

n. 5/2007

**Strumenti per la documentazione e diffusione dei  
microdati d'indagine: il Microdata Management  
Toolkit**

*G. Barcaroli, T. Pellicciotti*

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

**n. 5/2007**

**Strumenti per la documentazione e diffusione dei  
microdati d'indagine: il Microdata Management  
Toolkit**

*G. Barcaroli(\*), T. Pellicciotti(\*\*)*

(\*) ISTAT - Servizio Metodologie, tecnologie e software per la produzione dell'informazione statistica  
(\*\*) ISTAT - Ufficio delle relazioni internazionali e della cooperazione internazionale

**Contributi e Documenti Istat 2007**

Istituto Nazionale di Statistica  
Servizio Produzione Editoriale

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

## Introduzione

Ogni rilevazione, sia essa di carattere censuario che campionario, condotta dall'Istituto Nazionale di Statistica, dà luogo ad una collezione di dati elementari che rappresenta il massimo della informazione disponibile relativamente ai fenomeni per il cui studio la rilevazione viene condotta.

E' chiaro che per un utente finale "esperto", in grado cioè di effettuare elaborazioni in proprio sui dati, la soluzione ideale è quella di disporre direttamente di tali dati, anziché usufruire delle aggregazioni e delle analisi standard messe a disposizione dall'Istituto attraverso i canali tradizionali delle pubblicazioni cartacee, o, più avanzati, costituiti dai *data warehouse* statistici, che, pensati per la collettività di utenti nel suo complesso, potrebbero non soddisfare completamente il fabbisogno informativo dei singoli.

Tenendo conto di questa fascia di utenza, una volta garantito il rispetto del vincolo della riservatezza dei dati, l'ISTAT offre collezioni campionarie di dati elementari relative ad alcune indagini. Tra queste, citiamo il Censimento della Popolazione (1991 e 2001), le rilevazioni trimestrali e continue sulle Forze Lavoro, l'indagine sui Consumi delle Famiglie, quella sulle Condizioni di vita (EU-SILC), le indagini Multiscopo sulle Famiglie, la campionaria sulle Nascite, sull'Inserimento Professionale dei Laureati e sui Percorsi di Studio e di Lavoro dei Diplomatici.

Attualmente, tali collezioni campionarie anonimizzate, note come "File di microdati", possono essere rilasciate per fini di studio e di ricerca su richiesta motivata e previa autorizzazione del Presidente dell'Istituto. L'utente interessato sottometta la propria richiesta relativamente ad uno o più di tali file, ed ottiene un CD contenente il dataset, unitamente ad una documentazione (documenti word o pdf) relativa al processo di produzione del file dei microdati (disegno di campionamento, questionari utilizzati, raccolta e trattamento dei dati, costruzione delle stime, tecniche utilizzate per la riduzione del rischio di violazione della riservatezza), alla qualità dei dati (errori campionari), alle classificazioni utilizzate (codici e decodifiche), ed alla organizzazione logico-fisica dei dati (tracciato dei record).

Queste modalità di fornitura non sono però esenti da limiti. Ad esempio, di *completezza*: non sempre sono disponibili tutte le informazioni necessarie, non essendo stato definito uno standard in tal senso. Ma soprattutto di *operatività*: la documentazione necessaria per comprendere il contenuto dei file e poterlo elaborare è conservata in documenti a parte, non immediatamente processabili. E' quindi a carico dell'utente lo sviluppo di applicazioni che permettano di trattare i dati ricevuti, con conseguente scrittura in specifici programmi della definizione delle variabili assieme alle modalità ammissibili per ciascuna di esse.

Sono allo studio, da parte dell'Istituto, soluzioni di fornitura di microdati all'utenza esterna, alternative a quella attuale (ISTAT, 2006). In primo luogo verrebbero differenziate le tipologie di file da rendere disponibili: da una parte, *file ad uso pubblico*, disponibili a chiunque, e *file per la ricerca*, disponibili a seguito di presentazione di un progetto di ricerca e sottoscrizione di contratto. Entrambi verrebbero offerti sul sito web ufficiale dell'Istituto, "*con procedure differenziate a seconda dell'utenza: i file ad uso pubblico potrebbero essere scaricabili a seguito di una semplice procedura di iscrizione al sito, mentre i file per la ricerca potrebbero essere accessibili solo a fronte di un'autenticazione... a conclusione dell'iter di autorizzazione*".

Al di là della distinzione della tipologia dei file, per entrambi l'obiettivo è quello di renderli disponibili su web: non più dataset da fornire offline a richiesta, ma dati immediatamente visibili e ottenibili con tutta la necessaria documentazione, concettuale ed operativa, una volta ottemperate le formalità (differenziate) necessarie. In questo quadro sarebbe auspicabile che i dati, una volta scaricati, fossero immediatamente elaborabili, senza necessità di sviluppo ad hoc da parte dell'utente, mediante uno tra i pacchetti statistici maggiormente utilizzati nell'ambiente della ricerca.

Se questo è vero, è a questo punto di grande interesse la possibilità di utilizzare strumenti di tipo generalizzato in grado di garantire quanto sopra in modo ottimale e a costi implementativi minimali per l'Istituto.

In questo lavoro riportiamo l'esperienza effettuata nell'ambito del progetto di cooperazione con gli Istituti statistici della Bosnia Erzegovina nel triennio 2003-2005, per la conduzione dell'indagine sui

consumi (2004 BiH Household Budget Survey), nella parte riguardante l'utilizzo di un software generalizzato (Microdata Management Toolkit) per la costruzione del sito web contenente i microdati dell'indagine coi relativi metadati. Riteniamo che l'esperienza svolta possa fornire indicazioni utili anche per l'Istituto, relativamente alle innovazioni che si intendono introdurre in tale ambito.

Il Microdata Management Toolkit è stato sviluppato dalla World Bank Data Group per l'International Household Survey Network (IHSN), affinché fosse utilizzato in primo luogo dagli enti produttori di statistiche ufficiali, con l'obiettivo di mettere in condivisione con la comunità dei ricercatori i dati raccolti, rispondendo a standard internazionali di documentazione.

Il Toolkit si compone di tre diversi strumenti:

- il *Metadata Editor*, che permette ai responsabili di una data indagine di descrivere dati e metadati di interesse;
- il *CD-Builder*, che a partire da quanto ottenuto con il Metadata Editor, permette di costruire le pagine HTML (da inserire in CD-Rom o in pagine web) in cui rendere disponibili i dati e la relativa documentazione;
- il *Nesstar Explorer*, che consente di navigare tra dati e metadati e di scaricarli secondo i formati più convenienti (in formato testo o Excel, ma anche SAS, SPSS, Stata, Statistica, dBase, o anche in XML).

Nell'esperienza bosniaca, i dati raccolti sono stati dapprima elaborati al fine di trattare correttamente la mancata risposta parziale (con opportune procedure di localizzazione e imputazione degli errori e dei valori mancanti) e quella totale (con riponderazione dei dati). I dati così ottenuti sono stati analizzati mediante il software  $\mu$ -ARGUS per un'opportuna manipolazione al fine di contenere il rischio di violazione della riservatezza al di sotto di una prefissata soglia. A questo punto, sono stati utilizzati i vari strumenti del Microdata Management Toolkit per costruire pagine web in grado di consentire l'accesso ai microdati da parte dell'utenza interessata (BiH Statistical Institutes and ISTAT, 2005).

Il risultato finale è immediatamente visibile all'interno del sito ufficiale dell'Agenzia di Statistica della Bosnia Erzegovina (BHAS).<sup>1</sup>

Nel presente lavoro, dopo un richiamo agli standard internazionali di documentazione e ad una breve illustrazione delle caratteristiche delle diverse componenti del *toolkit*, esemplificheremo ciascuno dei passi di cui sopra, riportando, come già detto, l'esperienza vissuta con l'indagine sui Consumi in Bosnia Erzegovina, e mostrando come il carico di attività richiesto sia tutto sommato relativamente limitato, una volta appresi i principi fondamentali di funzionamento degli strumenti di cui si compone il Toolkit.

Il lavoro è così strutturato:

- nel paragrafo 1, vengono descritti gli standard internazionali di documentazione implementati all'interno del toolkit, e cioè il *Data Documentation Initiative* (DDI), ed il *Dublin Core Metadata Initiative* (DCMI), utilizzati rispettivamente per documentare i dati ed i processi necessari per ottenerli (DDI), e le risorse associate (questionari, programmi elaborativi, siti web, ecc.);
- nel paragrafo 2 viene data una descrizione dell'architettura del toolkit, e delle modalità di funzionamento di ciascuna componente (*Metadata Editor*, *CD-Builder*, *Nesstar Explorer*);
- nel paragrafo 3 viene riportato, come caso di studio, l'intero processo di costruzione e messa a disposizione del file dei microdati dell'indagine sui consumi effettuata in Bosnia Erzegovina nel 2004 (con l'assistenza tecnica prestata dall'ISTAT nel quadro della cooperazione internazionale): dalla preparazione del file di microdati "sicuro" (mediante utilizzo del software  $\mu$ -Argus), al caricamento dei dati e la definizione dei relativi metadati (mediante l'utilizzo del Metadata Editor), e la generazione delle pagine web che consentono la consultazione ed il download del file così ottenuto (mediante il CD-Builder);
- infine, nel paragrafo 4 vengono riportate delle considerazioni conclusive.

---

<sup>1</sup> [http://www.bhas.ba/Engleski/en\\_ankete.html](http://www.bhas.ba/Engleski/en_ankete.html)

## 1. Standard internazionali di documentazione dei microdati

Il Microdata Management Toolkit fa riferimento a due diversi modelli di per la documentazione dei microdati: il DDI (Data Documentation Initiative) ed il DCMI (Dublin Core Metadata Initiative), entrambi basati sull'utilizzo dell'Extensible Markup Language (XML).

### 1.1. Data Documentation Initiative (DDI)

Il modello per la specificazione dei metadati noto come Data Documentation Initiative è stato definito all'interno dell'Inter-university Consortium for Political and Social Research (ICPSR), che raggruppa 500 tra università e college. E' attualmente supportato da una "alleanza" di istituzioni nordamericane ed europee, tra cui molti tra i più grandi produttori di dati nel mondo<sup>2</sup>.

Il modello DDI è stato disegnato tenendo conto dei requisiti di documentazione di dati statistici generati da indagini campionarie e censuarie, archivi amministrativi, piani sperimentali (Blank et al, 2004). Le unità di analisi possono essere individui, famiglie, imprese, istituzioni, transazioni, nazioni, ed ogni altro oggetto di interesse scientifico, economico o sociale. Le osservazioni possono essere cross-sectional, oppure longitudinali, o di entrambi i tipi.

Mediante il modello DDI è anche possibile fornire documentazione relativa alle metodologie ed ai processi impiegati nelle varie fasi di indagine: strategia campionaria, tecniche di raccolta dei dati, procedure di trattamento, domini di studio, ecc.

I metadati definiti secondo il modello DDI sono organizzati in cinque diverse sezioni:

Sezione 1 - *Descrizione del Documento*: in questa sezione viene descritta l'organizzazione dei metadati dello studio

Sezione 2 - *Descrizione dello Studio* (indagine campionaria, censimento o altro): vengono fornite informazioni sulla/e istituzione/i responsabile/i, viene data una descrizione sintetica del contenuto dei dati, e vengono descritti i più importanti processi di raccolta, trattamento ed elaborazione dei dati.

Sezione 3 - *Descrizione dei Data File*: in questa sezione vengono riportate informazioni di carattere generale relative ad ogni dataset di interesse: contenuto, numero di variabili e numero di record, ecc.

Sezione 4 - *Descrizione delle Variabili*: per ogni variabile viene riportata informazione dettagliata relativa al significato, alla distribuzione, ad eventuali metodi di derivazione ed imputazione impiegati.

Sezione 5 - *Altro Materiale*: in questa sezione è possibile rendere disponibile informazione non strutturata, in forma di documenti (questionari, classificazioni, rapporti tecnici e risultati di analisi statistiche, manuali per i rilevatori, ecc.) o strumenti impiegati nel processo di produzione (programmi di elaborazione, mappe, ecc.).

### 1.2. Dublin Core Metadata Initiative (DCMI)

Gli standard di documentazione del modello DCMI prevedono due possibili livelli: il *Semplice* ed il *Qualificato*. Il Toolkit ne implementa solo la prima versione, quella semplice (DCMI, 2006).

