

**n. 4/2009**

## **L'Osservatorio Tecnologico per i Software generalizzati (OTS) nel 2008**

*D. Pagliuca, R. Cianchetta, M. Broccoli, T. Buglielli,  
R. Di Giuseppe e D. Zardetto*



---

# DOCUMENTI ISTAT

---

**n. 4/2009**

## **L'Osservatorio Tecnologico per i Software generalizzati (OTS) nel 2008**

*D. Pagliuca(\*), R. Cianchetta(\*), M. Broccoli(\*), T. Buglielli(\*),  
T. Buglielli(\*), R. Di Giuseppe(\*) e D. Zardetto(\*)*

(\*) ISTAT – Direzione centrale programmazione e contabilità

I lavori pubblicati riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto

**Contributi e Documenti Istat 2009**

Istituto Nazionale di Statistica  
Servizio Editoria – Centro stampa  
Via Tuscolana, 1788 - 00173

# L'Osservatorio Tecnologico per i Software generalizzati (OTS) nel 2008

Daniela Pagliuca, ISTAT, Direzione Centrale per le tecnologie e il supporto metodologico  
Raffaella Cianchetta, ISTAT, Direzione Centrale per le tecnologie e il supporto metodologico  
Marco Broccoli, ISTAT, Direzione Centrale per le tecnologie e il supporto metodologico  
Teresa Buglielli, ISTAT, Direzione Centrale per le tecnologie e il supporto metodologico  
Roberto Di Giuseppe, ISTAT, Direzione Centrale per le tecnologie e il supporto metodologico  
Diego Zardetto, ISTAT, Direzione Centrale per le tecnologie e il supporto metodologico

**Sommario:** Il presente lavoro descrive l'Osservatorio Tecnologico per i Software generalizzati (OTS), nato allo scopo di raccogliere e distribuire il software generalizzato utile per la produzione statistica sia in Istat che in ambito Sistan; in particolare illustra le pagine web corrispondenti e introduce le procedure di raccolta e diffusione dei software.

L'OTS è attivo, nella sua attuale configurazione, da fine 2005 nella rete intranet dell'Istituto; da Ottobre 2006 nella rete internet sul sito [www.istat.it](http://www.istat.it) da Ottobre 2006.

**Parole chiave:** software generalizzati, strumenti generalizzati

---

Le collane esistenti presso l'ISTAT - Contributi e Documenti - 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 e del Sistan, o da studiosi esterni.

I lavori pubblicati Contributi Istat vengono fatti circolare allo scopo di suscitare la discussione attorno ai risultati preliminari di ricerca in corso.

I Documenti Istat hanno lo scopo di fornire indicazioni circa le linee, i progressi ed i miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

I lavori pubblicati riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto.



## Indice

1. L'OTS - Osservatorio Tecnologico per i Software generalizzati <i>di Daniela Pagliuca</i>	pag.	9
2. La storia del progetto di realizzazione dell'Osservatorio <i>di Daniela Pagliuca</i>		10
3. Cosa è reperibile attraverso l'OTS e quali sono le aree di interesse <i>di Daniela Pagliuca</i>		11
3.1 La procedura di raccolta e diffusione dei software generalizzati		12
4. La European Union Public Licence (EURL) <i>di Raffaella Cianchetta</i>		14
5. La struttura dell'OTS e le pagine del sito <i>di Raffaella Cianchetta</i>		17
5.1. Descrizione dell'OTS per l'intranet		19
5.2. Confronto tra le pagine internet relative ai software generalizzati per la produzione statistica e le pagine intranet		27
6. I Software dell'OTS nella configurazione attuale – aggiornamento 2008		29
6.1 L'elenco completo dei software attualmente diffusi tramite l'OTS <i>di Daniela Pagliuca</i>		29
6.2 Nuovo software 2008: Mauss R (Multivariate Allocation of Units in Sampling Surveys versione open source) <i>di Teresa Buglielli e Daniela Pagliuca</i>		31
6.3 Nuovo software 2008: Ever (Estimation of Variance by Efficient Replication) <i>di Diego Zardetto</i>		33
6.4 Nuovo software 2008: Concord-Java (CONtrollo e CORrezione dei Dati - versione open source) <i>di Teresa Buglielli</i>		37
6.5 Una nuova release dei software Mauss e Genesees versioni SAS su Server Linux <i>di Roberto Di Giuseppe</i>		41
7. Procedura di notifica della registrazione utenti e rilascio concessione della password – Aggiornamento Dicembre 2008 <i>di Marco Broccoli</i>		42
8. Lo sviluppo dell'Osservatorio tecnologico e delle attività legate ai software generalizzati <i>di Daniela Pagliuca</i>		45
9. Informazioni utili per ricognizioni internazionali <i>di Daniela Pagliuca</i>		46

## Appendici

Appendice 1: Lista conferenze internazionali ISI / IASC	49
Appendice 2: Lista dei Software Register ASC	51
Appendice 3: Summary Of Survey Analysis Software	53
Appendice 4: Link utili per il sito dell'OTS e per informazioni sulla licenza EURL	55





## 1. L'OTS - Osservatorio Tecnologico per i Software generalizzati

Questo documento ha come fine quello di presentare l'Osservatorio Tecnologico per i Software generalizzati (OTS), nato allo scopo di raccogliere e distribuire il software generalizzato utile per la produzione statistica sia in Istat che in ambito Sistan.

L'Istituto Nazionale di Statistica si occupa di produrre dati ed in relazione a ciò si occupa anche degli aspetti metodologici connessi alle diverse fasi di una indagine statistica e dell'individuazione di quegli strumenti software generalizzati utili alla produzione, che implementano i metodi stessi.

Per *software generalizzati per la produzione statistica* si intendono quei sistemi disegnati per garantire le funzionalità produttive delle indagini statistiche, che abbiano le seguenti caratteristiche:

- implementino le metodologie e le tecniche statistiche più avanzate;
- siano utilizzabili senza o con limitata necessità di ulteriore sviluppo software;
- siano dotati di adeguata documentazione e di interfaccia di tipo amichevole, che ne rendano possibile l'utilizzo anche a utenti non esperti da un punto di vista informatico.

In particolare una struttura centralizzata dell'Istituto, occupandosi specificatamente di strumenti generalizzati, ha il compito di renderli disponibili, ovvero di raccogliarli e distribuirli: in collaborazione con altri settori interessati, questa struttura svolge una attività continua e sistematica di ricognizione, analisi, sperimentazione, valutazione e selezione, che permette di reperire il meglio del software esistente, o, in alternativa, di curare lo sviluppo di nuovo software generalizzato.

Per facilitare tale compito utilizza una "vetrina" virtuale, ovvero l'Osservatorio per i software generalizzati", tramite la quale è possibile raccogliere e diffondere in modo controllato le informazioni sui software e, ove possibile, permettere di effettuare il download dei software stessi.

In questo documento viene illustrato l'OTS nella sua ultima versione, aggiornata a Dicembre 2008.

Dopo questo primo paragrafo introduttivo, nel prossimo paragrafo (paragrafo 2) viene introdotta la storia del progetto di realizzazione dell'Osservatorio che ha portato all'attuale configurazione;

nel paragrafo 3 viene illustrato cosa è reperibile nell'Osservatorio;

nel paragrafo 4 viene presentata la licenza EUPL che l'Istat ha adottato per i software *open source*;

nel paragrafo 5 vengono descritte le pagine del sito dell'Osservatorio, sia nella versione internet che intranet;

nel 6 vengono presentati i software che attualmente sono di interesse dell'Osservatorio e in particolare - oltre alla lista completa di quelli reperibili - vengono evidenziati i nuovi rilasci del 2008 ;

nel paragrafo 7 viene descritta la procedura attualmente utilizzata per la registrazione delle utenze;

nel paragrafo 8 vengono illustrati alcuni passi previsti per lo sviluppo ulteriore dell'OTS; infine nel paragrafo 9 vengono presentati alcuni istituti e associazioni, e i relativi siti internazionali tramite i quali è possibile reperire informazioni interessanti circa gli strumenti generalizzati per la produzione statistica.

Le appendici rappresentano utili approfondimenti legati alle informazioni reperibili a livello internazionale e evidenziano i link connessi.

## 2. La storia del progetto di realizzazione dell'Osservatorio

Nell'attivare l'OTS si è partiti da quanto realizzato nell'ambito di un progetto rientrante nei piani dell'Istituto del 2002, specifico per la manutenzione e assistenza dei software generalizzati: in quell'anno infatti tale progetto ha portato alla messa in opera di un sito, collocato nella intranet dell'Istituto e raggiungibile dunque dalla sola utenza Istat, utile a diffondere i prodotti sviluppati all'interno della struttura preposta allo sviluppo dei software generalizzati.

