelli – *La regionalizzazione del Laboratorio ADELE: un'ipotesi di sistema distribuito per l'accesso ai dati elementari*
- 10/2006 – Alessandra Bugio, Claudia De Vitiis, Stefano Falorsi, Lidia Gargiulo, Emilio Gianicolo e Alessandro Pallara – *La stima di indicatori per domini sub-regionali con i dati dell'indagine: condizioni di salute e ricorso ai servizi sanitari*
- 11/2006 – Sonia Vittozzi, Paola Giacchè, Achille Zuchegna, Piero Crivelli, Patrizia Collesi, Valerio Tiberi, Alexia Sasso, Maurizio Bonsignori, Giuseppe Stassi e Giovanni A. Barbieri – *Progetto di articolazione della produzione editoriale in collane e settori*
- 12/2006 – Alessandra Coli, Francesca Tartamella, Giuseppe Sacco, Ivan Faiella, Marcello D'Orazio, Marco Di Zio, Mauro Scanu, Isabella Siciliani, Sara Colombini e Alessandra Masi – *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'Indagine Banca d'Italia sui bilanci delle famiglie italiane*
- 13/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Intrastat*
- 14/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Extrastat*
- 15/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: comparazione tra rilevazione Intrastat ed Extrastat*
- 16/2006 – Fabio M. Rapiti – *Short term statistics quality Reporting: the LCI National Quality Report 2004*
- 17/2006 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Valutazione delle possibilità di uso di dati fiscali a supporto della rilevazione PMI*
- 18/2006 – Mauro Albani – *La nuova procedura per il trattamento dei dati dell'indagine Istat sulla criminalità*
- 19/2006 – Alessandra Capobianchi – *Review dei sistemi di accesso remoto: schematizzazione e analisi comparativa*
- 20/2006 – Francesco Altarocca – *Gli strumenti informatici nella raccolta dei dati di indagini statistiche: il caso della Rilevazione sperimentale delle tecnologie informatiche e della comunicazione nelle Pubbliche Amministrazioni locali*
- 1/2007 – Giuseppe Stassi – *La politica editoriale dell'Istat nel periodo 1996-2004: collane, settori, modalità di diffusione*
- 2/2007 – Daniela Ichim – *Microdata anonymisation of the Community Innovation Survey data: a density based clustering approach for risk assessment*
- 3/2007 – Ugo Guarnera, Orietta Luzi e Irene Tommasi – *La nuova procedura di controllo e correzione degli errori e delle mancate risposte parziali nell'indagine sui Risultati Economici delle Aziende Agricole (REA)*
- 4/2007 – Vincenzo Spinelli – *Processo di Acquisizione e Trattamento Informativo degli Archivi relativi al Modello di Dichiarazione 770*
- 5/2007 – Anna Di Carlo, Maria Picci, Laura Posta, Michaela Raffone, Giuseppe Stassi e Fiorella Tortora – *La progettazione dei Censimenti generali 2010-2011: I - Analisi, valutazione e proposte in merito ad atti di normazione e finanziamento*
- 6/2007 – Silvia Bruzzone, Antonia Manzari, Marilena Pappagallo e Alessandra Reale – *Indagine sulle Cause di Morte: Nuova procedura automatica per il controllo e la correzione delle variabili demo-sociali*
- 7/2007 – Maura Giacommo, Carlo Vaccari e Monica Scannapieco – *Indagine sulle Scelte Tecnologiche degli Istituti Nazionali di Statistica*
- 8/2007 – Lamberto Pizzicannella – *Sviluppo del processo di acquisizione e trattamento informatico degli archivi relativi al modello di dichiarazione 770. Anni 2004 - 2005*