Il DCMI ha come finalità quella di "promuovere l'adozione di standard interoperabili di metadati, e sviluppare vocabolari specializzati per la descrizione di risorse che rendano possibili sistemi informativi più intelligenti".

Tenendo conto il concetto di *risorsa* sopra definito, il DCMI nella sua versione "semplice" prevede la possibilità di organizzare i metadati secondo 15 differenti elementi:

1. *Titolo*: il nome della risorsa
2. *Soggetto*: argomento cui la risorsa è riferibile
3. *Descrizione*: Descrizione sintetica della risorsa

---

<sup>2</sup> Per informazioni più dettagliate: <http://www.ddialliance.org/DDI>

4. *Tipo*: qualificativo della risorsa (un questionario, un programma, una mappa, ecc.)
5. *Fonte*: riferimento alla risorsa (indirizzo web, identificativo di un documento PDF, ecc.)
6. *Relazione*: riferimento ad una risorsa correlata
7. *Copertura*: tipicamente, localizzazione spaziale (ad esempio, un paese) o un periodo temporale (una data o un intervallo)
8. *Creatore*: persona, organizzazione o servizio responsabile del contenuto della risorsa
9. *Editore*: persona, organizzazione o servizio responsabile della messa a disposizione della risorsa
10. *Contributore*: persona, organizzazione o servizio che ha contribuito al contenuto della risorsa
11. *Diritti*: eventuale copyright della risorsa
12. *Data*: riferimento temporale associato ad un evento nel ciclo di vita della risorsa (tipicamente, data di creazione o di disponibilità della risorsa)
13. *Formato*: qualificativo di componenti hardware o software necessari per dispiegare o rendere disponibile la risorsa
14. *Identificativo*: riferimento univoco alla risorsa all'interno di un dato contesto (ad esempio, URL o ISBN)
15. *Linguaggio*: linguaggio mediante il quale è espresso il contenuto della risorsa

### 1.3. Utilizzo di template di metadati all'interno del Toolkit

Il modello DDI contiene centinaia di elementi, disegnati per rispondere ai requisiti delle situazioni più diverse. Nei casi concreti, solo un sottoinsieme limitato di tali elementi si rende necessario per documentare uno studio o un particolare dataset.

Similmente, non tutti gli elementi del modello DCMI sono necessari per documentare delle risorse correlate ad uno studio.

La componente Metadata Editor del Microdata Management Toolkit mette a disposizione due differenti tipi di template per permettere all'utente di operare una selezione dei metadati di documentazione più rilevanti:

- *Template di Studio*: per identificare gli elementi del modello DDI da utilizzare per documentare uno studio (indagine, censimento) ed i relativi data files;
- *Template di Descrizione delle Risorse*: per specificare gli elementi del modello DCMI da utilizzare nella documentazione di risorse quali documenti, programmi di elaborazione, mappe o altro materiale.

Ad esempio, l'International Household Survey Network (IHSN)<sup>3</sup>, sponsorizzato dalla World Bank, ha definito uno Study Template ed un Resource Description Template particolarmente adatti a documentare indagini e dati su famiglie ed individui, da importare all'interno del Metadata Editor. Poiché il CD-Builder richiede obbligatoriamente alcuni elementi contenuti nei template IHSN, il loro utilizzo è altamente raccomandato.

Un template di studio ha tre livelli principali:

1. il primo livello è associato ad una specifica di descrizione DDI (descrizione del Documento, dello Studio, dei File, delle Variabili, di Altro Materiale);
2. il secondo livello permette di raggruppare le informazioni secondo Gruppi, non corrispondenti a standard DDI, ma convenientemente scelti per la particolare applicazione;
3. il terzo livello è riferito a elementi DDI di dettaglio.

Ad esempio, lo Study Template dell'IHSN prevede, per la descrizione dello studio (primo livello), i seguenti gruppi (secondo livello):

1. identificativo
2. versione

---

<sup>3</sup> Vedi il sito: <http://www.internationalsurveynetwork.org/home/>

3. overview
4. scopo
5. copertura
6. produttori e sponsor
7. campione
8. raccolta dati
9. elaborazione dati
10. validazione dati
11. accesso ai dati
12. copyright
13. contatti

Ci sembra interessante riportare il dettaglio di terzo livello relativo ai gruppi 7, 8, 9 e 10 relativi in modo specifico al processo di produzione dei dati validati:

Gruppo 7 (“campione”):

1. disegno del campione
2. deviazioni dal disegno teorico
3. tassi di risposta
4. ponderazione

Gruppo 8 (“raccolta dati”):

1. periodo di raccolta
2. periodo di riferimento
3. questionari
4. rilevatori
5. supervisori

Gruppo 9 (“elaborazione dati”):

1. data editing

Gruppo 10 (“validazione dati”):

2. stima degli errori campionari
3. altre forme di valutazione della qualità dei dati

## **2. Architettura e funzionamento del Toolkit**

L'architettura del Toolkit sviluppata dal Data Development Group della Banca Mondiale risponde all'obiettivo principale di superare gli ostacoli di natura tecnica che, nei paesi in via di sviluppo ma spesso anche per i produttori di statistica più avanzati, rallentano le fasi conclusive della lavorazione nonché l'ottimizzazione e la massimizzazione dello sfruttamento dei dati di un'indagine e il loro ritorno ad un pubblico più vasto dei soli produttori e finanziatori. Esso infatti permette di organizzare e conservare tutta la documentazione relativa ad un'indagine di carattere sia campionario che censuario mantenendo dati, microdati e relativi metadati in un unico file contenitore; è di facile utilizzo in tutte le sue componenti e i processi che portano alla generazione del prodotto finale sono in larga parte automatizzati (Dupriez, 2006).

Inoltre, l'adesione e conformità della sua struttura ai principali standard internazionali, quali quelli definiti nella DDI e il Dublin Core promuovono la diffusione e l'adozione degli stessi su scala sempre più vasta.

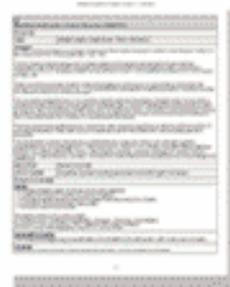
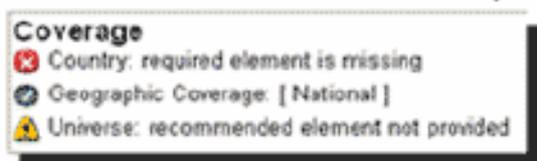
Come già accennato nell'introduzione, il Toolkit si compone di due strumenti, il Metadata Editor e il CD Builder, di facile installazione ed entrambi operanti e compatibili con l'ambiente Windows, ed è corredato da una applicazione libera, Nesstar Explorer, un software per la lettura e navigazione dei dati che ne permette inoltre l'esportazione in vari formati standard quali SAS, SPSS, STATA etc.

- 1 With the **Metadata Editor**, import data from various standard format (SPSS, STATA, ASCII, others), and provide comprehensive metadata in user-friendly screens.



Data and metadata become one entity, saved in a single file. The free **Nesstar Explorer** program allows users to view metadata and re-export data to various common formats.

- 2 Generate various "Metadata Diagnostic" reports...



- 3 ... and automatically produce detailed survey documentation in PDF format.

- 4 With the **CD-ROM Builder**, produce a user-friendly html-output for sharing and preserving your data and metadata.



Fig. 1 – Il flusso delle attività descritto nel sito dell'International Household Survey Network

Il flusso delle attività (vedi fig.1) prevede l'utilizzo in una prima fase dell'applicazione Metadata Editor, che produce un file Nesstar; questo viene importato e rifinito nell'applicazione CD Builder, che genera l'output finale, cioè l'intera indagine completa di microdati e metadati relativi ad essa, in forma di CD, DVD o pagine web. Il contenuto generato può essere quindi letto e navigato per mezzo del programma di lettura Nesstar Explorer, fornito insieme a tutta la documentazione nel momento della creazione del CD finale. Di seguito si riporta descrizione del funzionamento delle singole applicazioni che compongono il Toolkit.

## 2.1. Metadata Editor

Metadata Editor è uno strumento editor di notevole facilità di approccio e di utilizzo che viene applicato per documentare set di microdati. E' l'unico degli strumenti del Toolkit che richiede una licenza<sup>4</sup>, poiché è costruito e sviluppato su un'applicazione commerciale.

Il processo di documentazione dell'indagine prende avvio dall'importazione di un file di dati (dataset), cioè uno *studio*, nei formati statistici standard (generalmente SPSS, STATA and STATISTICA; l'attuale versione del programma non permette l'importazione diretta di file SAS, pertanto set di dati in questo formato vanno convertiti in uno dei formati precedenti; ancora più onerosa e lunga al momento è la procedura per l'esportazione di file ASCII, poiché i file di dati in questo formato non contengono alcuni elementi invece richiesti dal programma, ad esempio, per i nomi delle variabili). Il programma lavora su un *progetto*, che corrisponde ad un'indagine, ad un censimento, o comunque ad un set di microdati, inclusi tutti i dati e metadati relativi ad uno *studio*. Le informazioni statistiche derivanti da un'indagine ed importate tramite l'Editor vengono infatti integrate da un ampio set di metadati: informazioni generali sull'indagine (nome, obiettivo, contesto socio-economico in cui prende avvio), produttori e finanziatori, informazioni di carattere metodologico quali ad esempio strategia e disegno di campionamento, modalità di raccolta dati, periodi di riferimento e di realizzazione, questionari utilizzati, metodi di valutazione, controllo e imputazione dei dati; documenti disponibili di corredo all'indagine, elenco e descrizione dei programmi informatici utilizzati e altri materiali di varia natura.

Rispetto ai numerosi elementi contenuti nelle specifiche del DDI e del Dublin Core, il Metadata Editor permette di selezionare gli elementi più rilevanti utilizzando delle strutture (*templates*) definite ed applicate per selezionare ed organizzare gli elementi richiesti e necessari, cui si accede dal menu *Documentation* dell'applicazione (vedi fig.2).

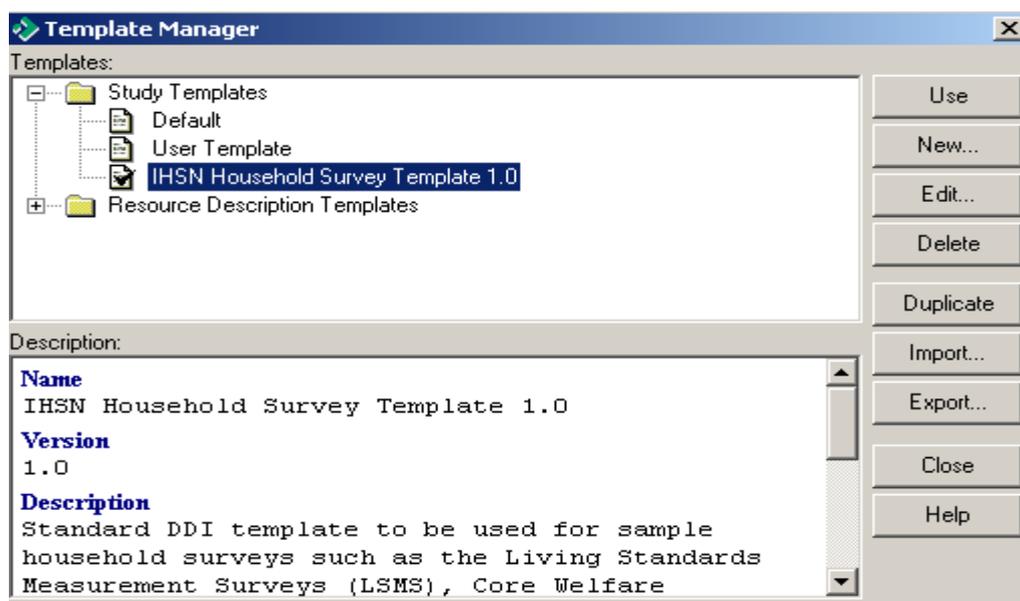


Fig. 2 – Gestione dei template

Il Metadata Editor ne prevede due tipi, una per lo Studio e una per le Risorse Esterne (*Study Template* e *External Resources Template*). La struttura utilizzata ad esempio per i dati dell'indagine svolta in Bosnia Erzegovina è l'*IHSN Household Survey Template*.

I templates sono predefiniti ma personalizzabili: la funzione Template Editor (fig.3) consente di scegliere e modificare la struttura più adatta al tipo di indagine in oggetto, cioè quella contenente tutti gli elementi più appropriati alla sua documentazione.