Con la creazione di quel primo spazio virtuale sono state attivate una serie di procedure standard per permettere alla utenza dei software generalizzati di effettuare il download degli strumenti di interesse, prevedendo una registrazione delle utenze (si richiedeva l'invio di una email per ottenere una password di utilizzo) allo scopo di garantire l'adeguata assistenza.

Il successivo passaggio ha poi portato all'ampliamento delle finalità attribuite a questo sito intranet: si è pensato ad uno spazio virtuale, sempre interno all'Istituto, che garantisse la reperibilità dei prodotti e delle informazioni di software generalizzati, non solo sviluppati all'interno della struttura centrale, ma adeguatamente selezionati per essere messi a disposizione della utenza Istat.

Successivamente, grazie alla ottima ricezione da parte dell'utenza, si è poi deciso di attivare tale sito anche nella rete internet per rivolgerlo alla utenza esterna ed in particolare al Sistan.

Per dare qualche informazione sui tempi, è utile evidenziare che l'occasione favorevole per l'estensione delle pagine dell'unità si è presentata con la revisione generale delle pagine intranet dell'Istituto nel corso del 2005: è stata dunque creata la prima versione dell'Osservatorio Tecnologico per i Software generalizzati (OTS) sulla intranet e successivamente sul sito dell'Istituto ([www.istat.it](http://www.istat.it)), questa ultima depurata dei soli prodotti per i quali esiste una licenza che ne autorizza l'utilizzo solo all'interno dell'Istat stesso, tale per cui non ne è concessa la diffusione all'esterno.

Questo passaggio ha di fatto definito un vero e proprio "Osservatorio" dei software generalizzati.

In effetti i cambiamenti avvenuti hanno portato a definire quale obiettivo principale quello di offrire all'utenza, sia interna che esterna, un "punto di raccolta" aggiornato e controllato in modo centralizzato e per far ciò è divenuto fondamentale occuparsi di verificare costantemente il software statistico disponibile, prodotto anche all'esterno dell'Istituto, nell'ottica di garantire una offerta sempre più completa ed aggiornata.

Sia per l'attività di implementazione che per la selezione del software esistente, è divenuto anche fondamentale considerare gli strumenti *open source*, più facilmente reperibili e divulgabili in realtà territoriali differenti: nel diffondere informazioni e software all'esterno, soprattutto in ambito Sistan, è più facile raggiungere quegli utenti che rappresentano piccole realtà sul territorio (dotati dunque di strumenti informatici spesso ridotti o comunque diversi da quelli adoperati in Istat) evitando l'uso di software proprietario (SAS principalmente) e evitando di proporre soluzioni dispendiose e difficilmente modificabili in modo autonomo.

Nella rete intranet dell'Istituto l'OTS, nella configurazione attuale, è attivo da fine 2005; è attivo sul sito [www.istat.it](http://www.istat.it) da Ottobre 2006.

### 3. Cosa è reperibile attraverso l'OTS e quali sono le aree di interesse

Tramite l'OTS l'Istat rende disponibili, per ogni fase dell'indagine statistica, dal disegno del campione alla diffusione e all'analisi dei dati, alcuni software generalizzati per la produzione statistica.

L'Osservatorio mette a disposizione :

- le informazioni sui software generalizzati
- la documentazione collegata
- i rispettivi referenti in Istat
- il download dei documenti e dei software disponibili.

Per ciascun software vengono descritte le principali informazioni; per i software sviluppati dall'Istituto, e ove possibile, è in effetti anche disponibile il *download*; per gli altri casi, invece, ne vengono indicate le modalità di reperimento;

è possibile trovare i giusti contatti per ciascuno dei software generalizzati elencati. Per ogni software è infatti indicato il referente in Istat: struttura di riferimento e/o persona che cura l'assistenza tecnica o metodologica;

ove possibile i software sono dotati di una opportuna documentazione (help in linea e manuale d'uso comprensivo di linee guida metodologiche) che illustra ed esemplifica le potenzialità delle applicazioni.

Nel prossimo paragrafo 3.1 si potranno leggere informazioni più dettagliate a riguardo, in quanto viene presentata la procedura di raccolta attivata per inserire un nuovo software nell'Osservatorio.

Le aree principali di interesse dell'OTS sono le seguenti:

#### 1. CAMPIONE

- Disegno
- Selezione
- Calcolo dei coefficienti di espansione
- Calcolo degli errori campionari

#### 2. ACQUISIZIONE DATI

- Codifica automatica
- Gestione acquisizione dati con più tecniche di rilevazione

#### 3. CONTROLLO E CORREZIONE

- Controllo e correzione variabili qualitative
- Controllo e correzione variabili quantitative
- Controllo e correzione variabili unità complesse

#### 4. DIFFUSIONE DEI DATI

#### 5. TUTELA RISERVATEZZA

- Tutela dati elementari
- Tutela dati aggregati

#### 6. ANALISI DEI DATI/RECORD LINKAGE

### 3.1 La procedura di raccolta e diffusione dei software generalizzati

Per acquisire le informazioni, è necessaria una attività continua e sistematica di ricognizione per reperire le informazioni sul software esistente.

Tale ricognizione deve essere effettuata sia a livello nazionale che internazionale.

Dal momento che è fondamentale la collaborazione con gli altri settori interessati dell'Istituto, la prima fonte di acquisizione delle informazioni è certamente interna, proveniente in particolare da coloro che hanno sviluppato o utilizzano gli strumenti stessi (spesso sono ricercatori o informatici che lavorano sia presso strutture in Istat specificatamente delegate ad occuparsi dei metodi, che presso le indagini stesse).

Altra fonte importante sono i siti internazionali e le conferenze.

Tra i convegni, ad esempio, si possono segnalare quelli più rivolti alla tecnologia, in particolare l' NTTS "New Techniques and Technologies for Statistics" (oggi congiunto all'ETK "Exchange of Technology and Know-how"), così come i diversi convegni organizzati dall'ASC, o ancor più dallo IASC e dall'ASA.

Oggi molte informazioni sono facilmente consultabili tramite internet e, per dettagli circa il reperimento delle informazioni a livello internazionale, si può leggere il successivo paragrafo 9 (e le Appendici 1,2,3), dove vengono riportate le indicazioni circa i siti web dei diversi istituti e delle associazioni.

In Istat, per raccogliere le informazioni utili da diffondere tramite l'Osservatorio, è stata definita una vera e propria procedura di raccolta, da attivare per inserire un nuovo software reputato idoneo che viene descritta nel seguito.

#### La procedura di raccolta dei software generalizzati

Si richiedono le specifiche informazioni che seguono:

##### **Informazioni generali**

---

- Sintesi informativa del software
- Eventuali link di collegamento (sia esterni che interni all'Istituto)

##### **Contatti**

---

- Evidenziare, ove possibile, la struttura di riferimento e la/le persona che rappresentano un contatto diretto utile a dare indicazioni sul prodotto, preferibilmente interno all'Istat (è possibile evidenziare solo la struttura, un referente unico, un referente metodologico e un referente informatico se è possibile specificarli etc.)

##### **Documentazione collegata al software**

---

- Inviare la bibliografia da evidenziare, eventuali articoli o lavori strettamente connessi al software (sia Istat che non), se possibile in formato elettronico per permettere di effettuare il download

##### **Download del software e del manuale utente allegato**

---

Ove possibile rendere disponibile il software.

Per far ciò è utile indicare tutte le giuste informazioni, quali:

- Requisiti hw e sw
- eventuale procedura di assistenza

Il software può essere inviato tramite un file compresso, dettagliando la procedura di espansione e di installazione per l'utenza

Le informazioni devono essere chiare e preferibilmente i file da inviare possono attenersi a quanto nel seguito consigliato

Occorre inviare :

- *Informazioni.doc* : un file word (o text) contenete le Informazioni (se ci sono link interni da inserire evidenziarli tra parentesi dopo la parola che rappresenta il link stesso)
- *i contatti* (nomi strutture, persone e indirizzi email)
- *elenco.doc*: un file word con l'elenco dei documenti (*elenco.doc*) che si desidera distribuire. Ciascuna voce dell'elenco è seguita dal nome tra parentesi del file.doc o .pdf che contiene il lavoro (o anche un file compresso unico, se si vuole distribuire assieme più documenti).
- *i file .doc e .pdf o compressi* relativi alla documentazione di cui sopra (*lavoro1.pdf lavoro2.pdf etc*)

Tutte le informazioni raccolte vengono inserite nell'Osservatorio

Per quanto concerne la diffusione dei software generalizzati, una volta individuati i sistemi candidati a divenire lo standard, sviluppati o esterni, vengono posti nelle pagine internet ed intranet a disposizione.