<sup>4</sup> v. <http://www.internationalsurvenetwork.org/home/?lvl1=tools&lvl2=documentation&lvl3=toolkit> per le condizioni di licenza e di utilizzo.

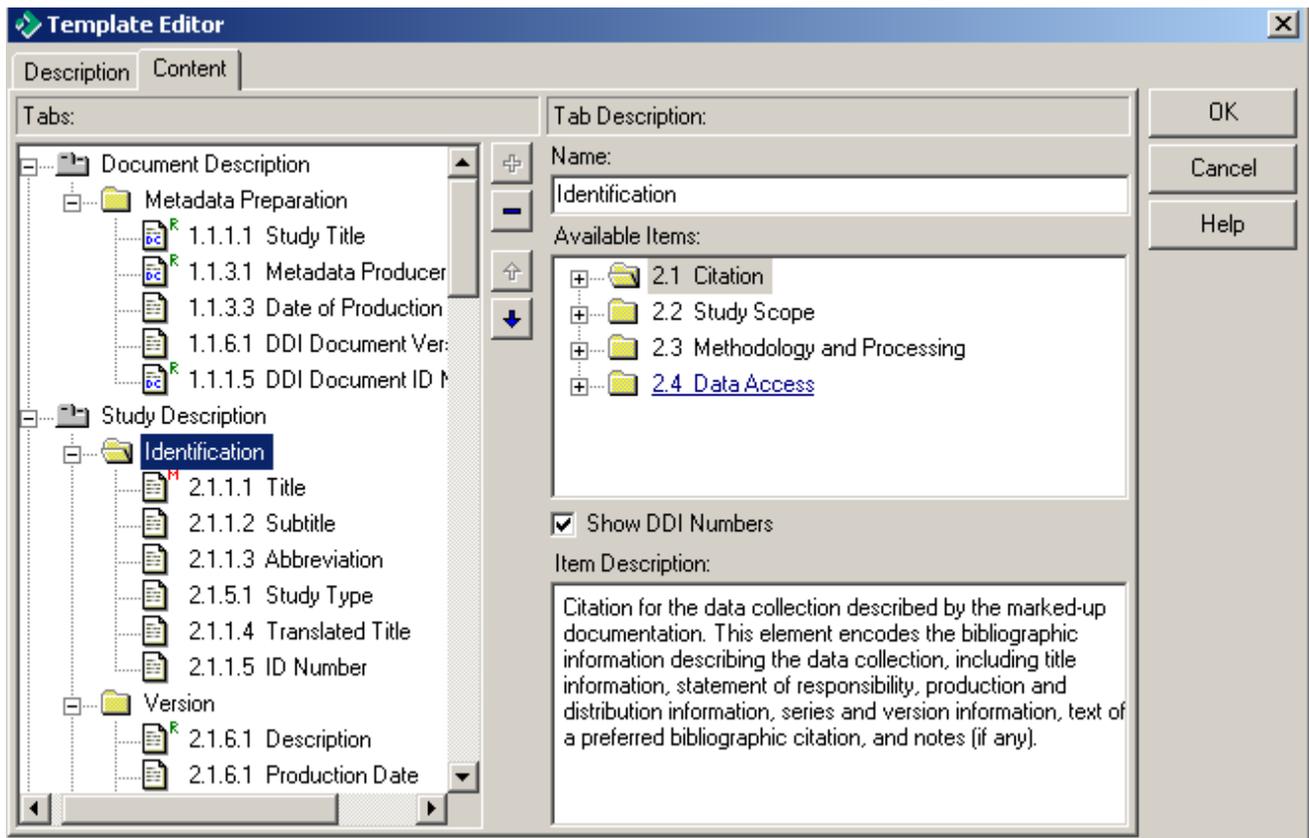


Fig. 3 – Template editor

Una volta scelto il template (essenziale anche perché verrà ripreso e richiesto successivamente dal programma CD Builder) e quindi importato lo studio, è possibile iniziare a documentare l'indagine, cioè ad aggiungere informazioni, metadati e risorse.

La prima specificazione che viene richiesta è la descrizione del documento che sarà l'output finale dell'Editor. Si apre poi la sezione di descrizione dello studio, *Study Description*, in cui gli elementi principali che descrivono lo studio sono raggruppati in tredici sottosezioni, anch'esse derivate dagli standard fissati da DDI):

1. identificazione (nome, tipo di indagine etc.);
2. versione;
3. descrizione generale del progetto;
4. obiettivo e settori di interesse;
5. copertura;
6. produttori e finanziatori;
7. campionamento: vengono date informazioni sul campione utilizzato, procedure e disegno di campionamento, deviazioni, tasso di risposta;
8. raccolta dati, vale a dire tempi e modalità di raccolta, modelli e questionari utilizzati, rilevatori e supervisori;
9. elaborazione dati, una breve descrizione delle procedure utilizzate per il controllo dei dati, identificazione degli errori e la loro correzione;
10. valutazione dei dati, come stima degli errori campionari ed altre forme;
11. accesso ai dati: tutela della riservatezza, politiche di accesso;
12. clausole di diritti e responsabilità;
13. contatti, per richiedere ulteriori informazioni.

La sezione successiva riguarda il dataset vero e proprio, vale a dire tutti i dati associati all'indagine, assieme alla descrizione dettagliata di ciascuna variabile (fig.4).

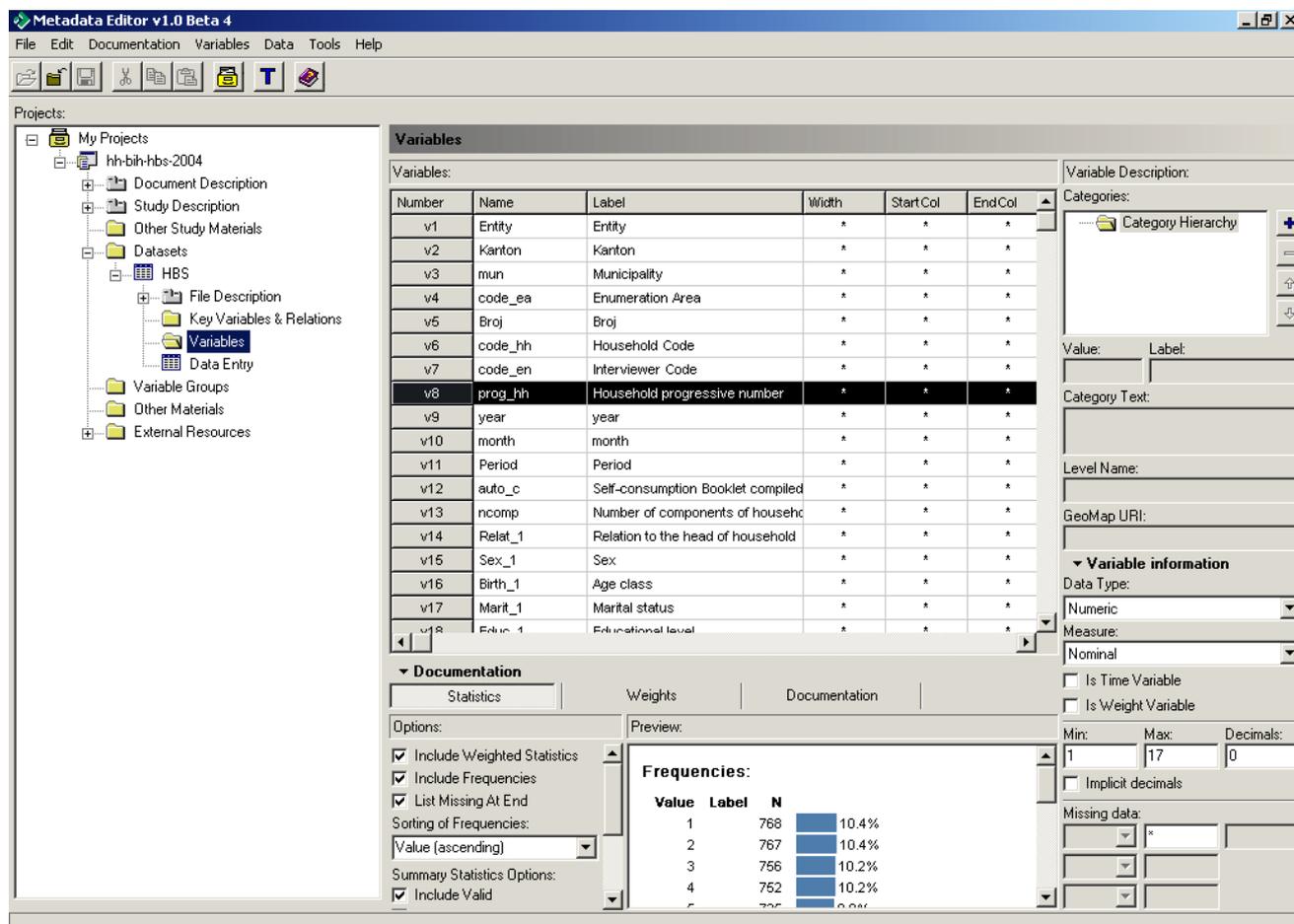


Fig. 4 – Struttura e contenuto del dataset

E' articolata in quattro elementi:

- descrizione del file, contenente gli elementi fondamentali per l'identificazione dei dati: produttori, versione, controlli eseguiti sul processo di elaborazione, valori mancanti.
- variabili chiave e loro relazioni,
- variabili, la sezione in cui vengono inseriti i metadati relativi a ciascuna variabile
- data entry, contenente i microdati.

La documentazione e i microdati relativi allo studio in questione, come già accennato, possono essere corredati ed integrati da tutta una serie di materiali che contribuiscono ad una maggiore comprensione e migliore interpretazione dell'informazione. Sono le cosiddette Risorse Esterne, cui è dedicata per intero l'ultima sezione, che fanno da complemento ai dati di indagine e che possono essere rappresentate da documenti di varia natura: rapporti tecnici e analitici, manuali, modelli e questionari utilizzati per la raccolta dati, tabelle statistiche, software e programmi informatici, fotografie, mappe, ed ogni altro materiale funzionale ed utile ad una più chiara ed ampia comprensione dell'indagine e dei suoi risultati. Vari sono anche i formati supportati dall'applicazione: essa può infatti incorporare file .doc, .xls, .ppt, .pdf, jpeg per le immagini. Per ogni risorsa esterna è prevista l'identificazione, vale a dire una serie di specifiche che ne definiscono origine, formato, autore.

L'impostazione predefinita dell'Editor non permette l'accesso da parte di più persone ad un file di progetto, tuttavia il programma supporta anche modalità di lavoro collaborativo e condiviso (*Shared projects*), come spesso viene ad essere quello di documentazione di un'indagine, in cui i diversi momenti dell'indagine sono affidati a settori o esperti diversi.

L'applicazione, inoltre, include procedure di validazione e controllo dei metadati e delle loro relazioni, e svolge controlli di coerenza e di qualità sulle informazioni contenute, con un ulteriore riflesso positivo sulla qualità dei dati.

L'output finale del processo di documentazione è un file Nesstar, che racchiude l'intera documentazione dello studio, tutti i dati e relativi metadati contenuti ed archiviati in un'unità che risponde ad una 'filosofia pdf', singola, indivisibile e non modificabile.

Vale la pena soffermarsi sulla possibilità, già accennata in precedenza, di rilasciare varie versioni del documento: a seconda del livello di anonimizzazione che si vuole ottenere e/o garantire nella diffusione dei dati, è possibile, e consigliabile, produrre diverse versioni del dataset: ad esempio, una versione 0. con dati grezzi anteriori alle procedure di controllo e correzione, una versione 1. ad uso interno dei produttori ed accessibile solo ad essi, ed una versione 2. con dati controllati ed anonimizzato per la diffusione all'esterno.

Infine, è anche possibile produrre automaticamente un report in formato pdf che contiene tutte le informazioni introdotte mediante il Metadata Editor (fig.5).

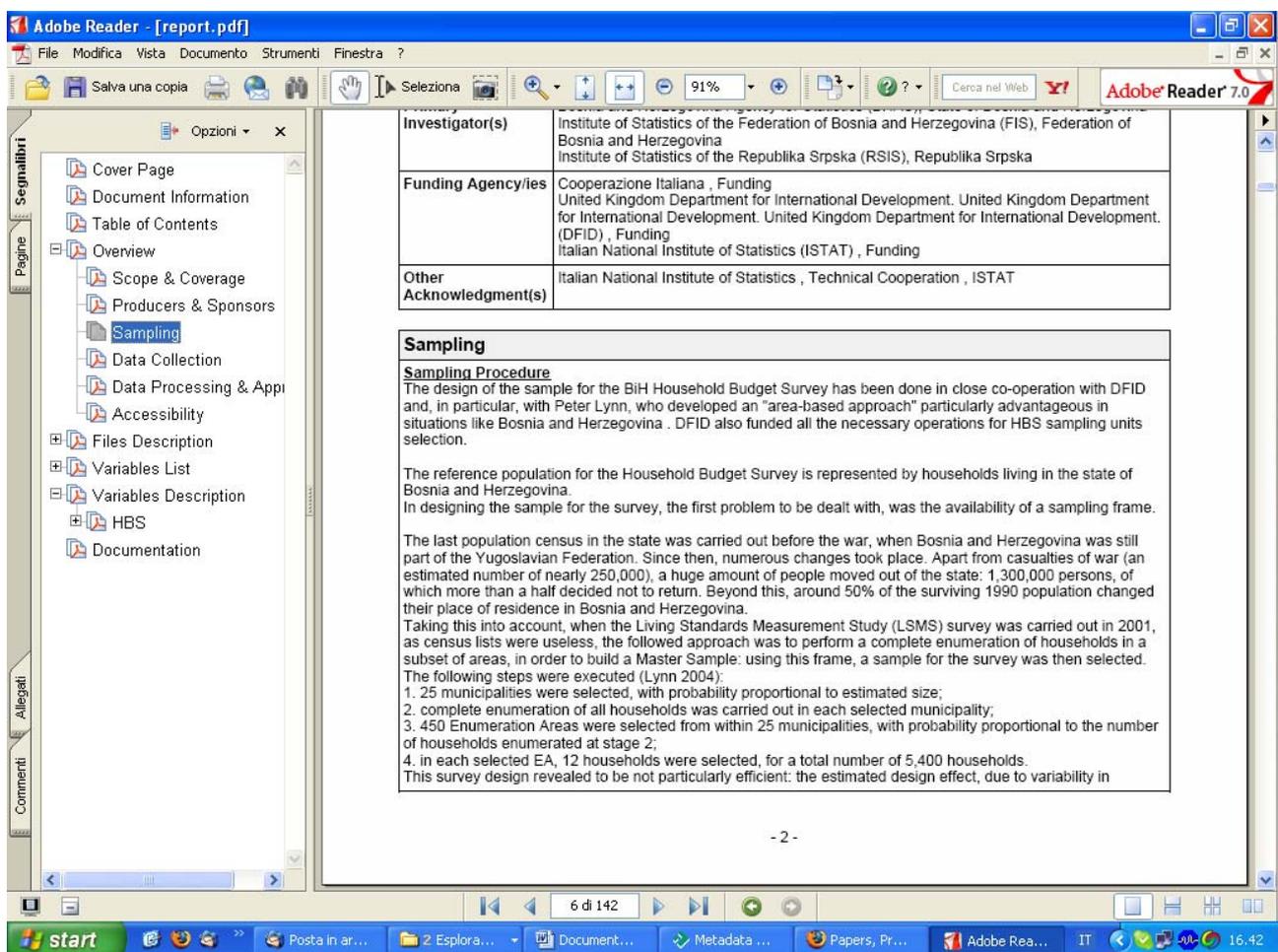


Fig. 5 – Il report automaticamente prodotto dal Metadata Editor

## 2.2. CD Builder

CD Builder, complementare dell'Editor, è l'applicazione che genera un contenuto che racchiude tutta la documentazione relativa all'indagine (dati e i metadati) in un unico output facilmente utilizzabile e navigabile. Questo può essere caricato su un server generando così un sito web per la diffusione online, per una facilità di accesso ai dati e alle indagini ancora maggiore, poiché all'utente è richiesto solo un browser per accedere ai dati.

L'applicazione utilizza il file Nesstar prodotto dall'Editor e partendo da tale dataset in modo estremamente semplice e rapido, genera un contenuto che tiene insieme dati e metadati, le risorse esterne, nonché eventuali materiale aggiuntivo (descrizioni, note, loghi). Partendo dal dataset rappresentato dal file Nesstar esso genera un output (fig.6) basato sul linguaggio HTML che può essere riversato su CD-Rom, DVD, o caricato su server, internet o intranet, pertanto facilmente navigabile e distribuibile; per le sue caratteristiche di completezza tale output può essere utilizzato anche, come già accennato nell'introduzione, ai fini di archivio, conservazione e 'memoria storica' di un istituto.

Fermo restando che il file Nesstar ottenuto tramite Editor non andrebbe modificato nella struttura o nei contenuti è tuttavia possibile personalizzare il contenuto del CD, soprattutto per quanto riguarda il suo aspetto, tramite il branding che identifica l'istituzione o ente responsabile dell'indagine, e la possibilità di modificare lo schema creato rimuovendo sezioni ed elementi non utilizzati, nonché di aggiungere materiali e descrizioni non strettamente correlato all'indagine.

Una volta completato il processo di costruzione e integrazione del file Nesstar, il passo successivo e finale della produzione di un output destinato alla vasta diffusione è la sua trasformazione in pagine in formato HTML, poi riversate su siti web per fini di diffusione, o su cdrom o dvd, per fini sia di diffusione che di archivio e conservazione.

Il processo prende avvio da un semplice passaggio di importazione, aprendo l'applicazione CD Builder e selezionando dal menu File l'opzione *New Cdrom Project* (fig.6): si apre la finestra di dialogo che permette di importare lo studio in formato Nesstar e di assegnare il branding, cioè una delle opzioni di personalizzazione del prodotto finale.

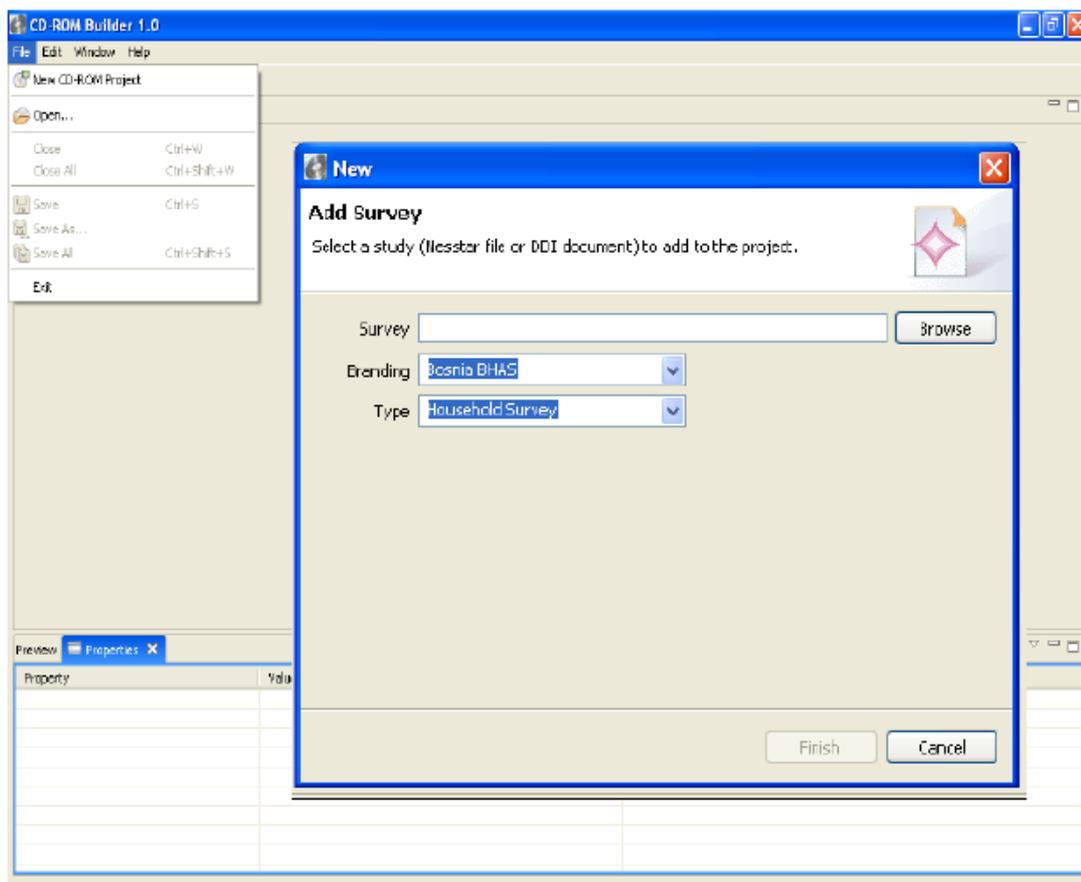


Fig. 6 – Finestra di importazione dello studio in formato Nesstar

L'applicazione crea automaticamente un *outline* che definisce il contenuto e la struttura che presenterà il Cdrom, popolata da tutti gli elementi precedentemente inclusi nel Metadata Editor:

A questo punto, il processo di generazione è in larga parte automatico. Il passo successivo infatti è la selezione menu *Tools > Export to Cdrom* (fig.7), dove viene richiesto di selezionare la posizione e la cartella in cui verranno salvati i contenuti del cdrom: il programma infatti non masterizza

automaticamente un Cdrom, ma crea un ‘master’ che dovrà essere trasferito successivamente su un cd, dvd o server per il caricamento su sito web.

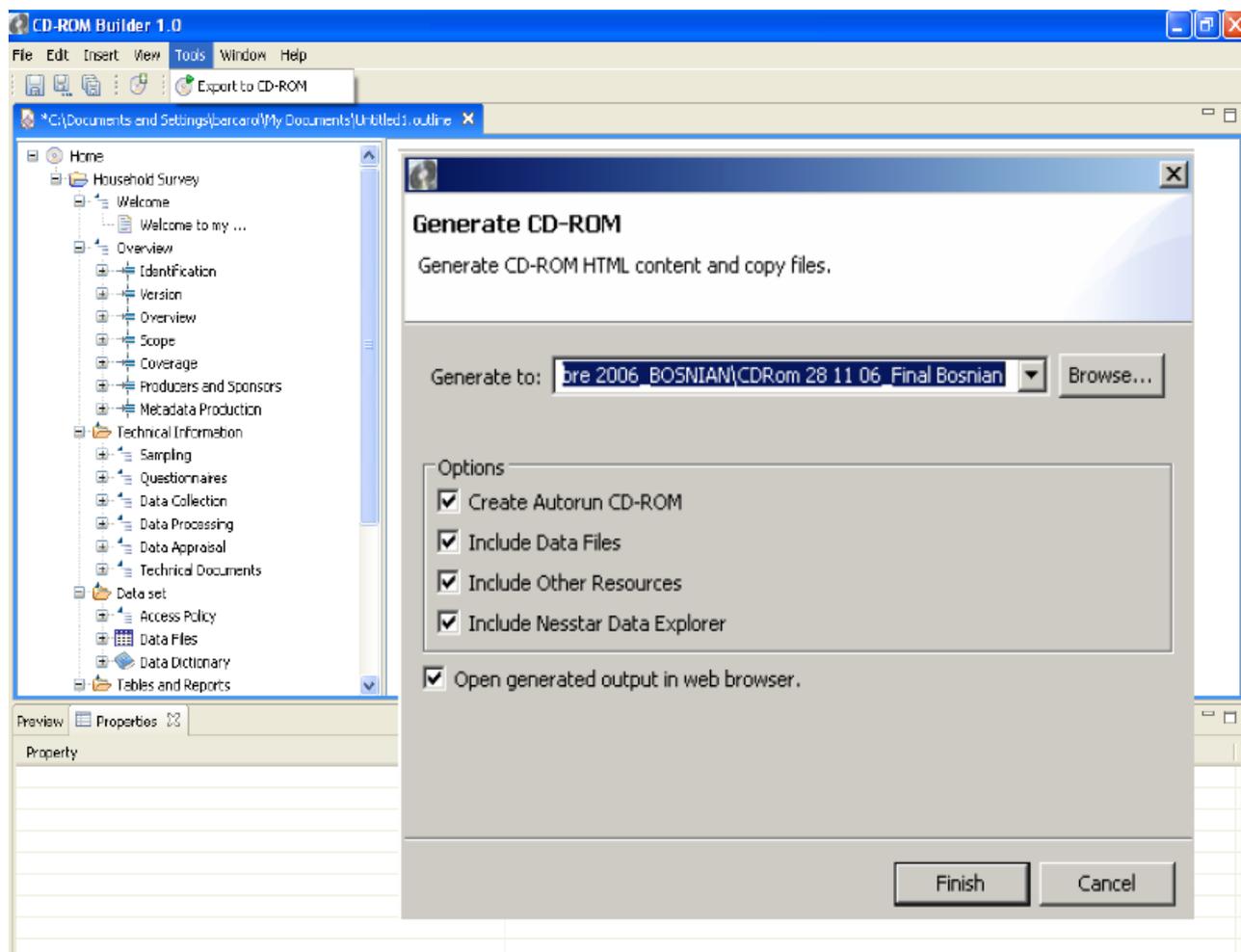


Fig. 7 – Funzione di creazione del CD-Rom

Viene inoltre richiesto di scegliere una serie di opzioni:

1. la selezione della prima opzione *Create Autorun Cdrom* fa sì che il cd parta automaticamente quando inserito nel computer dall'utente;
2. l'opzione *Include data files* include il file di microdati nel cdrom; va deselezionata nel caso in cui si voglia produrre un cd o caricare le pagine sul sito senza microdati inclusi;
3. l'opzione *Include other resources* copia tutte le risorse esterne di cui si è corredato il contenuto nell'Editor; ad esempio nel caso della Bosnia, come già accennato, sono stati inclusi un volume di analisi dei risultati finali e una nota metodologica in formato pdf, i modelli dei questionari utilizzati per la raccolta dei dati e le tavole statistiche in formato Excel;
4. la quarta opzione allega al cd l'applicazione Nesstar Explorer, necessaria all'utente per esportare il file di dati nel formato statistico desiderato (STATA, SPSS, ASCII etc).