Il primo step consiste generalmente nell'inserirle nelle pagine dell'OTS intranet e attendere eventuali ritorni.

Dopo un breve periodo si inviano tutte le informazioni alla struttura in Istat che si occupa di comunicazione esterna che – dopo l'eventuale revisione dei contenuti e del layout di presentazione – provvede ad inserirle nel sito dell'Istituto.

Come dettagliato nel successivo paragrafo 5.1 (in particolare si vedano figure 7 e 8), nella versione dell'OTS per la intranet vengono in particolare evidenziati quali software sono open source e l'utente può orientare la propria ricerca direttamente su quelli.

E' utile evidenziare che per la divulgazione dei software open source, è stata deciso di adottare la licenza EUPL che l'utente deve sottoscrivere. Tale licenza viene presentata in dettaglio nel prossimo paragrafo 4.

## 4. La European Union Public Licence (EURL).

La European Union Public Licence (EURL) è una licenza Open Source (FLOSS<sup>1</sup>) ed in quanto tale, si conforma alle libertà<sup>2</sup> garantite ai licenziatari espresse dalla FSF (Free Software Foundation) ed ai principi definiti dalla fondazione OSI (Open Source Initiative) nella Open Source Definition<sup>3</sup>.

La creazione della licenza EURL da parte della Commissione Europea, si è resa indispensabile a seguito della necessità di distribuire alle amministrazioni pubbliche europee, alcuni software prodotti nell'ambito dei programmi IDA<sup>4</sup> e IDABC<sup>5</sup>. Nell'intenzione della Commissione, infatti, tali applicazioni sarebbero dovute rimanere di tipo open e non proprietario.

L'esigenza di creare tale ulteriore licenza è sorta a seguito di un apposito studio, nell'ambito del programma IDABC, dal quale è emersa anche la necessità che essa risponda a determinati requisiti non presenti fino a quel momento in altre licenze:

- peculiarità della legge della Comunità Europea e degli Stati membri in merito alla terminologia applicabile ai diritti di copyright, alla puntuale determinazione delle limitazioni di responsabilità, delle garanzie e della giurisdizione competente;
- medesimo valore legale della licenza in tutte le lingue ufficiali dell'Unione Europea.

L'esistenza di problemi di incompatibilità, da un lato tra la terminologia e la legislazione europea e dall'altro tra le licenze Open Source esistenti, per la maggior parte statunitensi, poneva seri dubbi in merito alla loro applicazione nel contesto europeo.

Lo scopo della Commissione Europea, infatti, era di poter distribuire il proprio software sotto una licenza conforme alla legislazione comunitaria, potenziando l'interoperabilità legale nell'ambito del software open source mediante l'utilizzo di una licenza comune rispetto al proliferare di distinte licenze nazionali.

La prima bozza è stata pubblicata nel giugno 2005 e dopo una serie di interventi sul testo iniziale, apportati dalla comunità di utenti e di sviluppatori, si è giunti alla seconda bozza.

Il 9 gennaio 2007, dall'elaborazione di quest'ultima, la Commissione Europea è giunta, tramite la "Decision C(2006) 7108", all'approvazione della versione definitiva (v.1.0) inizialmente in tre lingue: inglese, francese e tedesco. Nel febbraio 2007, l'EURL è stata applicata al rilascio dei seguenti software: CIRCA e IPM sviluppati nel contesto del programma IDABC.

All'approvazione ha seguito il processo di traduzione, di coordinamento e di controllo della qualità della traduzione, articolato nelle fasi elencate di seguito:

- a) traduzione, da parte dei traduttori della Commissione Europea, nelle lingue ufficiali (ad esclusione dell'irlandese) dell'Unione Europea;
- b) nomina del partner contrattuale OSOR<sup>6</sup> in qualità di coordinatore delle attività e di garante del controllo di qualità della revisione;

---

<sup>1</sup>FLOSS (Free/Libre/Open Source Software).

<sup>2</sup>Libertà di eseguire il software, analizzarlo, ridistribuirne copie, migliorarlo e distribuirne le modifiche.

<sup>3</sup> <http://www.opensource.org/docs/osd>

<sup>4</sup> IDA - (Interchange of Data between Administrations) - Programma della Commissione Europea per gli anni 1995-2004, finalizzato alla promozione dell'interoperabilità tra le amministrazioni tramite l'accesso ai network intraeuropei.

<sup>5</sup> IDABC (Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens) - Programma della Commissione Europea per gli anni 2005-2009, gestito dal 1 gennaio 2007, all'interno del Directorate General for Informatics (DIGIT), da un'unità che ha la stessa denominazione del programma.

<sup>6</sup> Il Consorzio guidato da Unisys (Belgium) e costituito da Universidad Rey Juan Carlos, UNU-Merit, Gopa-Cartermill, si è aggiudicato il contratto per la costituzione di OSOR (Open Source Observatory and Repository): portale web avente la funzione di osservatorio e repository, del codice (sorgente ed oggetto) e della documentazione, delle applicazioni open source delle pubbliche amministrazioni della Comunità Europea. I servizi disponibili, riguardano anche l'accesso all'ambiente di sviluppo e la diffusione di casi studio e notizie internazionali su tematiche relative all'open source. Il nuovo portale rimpiazza ed amplia la precedente sezione, attiva dal 2004, del sito dell'IDABC relativa all'OSO (Open Source Observatory).

- c) revisione, correzione e commento delle traduzioni sulla base principalmente della versione inglese ed in via subordinata delle altre due già approvate, da parte di un gruppo selezionato di esperti in diritti di proprietà intellettuale nella misura di 2 per ciascuna lingua;
- d) riesame della versione rilasciata dagli esperti, da parte del gruppo dei coordinatori e della Commissione e loro ulteriore coinvolgimento nel caso di richiesta di chiarimenti;
- e) facoltà della Commissione Europea di interpellare un altro esperto prima di convalidare il testo finale relativo a tutte le versioni.

Il 9 gennaio 2008, la Commissione Europea ha convalidato la licenza per ulteriori 19 linguaggi, totalizzando complessivamente 22 versioni<sup>7</sup> nelle seguenti lingue: bulgaro, ceco, danese, olandese, estone, finlandese, greco, ungherese, italiano, lituano, lettone, maltese, polacco, portoghese, romeno, slovacco, sloveno, spagnolo, svedese, inglese, francese e tedesco.

Questa licenza ha stabilito alcuni primati:

- a) prima licenza open source europea sviluppata e adottata da un ente pubblico;
- b) validità legale in tutte le lingue ufficiali dell'Unione Europea sulla base del principio della diversità linguistica, come sancito dall'art. 22 della Carta dei Diritti Fondamentali dell'Unione Europea<sup>8</sup>;
- c) terminologia legislativa consona ai requisiti di legge degli stati appartenenti all'EU;
- d) compatibilità con le licenze open source più utilizzate.

In merito a quest'ultimo punto, una delle clausole contenute nella licenza concerne proprio la sua compatibilità con le seguenti licenze:

- la GPL v. 2;
- la OSL v. 2.1 e/o 3.0;
- la Eclipse Public Licence v. 1.0;
- la CeCill v. 2.0;
- la Common Public Licence v. 1.0.

L'altra clausola riguarda, invece, il copyleft: la redistribuzione del codice originale e/o la distribuzione di opere derivate, deve avvenire tramite la medesima licenza o una licenza che dia le stesse garanzie (anche essa copyleft).

Nel caso, però, di redistribuzione del codice originario e/o di distribuzione di opere derivate che prevedono la combinazione tra software licenziato con EUPL con software che sono licenziati tramite l'uso di una delle licenze copyleft compatibili sopra riportate, allora, la distribuzione del codice avverrà sulla base della licenza compatibile. Questo è il cosiddetto "effetto virale" che accompagna alcune licenze copyleft.

Non tutte le licenze FLOSS, quale ad es. la BSD<sup>9</sup> e la MIT<sup>10</sup>, hanno la limitazione del copyleft, si parla in tal caso di licenza permissiva, consentendone liberamente la redistribuzione del codice originario o la distribuzione di opere derivate mediante qualunque licenza. Nel caso, pertanto, di distribuzione di opere derivate che prevedono la combinazione tra software licenziato sotto EUPL e software licenziato sotto licenze permissive, la distribuzione del codice avverrà sotto la licenza EUPL.