Infine, selezionando *Open generated output in web browser* è possibile ottenere un'anteprima del prodotto finale sul browser installato sul computer.

Il processo di generazione a questo punto viene completato in pochi minuti, e ne risulta un output in formato Html di facile navigazione, utilizzo e consultazione.

## 2.3. Nesstar Explorer

Il file Nesstar generato da Metadata Editor non è leggibile dai software standard: il Toolkit è completato da Nesstar Explorer, un programma libero che viene installato automaticamente sul computer al momento dell'installazione del Toolkit oppure, in alternativa, viene fornito nel CD o nel sito per gli utenti esterni, poiché viene incluso automaticamente, selezionando l'apposita opzione, nell'output costruito da CD Builder.

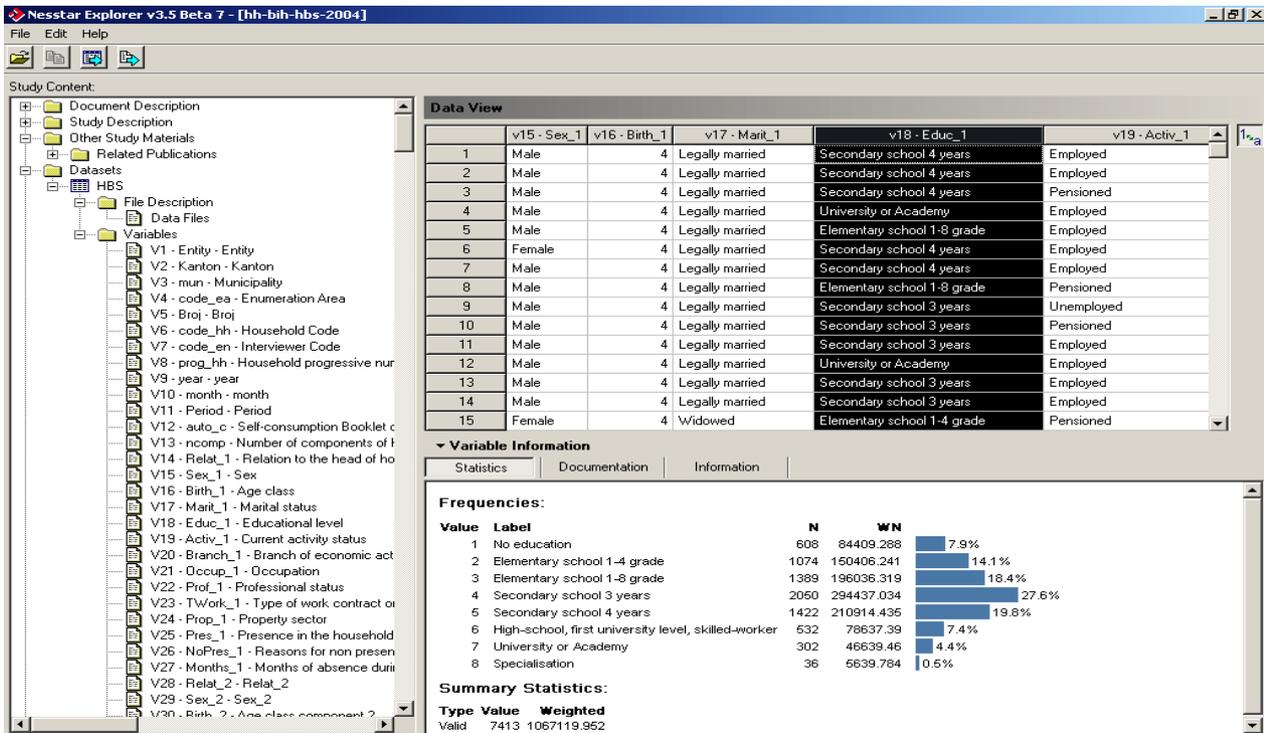


Fig. 8 – Navigazione nei microdati mediante Nesstar Explorer

Una volta lanciato, il programma chiede all'utente di selezionare il file Nesstar di interesse, e permette sia la navigazione dei dati in esso contenuti (fig.8) che la loro esportazione nei più comuni standard statistici quali ad esempio SAS, STATA, SPSS ecc., ai fini di successive analisi ed elaborazioni (fig.9). L'applicazione permette di esaminare tutte le variabili dell'indagine, nonché di visualizzare statistiche riepilogative e informazioni dettagliate per ciascuna variabile. Nell'applicazione sono disponibili anche tutte le informazioni riguardanti le risorse esterne, se precedentemente integrate nel file mediante il CD Builder. Permette inoltre di esportare, oltre ai dati, anche i metadati dello studio.

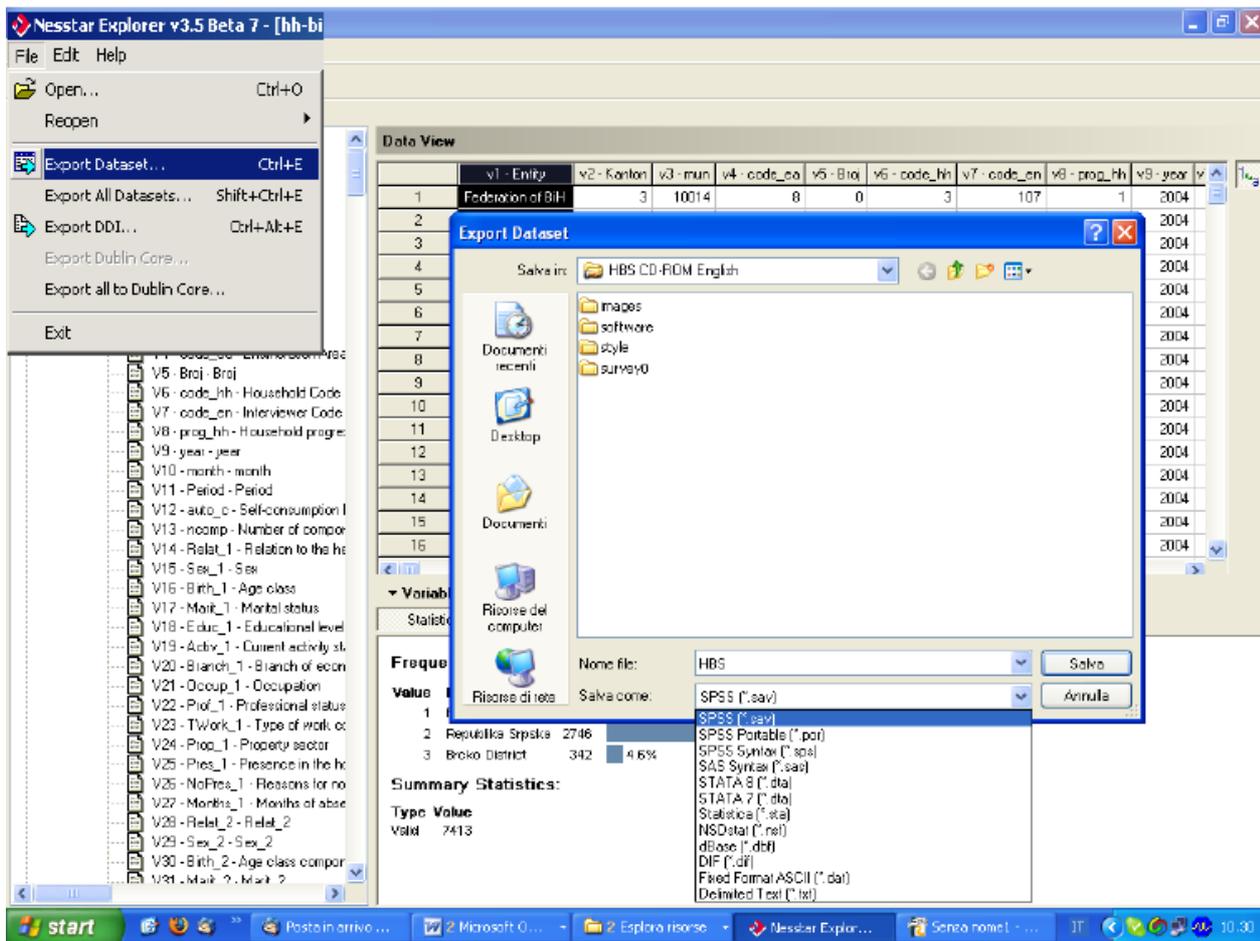


Fig. 9 – Esportazione dei microdati nei diversi formati

### 3. Un caso di studio: l'Indagine sui Consumi in Bosnia Erzegovina

Nel triennio 2003-2005 l'Istituto Nazionale di Statistica ha assicurato assistenza tecnica ai tre Istituti di Statistica della Bosnia Erzegovina (Agenzia di Statistica, Istituto della Federazione e Istituto della Repubblica Serba) per la conduzione dell'indagine sui Consumi (Household Budget Survey). Le operazioni di raccolta dati hanno riguardato tutto il 2004. Nel 2005 i dati sono stati sottoposti a procedure di validazione, e le stime di interesse sono state calcolate e diffuse, sia su supporto cartaceo che mediante un data warehouse esposto sul sito ufficiale dell'Agenzia<sup>5</sup>. Si decise di mettere a disposizione anche il file di microdati d'indagine per permettere ai ricercatori esterni agli Istituti (tra cui quelli di vari enti internazionali) di condurre in proprio le elaborazioni di interesse. A tal fine la World Bank offrì il proprio toolkit, in quei mesi ancora in fase di completamento. Fu rapidamente sviluppata e messa a disposizione sul sito dell'Agenzia una versione in lingua inglese del file, mentre tempi più lunghi ha richiesto lo sviluppo dell'analoga versione in lingua bosniaca.

Le attività svolte hanno riguardato:

1. la costruzione di un file di microdati protetti dal punto di vista della riservatezza, a partire dalla disponibilità dei microdati validati;
2. il caricamento del file dei microdati protetti all'interno del Metadata Editor, tool del Microdata Management Toolkit, e definizione dei relativi metadati;
3. utilizzo del tool CD-Builder per la generazione delle pagine HTML per la consultazione di dati e metadati, e caricamento sul server del sito web dell'Agenzia di Statistica.

<sup>5</sup> <http://www.bhas.ba>

### 3.1. Costruzione del file dei microdati protetti<sup>6</sup>

Nella produzione di un file di microdati “sicuro” dal punto di vista della protezione della riservatezza, da rendere disponibile all’utenza esterna agli Istituti di statistica della Bosnia Erzegovina, è stata adottata la stessa metodologia utilizzata dall’Istituto Nazionale di Statistica in Italia (Franconi, Seri, 2004).