Il testo della licenza, che consta di 15 articoli, essendo di tipo molto generale e non esistendo una clausola che ne limiti l'utilizzo, può essere utilizzato anche da licenzianti non istituzionali.

L'art. 1 riguarda la definizione della terminologia utilizzata nella licenza. Essa definisce diritti ed obblighi per il licenziatario, rispettivamente all'art. 2 ed all'art. 5, nonché le garanzie da parte di autori e contributori (art. 6). Anche tale licenza, come per tutte le altre, siano esse Open Source che proprietarie, esclude garanzie in merito alla qualità del software (art. 7) e responsabilità circa i danni di qualsiasi natura causati dal software (art. 8). La licenza, comunque, contempla dei casi nei quali emerge la responsabilità dell'autore: nel caso di dolo e nel caso di danni diretti a persone fisiche. L'EUPL in

---

<sup>7</sup> Tutte le 22 versioni, accompagnate da un preambolo che non è parte integrante della licenza, sono disponibili al seguente URL: <http://ec.europa.eu/idabc/en/document/7330>

<sup>8</sup> Pubblicata in GUCE 2000/C 364/01 il 18 dicembre 2000.

<sup>9</sup> Berkeley Software Distribution.

<sup>10</sup> Licenza creata dal Massachusetts Institute of Technology.

quanto documento ufficiale non è modificabile, vi è però la facoltà di stipulare pattuizioni aggiuntive (art. 9). Essa sancisce le modalità di comunicazione del codice sorgente (art. 3) e quelle di accettazione della licenza (art. 10). Non vi sono limitazioni al diritto d'autore (art. 4), per poter rilasciare il software sotto tale licenza si rende indispensabile il possesso dei diritti sullo stesso. L'art. 11 riguarda le informazioni prescritte dalla normativa, volte ad informare il pubblico nel caso di distribuzione o comunicazione con strumenti di comunicazione elettronica.

L'art. 12 stabilisce la cessazione della licenza ed i relativi diritti, l'art. 14 il foro competente nel caso di controversie, l'art. 13 riguarda le clausole nulle o non valide e le nuove versioni della licenza, infine, l'art. 15 la disciplina di legge applicabile.

Il Consorzio OSOR ha predisposto delle linee guida dell'EUPL: "Guidelines for users and developers"<sup>11</sup> aventi l'obiettivo di divulgare:

- le modalità di utilizzo del software distribuito sotto tale licenza;
- come sviluppare o integrare il software con tale licenza;
- le modalità d'uso della licenza per distribuire un proprio prodotto software.

L'EUPL può essere considerata come uno strumento per realizzare gli obiettivi previsti dall'Agenda di Lisbona<sup>12</sup> e dal progetto i2010Initiative (European Information Society in 2010) riguardante la ricerca e lo sviluppo nel settore ICT.

---

<sup>11</sup> "Guidelines for users and developers" - <http://osor.eu/eupl/eupl-guideline>

<sup>12</sup> Programma di riforme disposto dal Consiglio europeo ed approvato a Lisbona (23-24/03/2000).

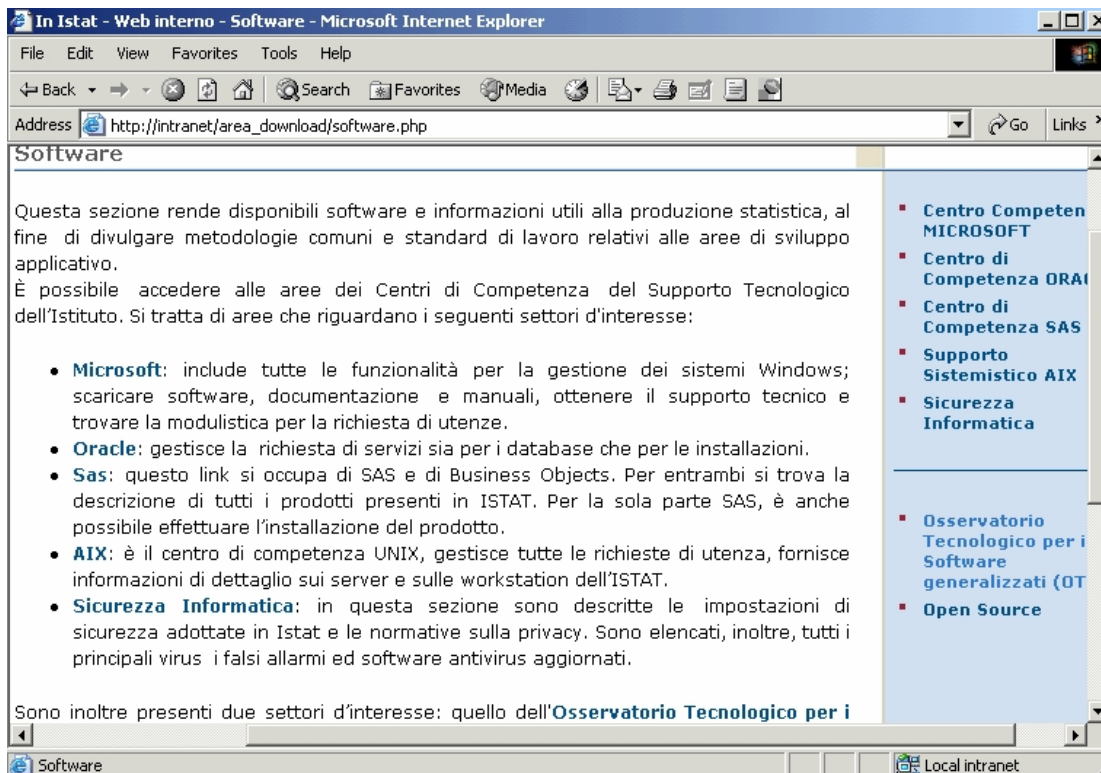


## 5. La struttura dell'OTS e le pagine del sito

L'Osservatorio Tecnologico dei Software generalizzati per la produzione statistica (da ora in poi OTS) è stato suddiviso in due prodotti: uno presente all'interno dell'intranet e l'altro all'interno del portale dell'Istituto.

Il primo, è rintracciabile accedendo al seguente link:

[http://intranet.istat.it/doutdes/SW\\_statistico\\_1/repository/index.html](http://intranet.istat.it/doutdes/SW_statistico_1/repository/index.html) oppure osservando il seguente percorso: partendo dalla homepage dell'intranet, all'interno dell'“area download” si seleziona “Software” e poi o si clicca sul link dell'Osservatorio Tecnologico per i Software generalizzati all'interno della stessa pagina o si sceglie sulla sezione di destra il link relativo all'OTS (Fig. 1).



**Figura 1:** Sezione “Software” dell’“Area download” dell’intranet: link dell’OTS

A differenza del primo, all'interno del portale dell'Istituto si è preferito evidenziare i prodotti software che sono esposti nella sezione Strumenti/Metodologie e Software/Software raggiungibile anche tramite il seguente link: <http://www.istat.it/strumenti/metodi/software/>. All'interno della pagina dedicata ai software generalizzati per la produzione statistica (Fig. 2), vi è un riferimento all'OTS ed un link ad esso associato che permette l'apertura di una pagina relativa alla descrizione dei contenuti informativi dell'OTS (Fig. 3).



Figura 2: Pagina del portale d'Istituto relativa ai “Software generalizzati per la produzione statistica”

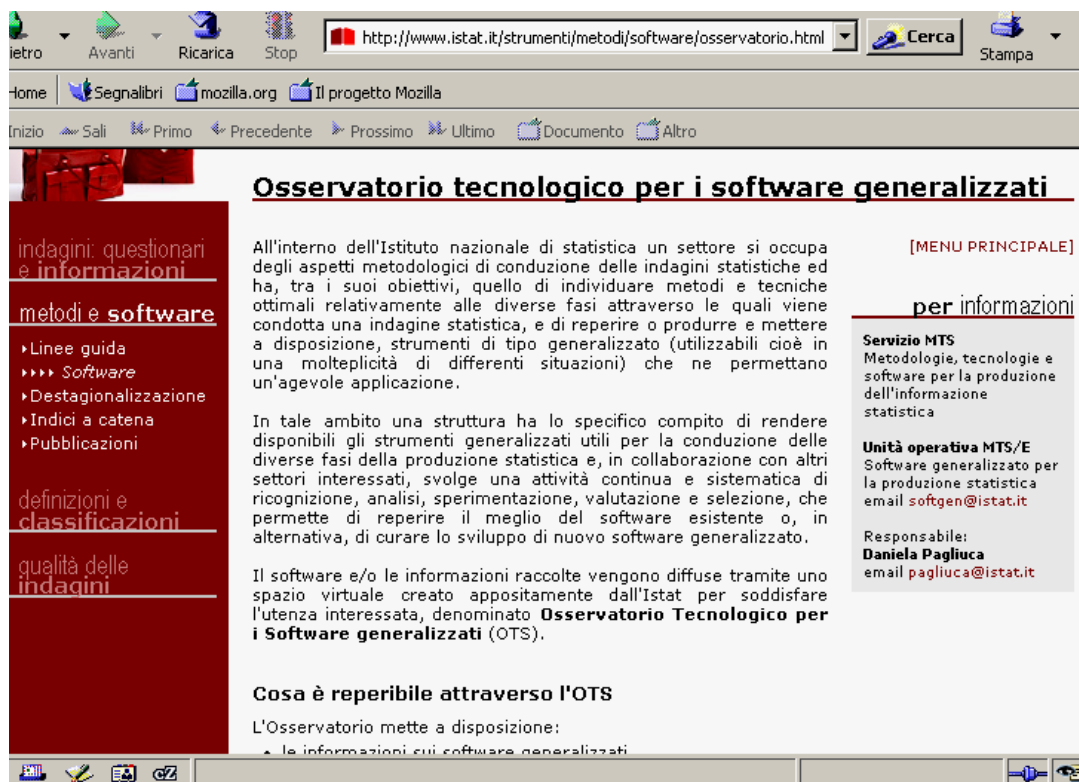


Figura 3: Pagina del portale d'Istituto relativa all'“Osservatorio tecnologico dei software generalizzati”

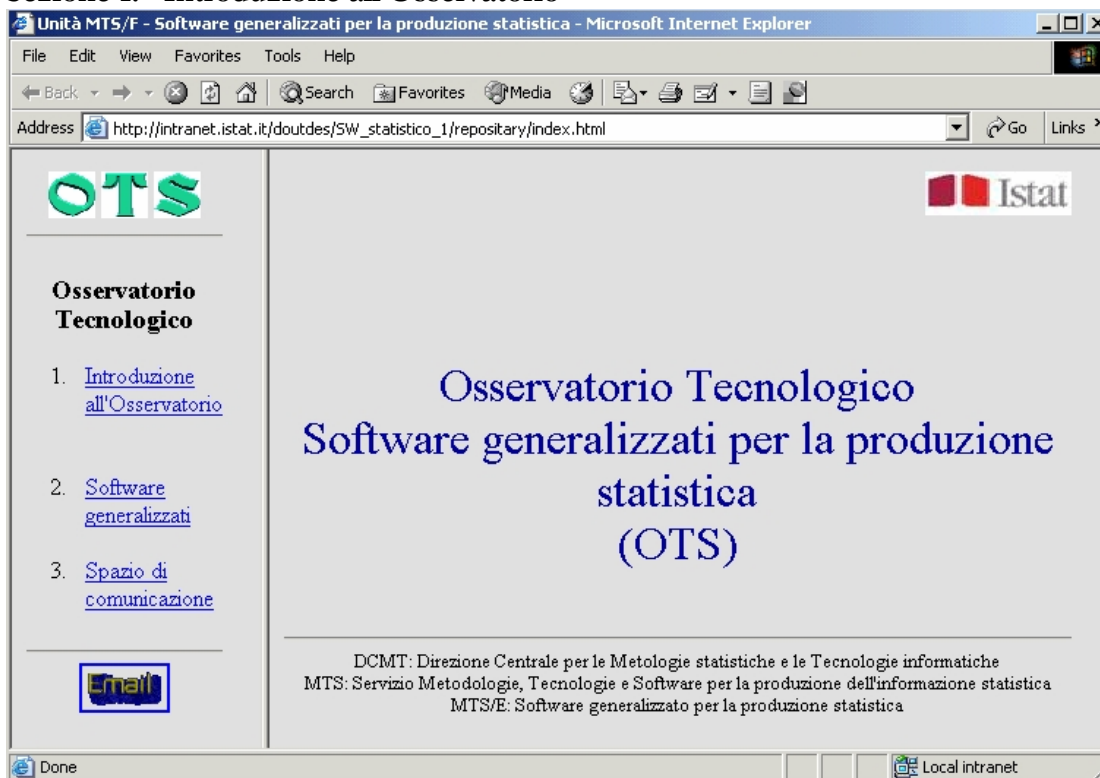
## 5.1 Descrizione dell'OTS per l'intranet

Nella homepage dell'OTS per l'intranet (Fig. 4) possiamo distinguere, nel frame di sinistra, i link alle seguenti sezioni:

- 1) Introduzione all'Osservatorio;
- 2) Software generalizzati;
- 3) Spazio di comunicazione.

Inoltre, è prevista la possibilità di inviare l'email all'U.O. MTS/E<sup>13</sup> cliccando sulla scritta EMail.

### Sezione 1: “Introduzione all'Osservatorio”



**Figura 4:** Homepage dell'OTS

Cliccando sul link della sezione 1 - “Introduzione all'Osservatorio”- si apre la pagina relativa alla descrizione dell'OTS (Fig. 5) individuandone la finalità, la fruibilità, le aree di interesse e la collocazione all'interno dell'Istituto.

---

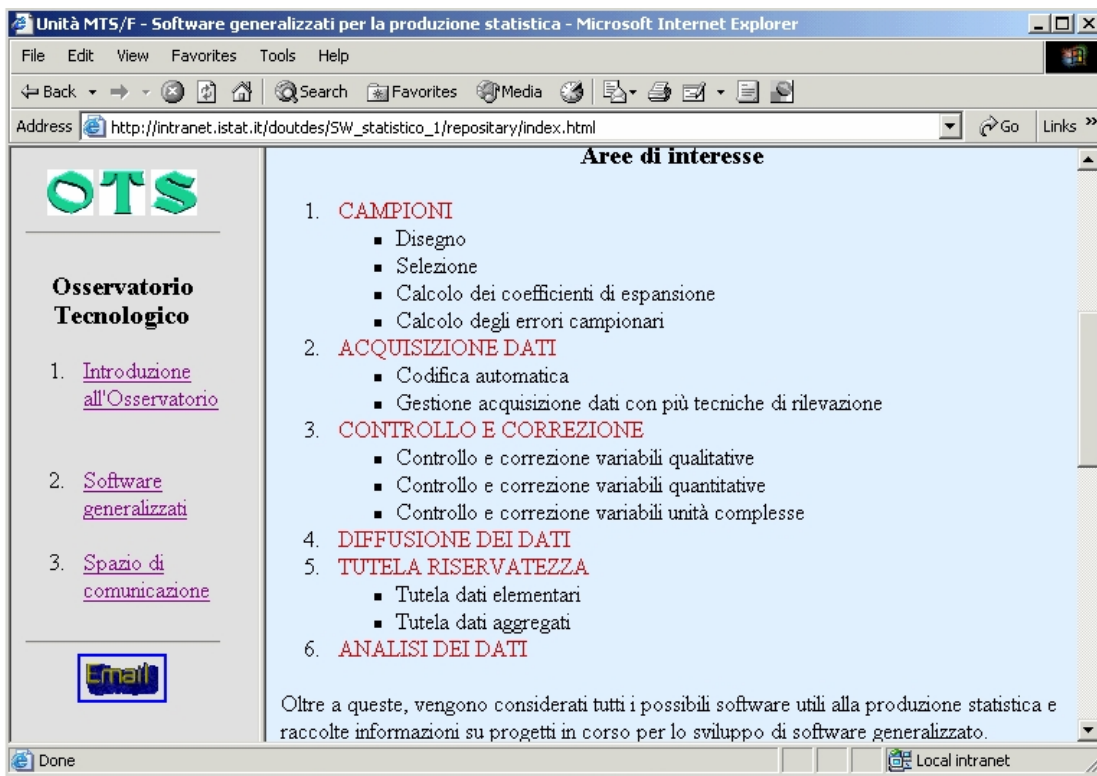
<sup>13</sup> Unità Operativa “Software generalizzato per la produzione statistica” appartenente al Servizio MTS (Servizio metodologie, tecnologie e software per la produzione dell'informazione Statistica)



**Figura 5:** Sezione 1 “Introduzione all’Osservatorio”

L’obiettivo dell’Osservatorio è quello di fungere da unico “centro di raccolta” disponibile, nella fattispecie, all’utenza interna, per il download dei software generalizzati per la produzione statistica sia di produzione interna che licenziati (ove possibile) corredati dai relativi manuali d’uso e/o dai documenti disponibili e per il reperimento di informazioni di carattere generale riguardanti i software, i requisiti hardware e software per la loro installazione, i referenti interni e la bibliografia d’interesse per le connesse funzionalità.