Tale metodologia si basa sulla valutazione del rischio di violazione effettuata a livello di singolo individuo, ipotizzando uno scenario nel quale un soggetto intenzionato a violare la riservatezza dei dati abbia a disposizione una base di dati contenente sia gli identificativi degli individui, che un insieme di variabili chiave che permettano di mettere in collegamento (*linkage*) tali identificativi con i singoli record del file rilasciato. Per ogni collegamento (*match*), attuato quando la combinazione dei valori delle variabili chiave corrispondenti all’identificativo di un individuo nella base dati è la stessa di quella di un record del file esposto, il relativo rischio di violazione (*disclosure risk*) è definito come la probabilità che tale collegamento sia corretto (cioè che il record corrisponda effettivamente all’identificativo nella base dati).

L’approccio seguito consente di stimare sia il rischio relativo ai singoli individui (*Base Individual Risk, BIR*), che il rischio associato al livello superiore delle famiglie (*Base Household Risk, BHR*). Valutando il rischio a livello di famiglia, implicitamente si tiene conto del rischio a livello di individuo, che non sarà mai superiore a quello valutato per la famiglia di appartenenza, in quanto quest’ultima possiede una quantità di informazione utile per il potenziale linkage che è somma di quella di ogni singolo componente: ne deriva che l’informazione relativa ad un singolo individuo è “sicura” se:

- il BHR è inferiore ad una determinata soglia  $s$ ;
- il BIR è inferiore alla stessa soglia  $s$ , divisa per il numero di componenti  $n$  della famiglia.

Questo implica un crescere del livello di protezione individuale al crescere della dimensione della famiglia, il che ha tanto più senso quanto più lo scenario di violazione assume che si cerchi di effettuare link a livello di individuo, sapendo a quale famiglia questo appartiene. In altre parole, per integrare entrambi i livelli individuale e familiare, l’analisi del rischio è condotta sul file dei record individuali dove sono inclusi anche gli identificativi delle famiglie.

Una volta stimato il rischio per il complesso dei record individuali del file (tenendo conto anche dei pesi campionari associati ad ognuno di essi), si procede a fissare una soglia di rischio accettabile, al di sotto della quale un record può essere dichiarato “sicuro”. Per ogni soglia di rischio, il numero di record che la eccedono può essere contenuto ricorrendo ad opportune azioni di ricodifica globale, consistenti nell’aggregare le modalità di alcune variabili identificative per il complesso dei record. Per tutti i record che eccedono tale soglia, vengono intraprese delle azioni di protezione, ricadenti nella categoria della soppressione locale.

Sia la valutazione del rischio, che le opportune manipolazioni dei dati per ricondurlo al di sotto di una soglia accettabile per tutte le unità individuali, sono attività effettuate mediante il software  $\mu$ -Argus<sup>7</sup>.

Nel nostro caso, il file dei Consumi dell’indagine BiH HBS 2004 contiene record relativi alle famiglie osservate. Le variabili definite come identificative sono: l’età, il sesso e lo stato civile dei singoli componenti, il numero dei componenti e la tipologia della famiglia, e l’entità (Federazione o Repubblica Serba) di appartenenza.

Il primo passo è consistito nell’effettuare una ricodifica globale della variabile “età del componente” al fine di sostituire tale variabile con la variabile derivata “classi di età” (0-5;6-17;18-34;35-64;65->), in quanto la disponibilità di valori puntuali avrebbe comportato livelli medi di rischio troppo elevati.

E’ stata quindi fissata una soglia di rischio pari a 0,01: questo significa che una osservazione individuale può essere considerata “sicura” se nella popolazione ve ne sono almeno 100 con la stessa combinazione di variabili identificative<sup>8</sup>.

<sup>6</sup> A questo paragrafo ha collaborato Giovanni Seri, dell’unità PSM/C (“Metodologie statistiche per la tutela della riservatezza”)

<sup>7</sup> Per informazioni sul progetto europeo CASC che ha prodotto  $\mu$ -Argus, vedi il sito: <http://neon.vb.cbs.nl/rsm/casc>

<sup>8</sup> Per tornare alla definizione iniziale, questo implica che se viene effettuato un collegamento tra un record individuale nel file rilasciato con una base dati di identificativi, la probabilità che l’associazione, stabilita tra i valori dell’osservazione ed un particolare identificativo nella base di dati, sia corretta risulta essere inferiore all’1%.

Il software  $\mu$ -Argus, dato l'insieme di variabili identificative e la soglia 0,01, ha individuato 95 record il cui rischio eccedeva tale soglia. Per tali record si è proceduto a effettuare soppressioni locali, in modo da ricondurli al di sotto di tale soglia. Nello scegliere le variabili rispetto alle quali effettuare soppressioni, è stato dato un peso a ciascuna di esse, evitando ad esempio di intervenire sulla variabile territoriale "entità" (assegnando peso 100, pari al massimo). Nella tabella seguente sono riportati gli interventi effettuati nel complesso.

Variabile	Peso	Numero di soppressioni
Sesso	70	1
Stato civile	50	102
Tipologia di famiglia	100	4
Entità	100	0
Classe di età	60	37
Totale	-	144

Il file dei microdati così trattati è quindi da considerarsi sicuro rispetto alla protezione della riservatezza, e può essere diffuso all'utenza esterna.

### 3.2. Caricamento del file dei microdati protetti e definizione dei metadati

Una volta prodotto il file dei microdati protetti, è possibile caricarlo all'interno del Metadata Editor per procedere alla definizione dei metadati.

Relativamente al formato dei dati in input al Metadata Editor, sono possibili diverse opzioni:

1. formato ASCII;
2. formato proprio di un particolare software statistico (SPSS, STATA, Statistica, SAS);
3. formato DDI XML (nel caso in cui dati e metadati siano già stati descritti con precedenti progetti di Metadata Editor).

Nel primo caso, è necessario ridefinire *ex novo* tutti i metadati operativi del tracciato record: il nome delle variabili, eventuali label, le colonne di inizio e di fine, l'ampiezza, i decimali. Solo dopo aver provveduto interattivamente a definire quanto sopra, è possibile importare i dati. Nel caso dei dati HBS della Bosnia sarebbe stato un lavoro estremamente oneroso, data la presenza di 574 variabili.

Poiché parte del trattamento dei dati è stato effettuato con programmi SAS, si è pertanto deciso di procedere al caricamento di un dataset con quel formato, sfruttando pertanto la definizione del tracciato record (nome ed etichette delle variabili, posizioni nel record) già utilizzata nell'ambiente SAS. A differenza degli altri software statistici, l'importazione di un dataset SAS non è altrettanto immediata: occorre prima eseguire, in ambiente SAS, una procedura costituita dai seguenti passi:

1. in SAS, selezionare *File > Export Data* e selezionare il dataset SAS da esportare (nel nostro caso, il dataset HBS.sasb7dat);
2. scegliere il Comma-Separated Values (.csv) come formato di esportazione;
3. preparare i files .sp1 and .sp2 richiesti dal Metadata Editor utilizzando il seguente programma SAS:

```
LIBNAME LIBRARY 'G:\HBS2004\';
LIBNAME IN 'G:\HBS2004\';
PROC CONTENTS DATA=IN.HBS;
RUN;
DM 'OUTPUT; FILE 'G:\HBS2004\HBS.SP1''';
DM 'OUTPUT; CLEAR';
```

```
TITLE 'HBS';
PROC FORMAT LIBRARY=LIBRARY FMTLIB;
RUN;
DM 'OUTPUT; FILE 'G:\HBS2004\HBS.SP2''';9
```

4. uscire da SAS e lanciare Metadata Editor;
5. selezionare *File > Import* nel menu principale;
6. selezionare SAS come formato di importazione, e fornire il nome e la locazione dei files nella finestra di dialogo. Il Metadata Editor controlla che i 3 necessari files (sp1, sp2 e csv) siano presenti, e procede ad importare i dati.

Le operazioni di cui sopra hanno permesso di importare i dati dell'indagine HBS (7413 record con 574 variabili) all'interno del Metadata Editor, trovando già inizializzati i metadati relativi al nome delle variabili ed al tracciato del record.

Il lavoro successivo è consistito in primo luogo nel completare la sezione relativa ai metadati operativi, quelli funzionali alla comprensione delle caratteristiche “informatiche”, e cioè:

- o le variabili “identificative”;
- o la caratterizzazione in “gruppi” delle variabili;

Rispetto alle prime, è stato indicato l'identificativo univoco, o chiave primaria, costituito dal codice della famiglia (da 1 a 7413).

Rispetto ai gruppi, è stata seguita la strutturazione in sezioni dei questionari di rilevazione. In particolare, è stata trattata la sezione relativa ai componenti della famiglia individuando un gruppo “periodico” (da 1 a 10) dove ogni istanza riporta variabili di carattere demo-sociale relative ad ogni componente. A seguire:

- o i dati sulle abitazioni (primaria e secondaria), contenente informazioni anche sullo status legale, sui beni durevoli, sull'affitto reale o stimato;
- o i dati di spesa organizzati in capitoli: mobilio ed articoli per la casa, abbigliamento, salute, trasporti e comunicazioni, tempo libero e cultura, cibo e bevande, spese varie ed autoconsumi.

Completati i metadati operativi, si è passati a fornire quelli di processo. Per far questo, ci si è avvalsi in modo particolare di quanto già descritto nella pubblicazione di carattere metodologico già disponibile<sup>10</sup>. Le descrizioni di fasi quali quelle relative al disegno del campione, la raccolta dei dati, il loro trattamento (controllo e correzione), la produzione delle stime ed il calcolo della varianza campionaria sono quindi state semplicemente inserite nelle relative sezioni con un semplice lavoro di copia ed incolla dai rapporti già prodotti.

Infine, sono state definite le “External Resources”, quei documenti utili ad ampliare la conoscenza sui dati a disposizione. In particolare, sono stati inseriti i tre questionari utilizzati nella fase di raccolta (nella versione .pdf), il già citato rapporto metodologico, e un ulteriore studio contenente l'analisi dei risultati<sup>11</sup>. Sono anche state messe a disposizione una serie di tabelle Excel che riportano informazioni di carattere aggregato utili per la comprensione dei fenomeni oggetto di studio.

### 3.3. La costruzione del CD-ROM e delle pagine HTML per la consultazione online.

Una volta importati i dati nel Metadata Editor dal dataset SAS, si è proceduto a:

- a) raggruppare opportunamente le variabili del dataset, seguendo la struttura del questionario;

<sup>9</sup> E' necessario prima lanciare un programma di creazione di un catalogo SAS nella directory 'G:\HBS2004\', eseguendo ad esempio il programma:

```
proc format library=hbs;
value entity 1='Federation BiH'
            2='Republika Srpska'
            3='Brcko District';
run;
```

<sup>10</sup> “The BiH Household Budget Survey – Methodology”, pubblicata congiuntamente da ISTAT e dai tre uffici di statistica della BiH (BHAS, FIS e RSIS).

<sup>11</sup> “The BiH Household Budget Survey – Final results”, anch'esso pubblicato congiuntamente

- b) inserire le notizie mancanti relativamente alle singole variabili del dataset (in particolare, le classificazioni);
- c) inserire la descrizione delle varie fasi del processo di produzione dei dati.

Ad esempio, essendo definiti fino a dieci componenti, tutte le notizie relative ad ognuno di essi sono state opportunamente raggruppate (fig.10).