Nella schermata che segue (Fig. 6), sono visualizzate le aree di interesse che non debbono essere considerate in senso limitativo, in quanto si può inserire qualunque software di tipo generalizzato vantaggioso per la produzione statistica.



**Figura 6:** Sezione 1 “Introduzione all’Osservatorio” - Aree d’interesse

In merito alla collocazione dell’Osservatorio, esso è stato creato ed è gestito all’interno dell’U.O. MTS/E.

### **Sezione 2: “Software generalizzati”**

Nella prima pagina della sezione 2 “Software generalizzati”(Fig. 7) è presentato un elenco delle aree all’interno delle quali sono stati sviluppati uno o più software. Sotto di esso, appaiono tre ulteriori sezioni: “Open Source Software (OSS)”, “Progetti in corso” e “Nuovi rilasci software”. Tramite la prima, si accede ad una tabella nella quale sono elencati i software open source, le associate aree di interesse ed i linguaggi nei quali sono stati sviluppati (Fig. 8). In questo modo, l’utente può avere una panoramica dei sw open source disponibili, senza cercarli all’interno delle varie aree.

Nella seconda, sono riportati i nomi dei progetti ancora in fase di implementazione e le informazioni al momento disponibili. Nella terza, sono segnalati i rilasci di nuovi software e delle ultime versioni, pertanto, è consigliabile agli utenti un accesso ricorrente (Fig. 9).

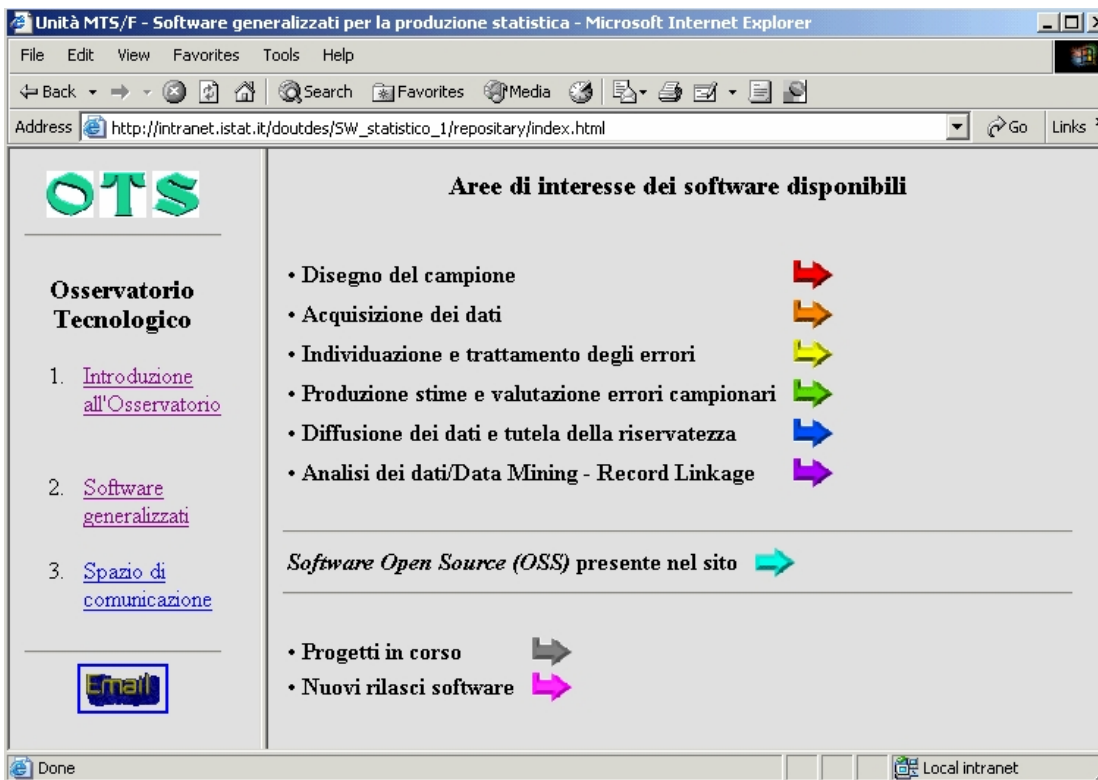


Figura 7: Sezione 2 “Software generalizzati” - elenco delle aree per le quali sono presenti software generalizzati

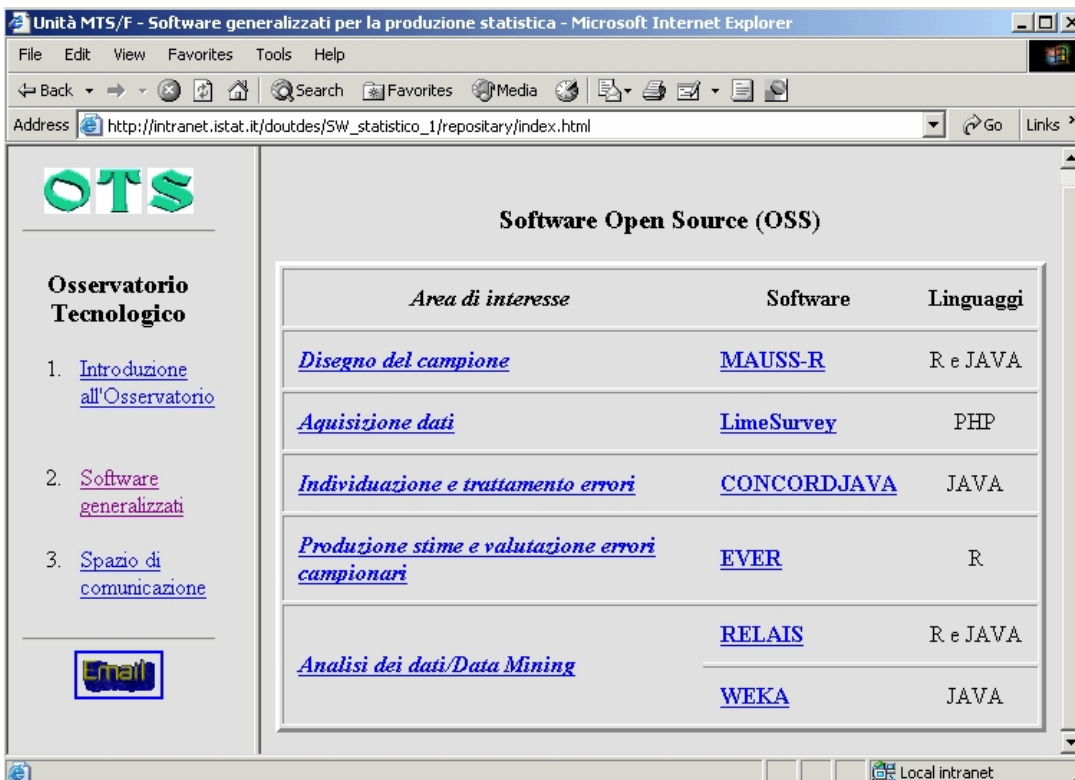
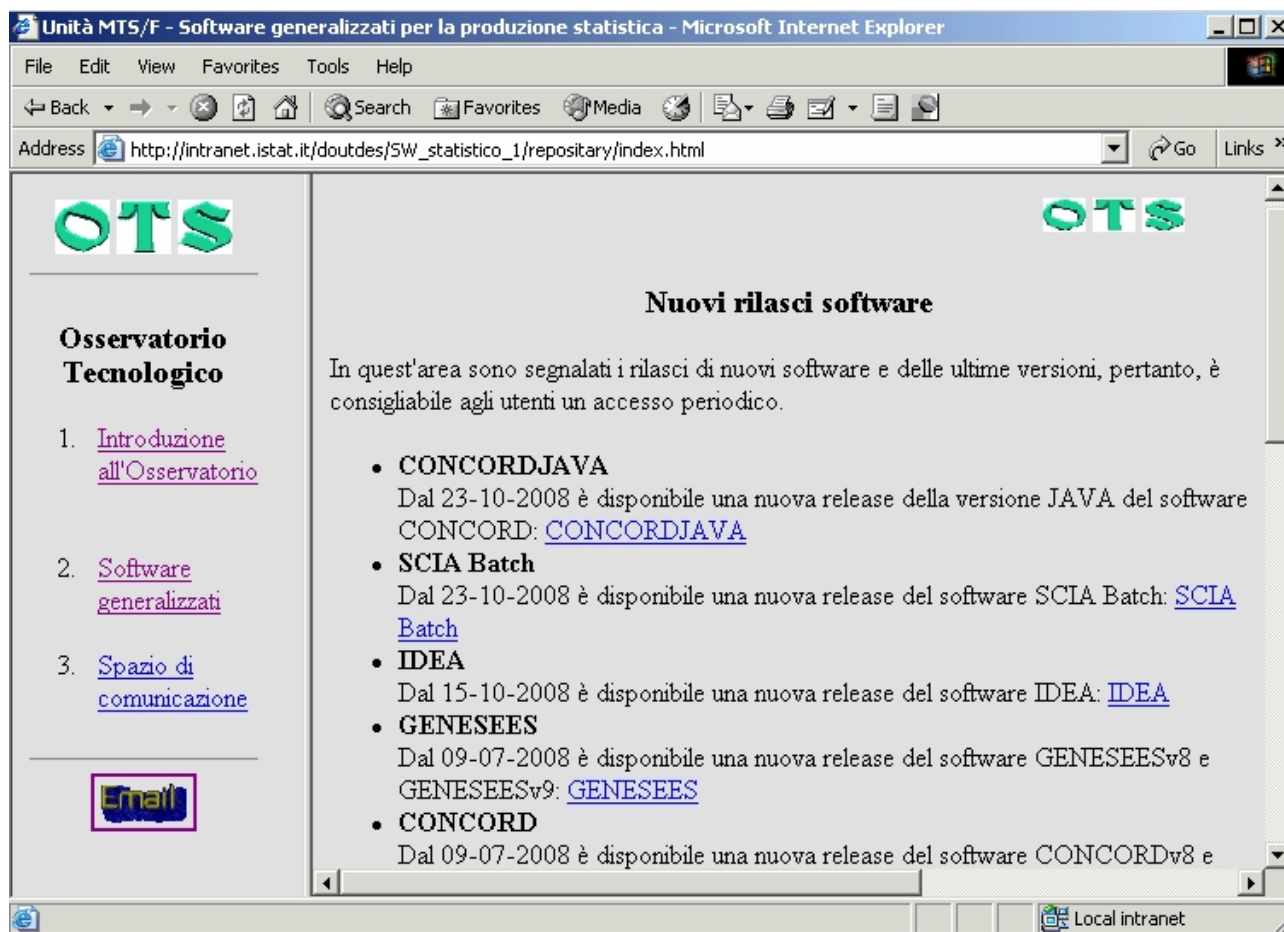


Figura 8: Sezione relativa al Software Open Source presente nel sito



**Figura 9:** Sezione relativa ai nuovi rilasci del software

Per ciascuna area sono presentati uno o più software (vedasi Fig. 10). Per ciascuno di essi, sono state individuati i link alle seguenti voci<sup>14</sup>:

- a) Informazioni generali sul software;
- b) Documentazione disponibile;
- c) Referenti, contatti in Istat;
- d) Download del software.

All'interno della pagina principale di ciascun prodotto, è presente una nota con la quale si invitano i referenti e gli utilizzatori del software ad inoltrare via email all'U.O. MTS/E (è presente il link ad essa associato) suggerimenti e/o variazioni relativamente ad informazioni inesatte o insufficienti, documentazione sul software, nominativi di altro personale interno competente sul prodotto per possibili contatti ed altro prodotto d'interesse per l'Istituto.

Nell'ambito dell'area "Produzione stime e valutazione errori campionari", viene esaminata ciascuna voce prendendo in esame, a titolo esemplificativo, il software GENESEES (vedasi Fig. 11).

<sup>14</sup> Per alcuni software, come ad es. GENESEES, in aggiunta è prevista anche la voce relativa alla disponibilità del sw su server in ambiente LINUX

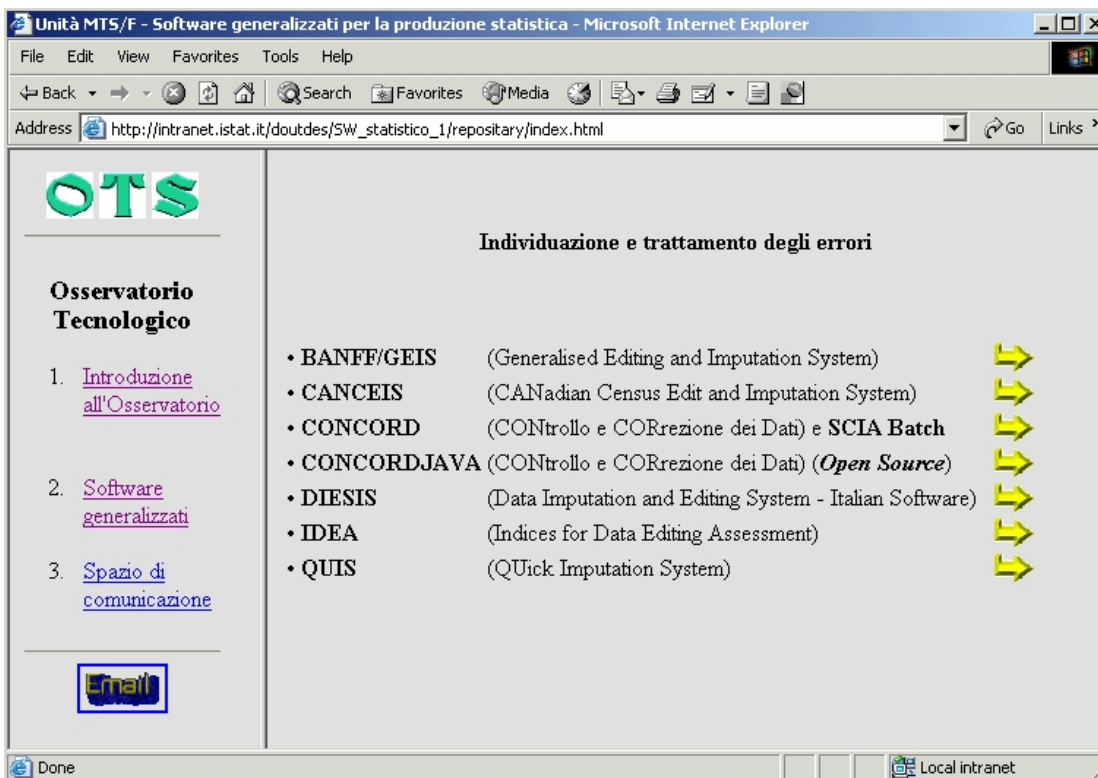


Figura 10: Sezione 2 “Software generalizzati” - elenco dei software dell’area “Individuazione e trattamento degli errori”