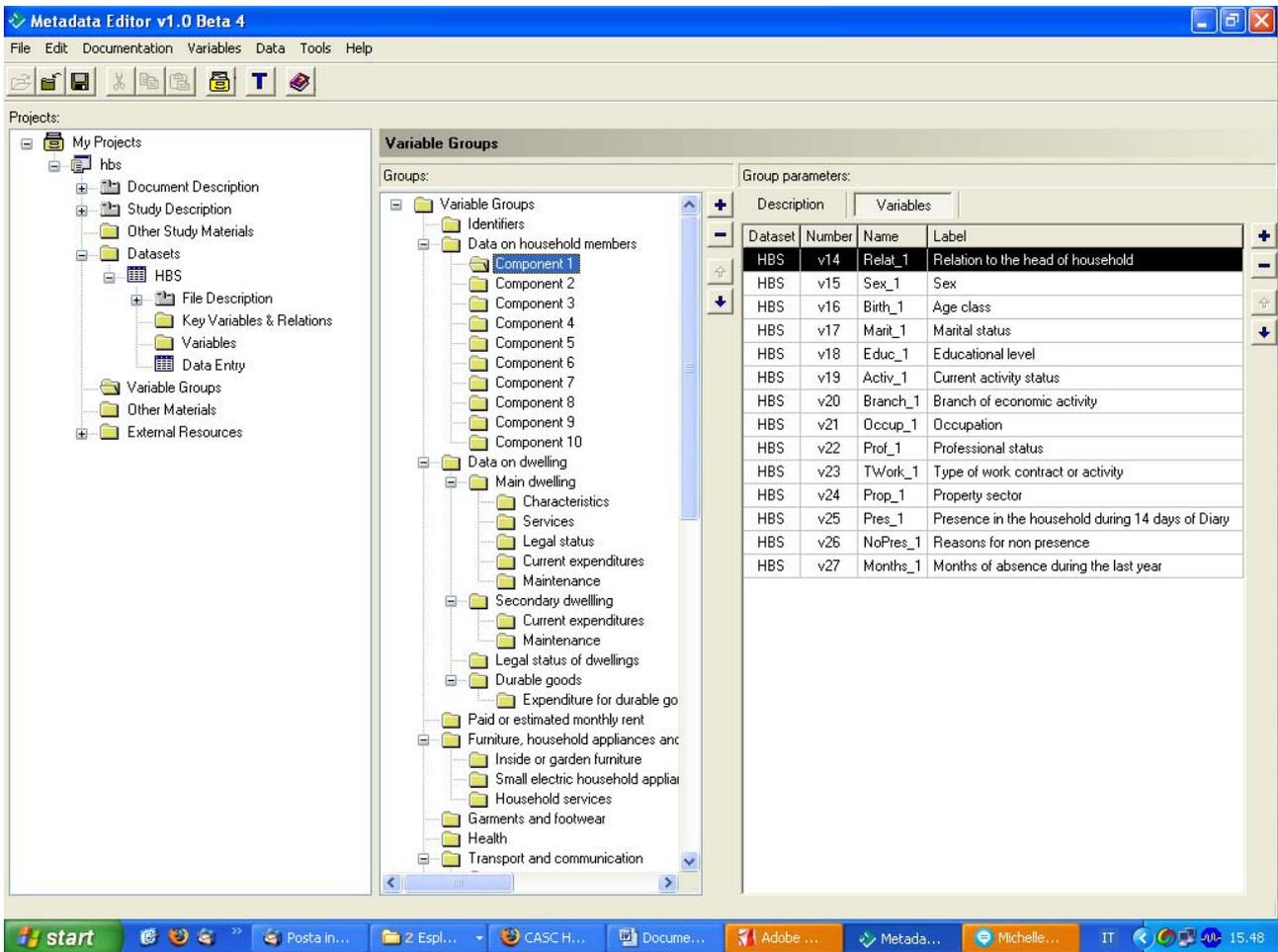


Fig. 10 – Strutturazione in gruppi delle variabili

Per ogni variabile del primo gruppo sono state definite manualmente le classificazioni, cioè l'insieme di valori ammissibili (descrizione e codifica per ognuno di essi). Una volta definite, le classificazioni possono essere riutilizzate dalle altre variabili che vi si riferiscono. Nella seguente figura è riportato un esempio di definizione di classificazione (per *Marital Status*). Nello stesso pannello (fig.11) è possibile definire il *tipo* della variabile (se numerica, alfanumerica o data), e la *misura* (scalare, ordinale, nominale).

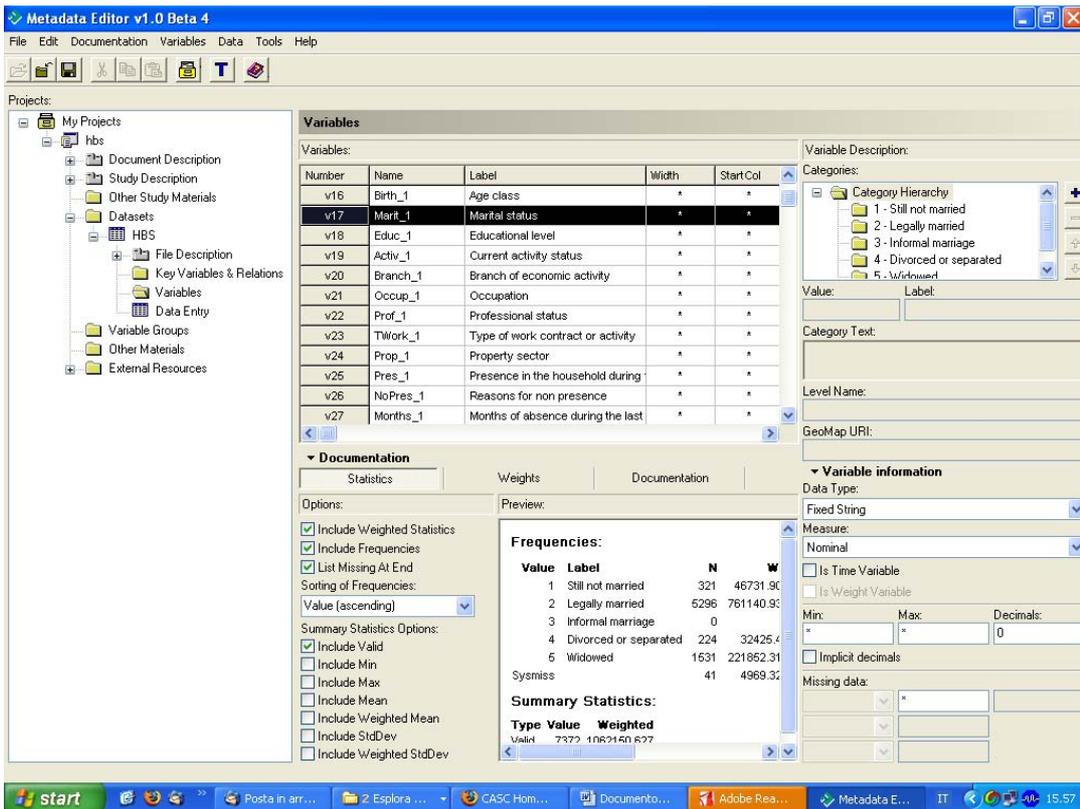


Fig. 11 – Trattamento delle singole variabili (definizione classificazioni e inserimento tipo e misura)

Una volta terminata la messa a punta dei metadati “operativi”, si è passati a quelli di carattere generale e di processo. In particolare, sono state implementate le seguenti sezioni: *Sampling, Data Collection, Data Processing, Data Appraisal, Data Access.*

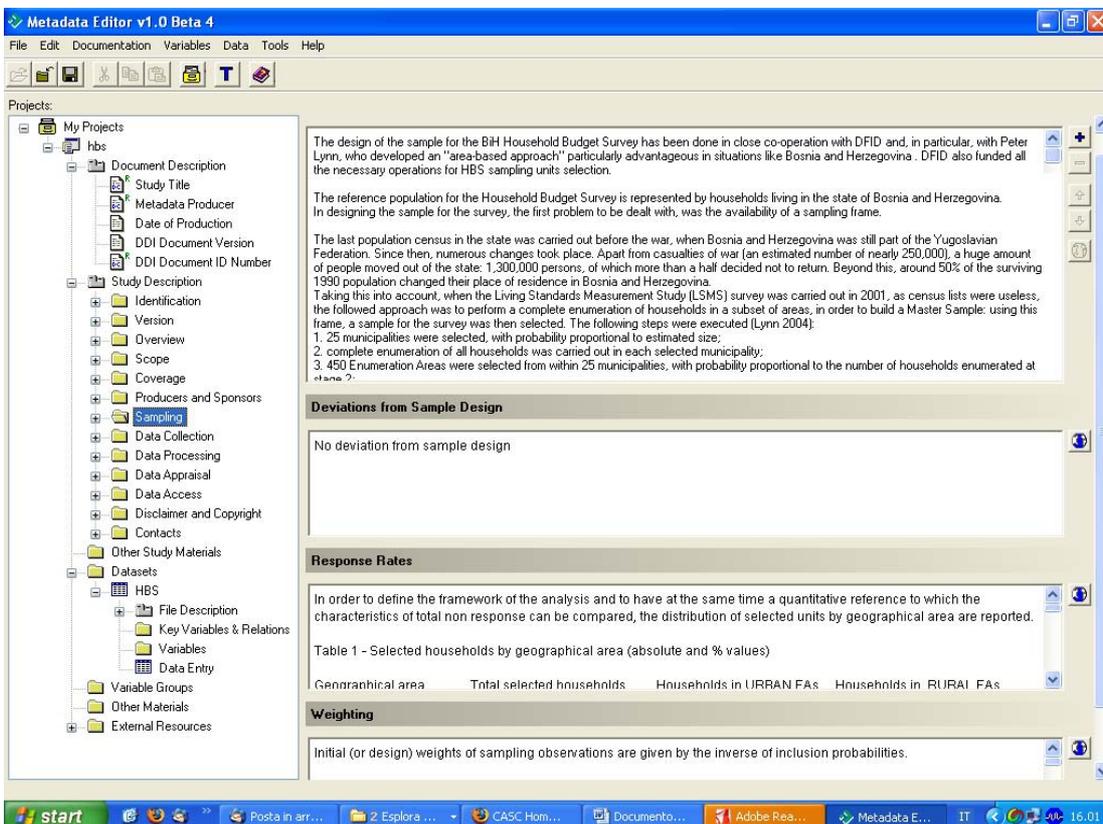


Fig 12 – Descrizione degli aspetti campionari dell'indagine HBS

A questo punto si è provveduto a definire le *External Resources*, tutto il materiale ausiliario cioè utile a corredare i dati. In una prima versione (fig.13), questo è consistito nei seguenti elementi:

- i tre questionari utilizzati dall'indagine, in formato pdf;
- la pubblicazione di carattere metodologico (pdf)
- le presentazioni PowerPoint utilizzate nel convegno di chiusura del progetto di cooperazione.

Nella versione finale, sono stati mantenuti i tre questionari e la pubblicazione di carattere metodologico, ma una volta provveduto ad inserire anche quella di analisi dei risultati (BiH Statistical Institutes and ISTAT 2005b), sono state eliminate le presentazioni.

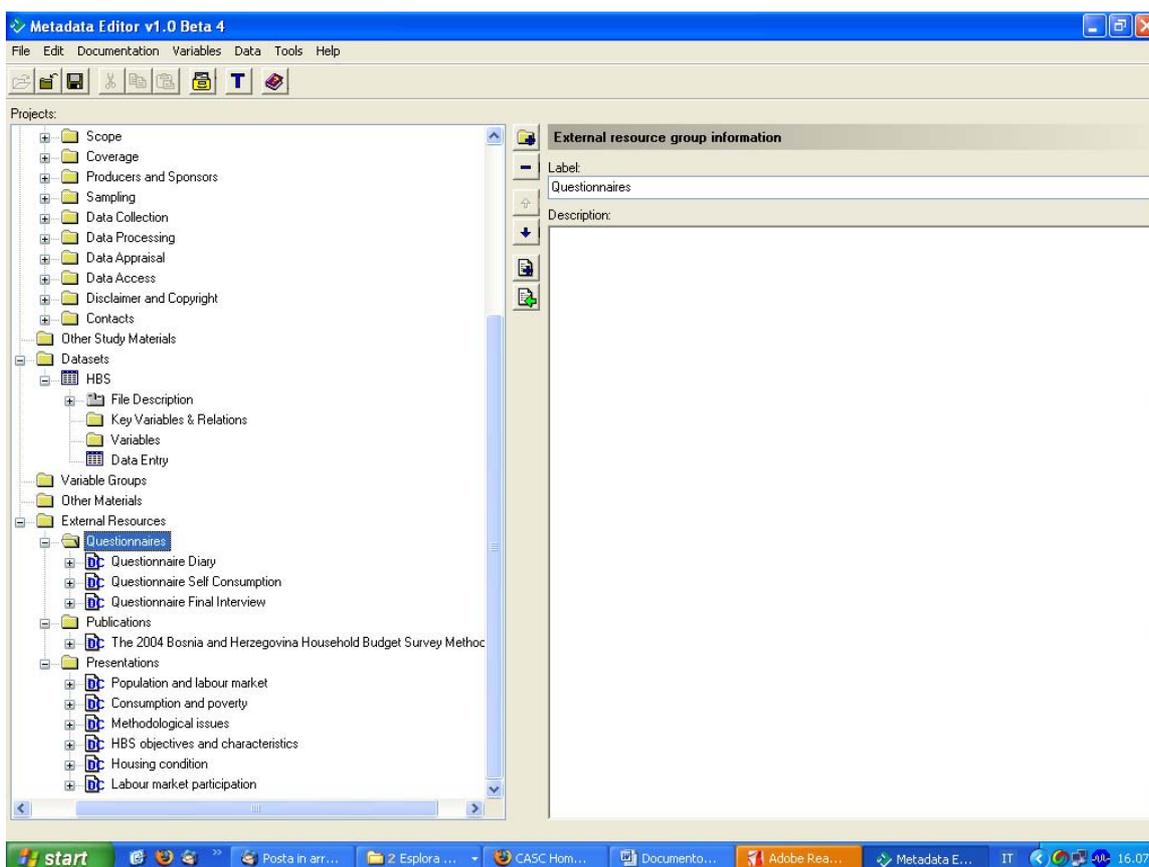


Fig 13 – *External resources*

Completate le attività di messa a punto dei metadati, si è provveduto ad utilizzare il CD-Builder per generare il CD-Rom, ma soprattutto le pagine web da esporre nel sito ufficiale della BHAS (BiH Agency for Statistics). L'aspetto di tali pagine è riportato nella figura 14.

Infine, si è proceduto a costruire la versione in bosniaco delle pagine web. A tal fine, è stato necessario ripartire dalla definizione dei metadati nel Microdata Editor, effettuando una traduzione di tutti i termini che lo richiedevano: label delle variabili, descrizioni dei singoli item delle classificazioni, nonché tutti i testi contenuti nelle descrizioni di processo.

Non è stato possibile procedere ad una traduzione dei termini in inglese utilizzati internamente dalle componenti del toolkit<sup>12</sup>.

Nella figura 15 è riportato un esempio di pagina web nella versione bosniaca.

<sup>12</sup> Una prossima versione del toolkit dovrebbe consentire questa possibilità

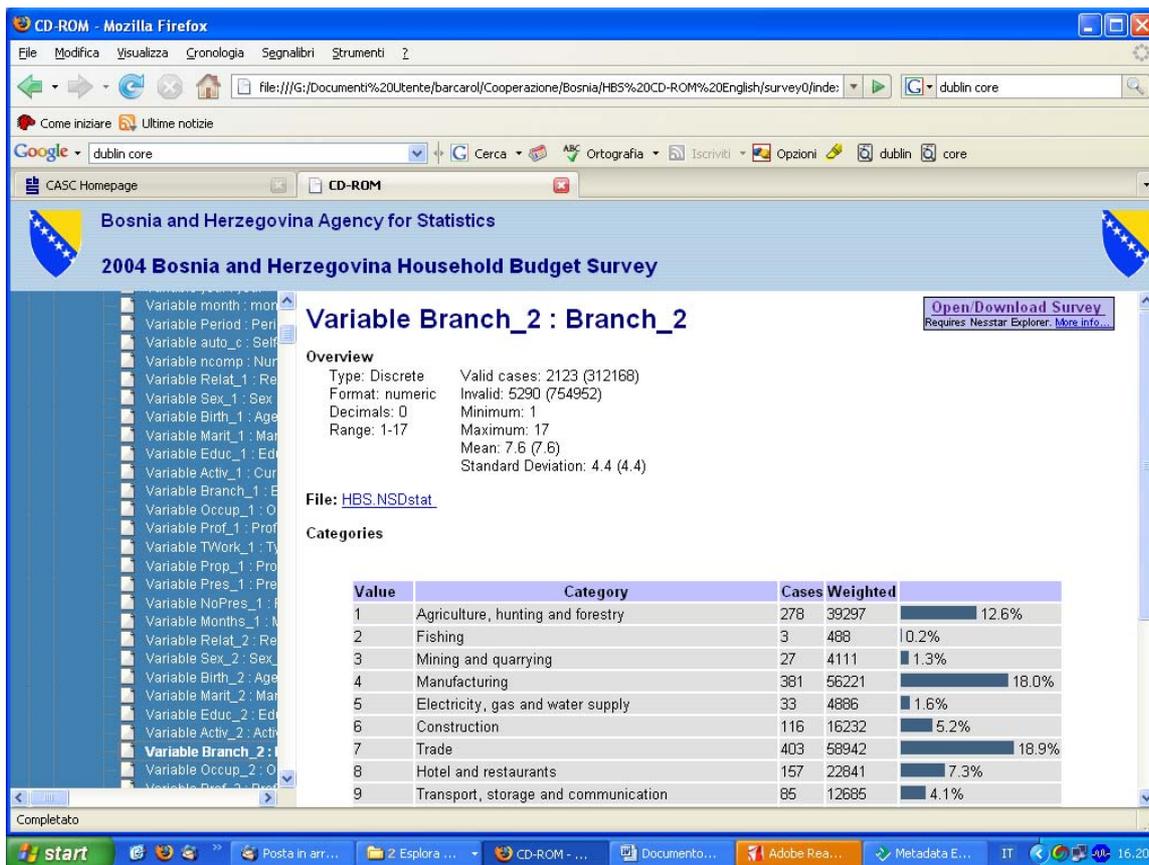


Fig. 14 – Pagine web (versione inglese) esposte nel sito BHAS

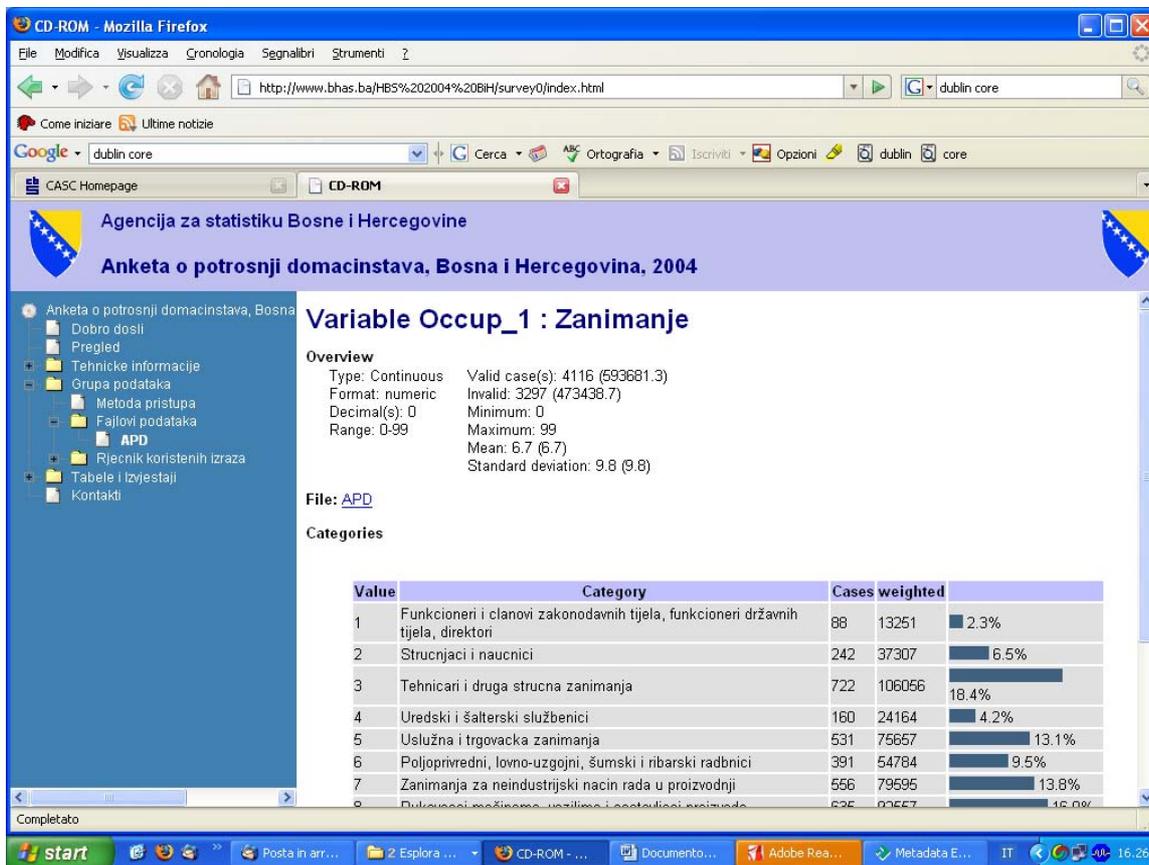


Fig 15 - Versione bosniaca delle pagine web

## 4. Conclusioni

L'esperienza di utilizzo del Microdata Management Toolkit dell'International Household Survey Network della World Bank, svolta nell'ambito del progetto di cooperazione ISTAT con gli Istituti della Bosnia-Erzegovina per la conduzione dell'Indagine sui consumi delle famiglie, ha dato indicazioni preziose per l'adozione di tali strumenti anche all'interno del nostro Istituto.

Il rispetto degli standard internazionali assicurato dal toolkit permette di ottenere un prodotto disseminabile agli utenti esterni ottimale sotto il profilo della completezza dei metadati. Questi ultimi sono sia di tipo concettuale-statistico, garantendo quindi la comprensione dei processi di produzione dei microdati validati, che informatico-operativo, permettendo di mettere a disposizione tutto quanto necessario per procedere alle elaborazioni di interesse. Inoltre, la semplicità di utilizzo permette di ridurre al minimo il carico di lavoro necessario per produrre i CD-Rom o il sito web di esposizione dei dati.

Sulla base dell'esperienza riportata, riteniamo utile avanzare una proposta di procedura standard di messa a disposizione di microdati per l'utenza esterna.

Il responsabile dell'indagine, una volta provveduto a introdurre il file dei dati validati della propria indagine all'interno di ARMIDA (Archivio dei Microdati Validati dell'Istituto), certificandone in tal modo l'ufficialità, dovrebbe eseguire i seguenti passi:

1. a partire dai dati validati, produrre le due versioni (file pubblico e file per ricerca) di microdati per l'utenza esterna, utilizzando il software  $\mu$ -ARGUS per la valutazione del rischio di violazione e per l'opportuna manipolazione (*data reduction* e/o *data perturbation*), con criteri più o meno restrittivi a seconda della tipologia del file;
2. per ogni file così prodotto, caricare i dati all'interno del Metadata Editor per associarvi l'opportuna documentazione concettuale e operativa: lo strumento consente di importare quanto già esistente (ad esempio classificazioni) evitando duplicazioni di attività;
3. l'output del Metadata Editor diventa input del CD-Builder per la costruzione delle pagine web utilizzabili dall'utente finale (precedentemente abilitato) per consultare dati e metadati;
4. a questo punto, l'utente finale può utilizzare Nesstar Explorer per navigare e scaricare i dati di interesse secondo i formati più convenienti.

Alcuni dei limiti riscontrati, in particolare:

- l'impossibilità di importare classificazioni dall'esterno quando queste siano già disponibili;
- la presenza di termini ancora in inglese nelle pagine web prodotte dal CD-Builder in una lingua diversa dall'inglese,

potrebbero essere risolti con sviluppo aggiuntivo concordato con il team della World Bank responsabile del *toolkit*.

## **Riferimenti**

BiH Statistical Institutes and ISTAT 2005a – “The BiH Household Budget Survey (methodology)”

BiH Statistical Institutes and ISTAT 2005b – “The BiH Household Budget Survey (final results)” - 2005

Blank, Grant, and Karsten B. Rasmussen 2004 - "The Data Documentation Initiative: The Value and Significance of a Worldwide Standard." *Social Science Computer Review* 22, No. 3 (August 2004): 307-318

DCMI 2006 – “Dublin Core Metadata Element Set, Version 1.1” – DCMI Recommendation (<http://dublincore.org/documents/dces>)

Dupriez, Olivier 2006– “Microdata Management Toolkit Version 1.0 – User’s Guide”– International Household Survey Network Washington D.C. - September 2006 (<http://www.surveynetwork.org/toolkit>)

Franconi Luisa, Seri Giovanni (a cura di) 2004 – “Metodologie e tecniche di tutela della riservatezza nel rilascio di informazione statistica” – Metodi e norme n.20 – ISTAT 2004

ISTAT 2006 - “Certificazione della qualità dell’informazione statistica e accesso ai dati – Materiale per la discussione” – *Ottava conferenza nazionale di statistica (Roma Novembre 2006)*

Miller, Kenneth, and Vardigan, Mary 2005 - "How Initiative Benefits the Research Community - the Data Documentation Initiative." Paper presented at the First International Conference on e-Social Science, Manchester, UK, June 2005

Norwegian Social Science Data Services 1999 - "Providing Global Access to Distributed Data Through Metadata Standardisation: The Parallel Stories of NESSTAR and the DDI." Working Paper No. 10, UN/ECE Work Session on Statistical Metadata, Geneva, Switzerland, September 22-24, 1999

## 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 Loporcario 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](http://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*