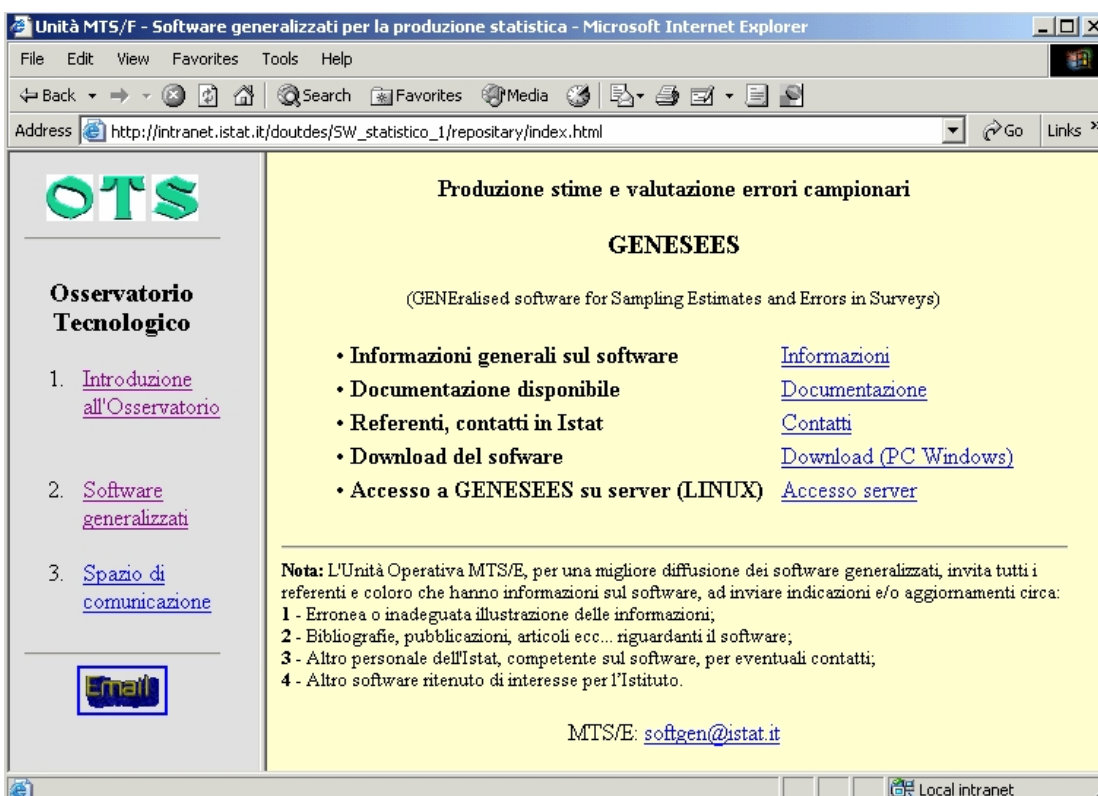


Figura 11: Sezione 2 “Software generalizzati” - pagina principale del software GENESEES

Nella voce “Informazioni generali sul software”, sono disponibili per tutti i software notizie generiche sul prodotto (gli sviluppatori, le funzioni disponibili e qualunque altra informazione d’interesse per gli utilizzatori).

Nella voce “Documentazione disponibile”, per il software GENESEES si riportano la bibliografia relativa al calcolo delle stime e degli errori campionari e la documentazione relativa a versioni



precedenti GENESEES v3.0, nonché i link ai manuali utenti ed alla documentazione relativa all'ultima versione v3.0.

In generale, per la maggior parte dei prodotti sviluppati in Istituto, nella voce "Documentazione disponibile", si riportano la bibliografia relativa alle argomentazioni oggetto del prodotto e/o alle funzioni sviluppate dal software, alla documentazione di corredo e ai manuali utente relativi ad un'eventuale versione precedente ed alla versione corrente. Nel caso di software sviluppati all'esterno, si riporta la bibliografia nonché eventuali link ai siti e/o alla documentazione di riferimento. In entrambi i casi, se si tratta di documentazione pubblicata dall'Istituto, vi è associato il relativo link.

In merito alla voce "Referenti, contatti in Istat", per tutti i software si effettua una distinzione tra referenti per problematiche metodologiche, per l'assistenza informatica e per informazioni generali nel caso in cui non coincidano con la stessa persona. Per ognuno di essi si riporta il nome e l'email, la denominazione della struttura di riferimento (Unità Operativa).

In merito alla voce "Download del software", tranne qualche eccezione<sup>15</sup>, sono descritti i requisiti hardware e software per poter effettuare il download e le connesse istruzioni operative. Nel caso di software open source sviluppato in Istituto, l'utente avrà accesso ad una maschera di identificazione e solo a seguito della registrazione avrà accesso alla pagina che ne permette il download.

Per alcuni prodotti con interfaccia SAS sviluppati in Istituto (MAUSS, GENESEES e CONCORD), viene indicata la necessità di una password per l'utilizzo primario del software fornita automaticamente tramite posta elettronica a seguito della compilazione della schermata di identificazione utente (cfr. par. 7).

In particolare, per il software GENESEES, si riportano nuovamente nella sezione di download sia i link per il download dei manuali utenti e della documentazione del sw, sia quelli relativi alla documentazione sui problemi ricorrenti nell'utilizzo del software succitato relativamente a due funzioni sviluppate dal software.

### **Sezione 3: "Spazio di comunicazione"**

Nella Sezione 3 - "Spazio di comunicazione" - si apre una pagina informativa (vedasi Fig. 12) in merito alla possibilità per gli utenti di comunicare attraverso una voce specifica - "Osservatorio Tecnologico per i Software generalizzati" - del forum di discussione dell'Istituto, inserita all'interno delle "ATTIVITA' DI LAVORO E ORGANIZZAZIONE". L'accesso al forum è possibile mediante il link predisposto, per mezzo del quale si accede alla sezione del forum suddetta (vedasi Fig. 13), che è possibile utilizzare per instaurare ed agevolare una collaborazione efficiente con gli utenti, fermo restando la possibilità di contattare direttamente l'U.O. MTS/E via email. Il forum potrà essere utilizzato dall'U.O. MTS/E anche come strumento per divulgare tematiche di particolare interesse per l'utenza.

---

<sup>15</sup> Per il download del software DIESIS, è necessario rivolgersi ai referenti Istat

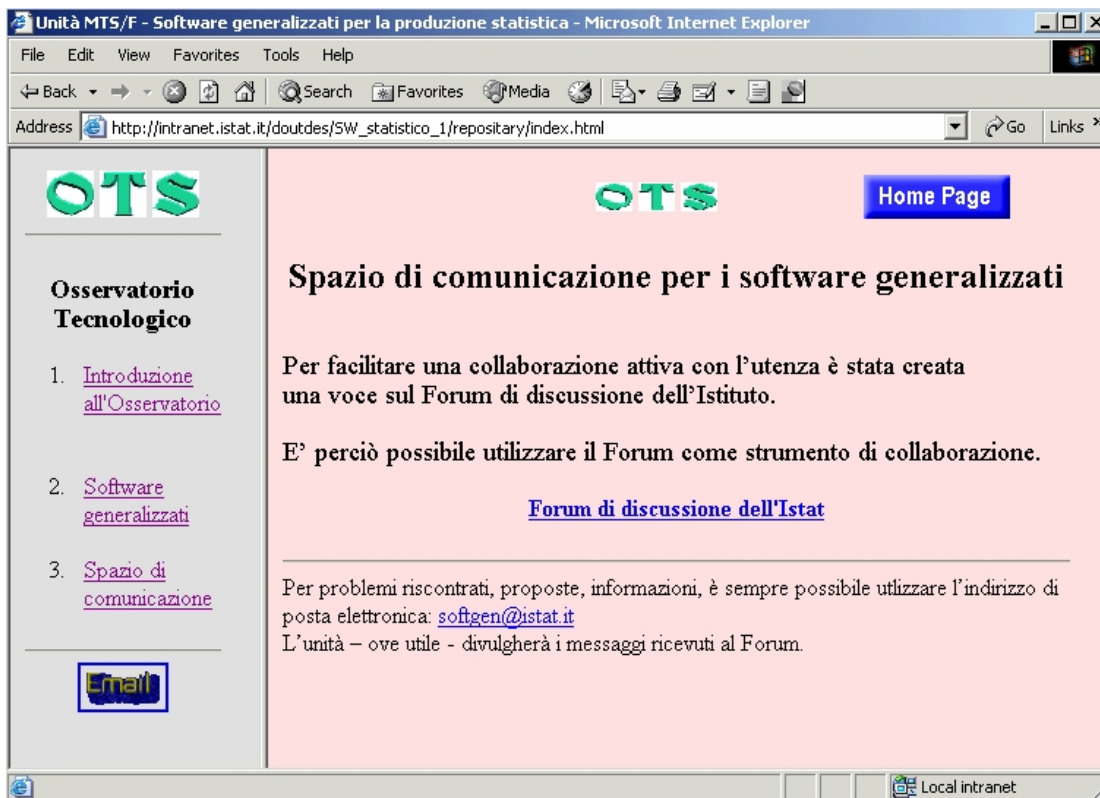


Figura 12: Sezione 3 “Spazio di comunicazione”



Figura 13: Sezione 3 dettaglio del forum di discussione

## 5.2 Confronto tra le pagine internet relative ai software generalizzati per la produzione statistica e le pagine intranet

Le pagine internet relative ai software generalizzati per la produzione statistica, sono state realizzate dall'Ufficio della Comunicazione (COM) sulla base delle pagine intranet dell'OTS create all'interno dell'U.O. MTS/E.

La versione internet delle pagine web relative ai software generalizzati per la produzione statistica, differisce da quella per intranet per vari aspetti.

Una prima differenza attiene al layout delle pagine nelle due versioni.

Nelle pagine per internet, infatti, sono presenti i link relativi alle seguenti voci:

- download del software;
- informazioni generali sul software;
- documentazione.

L'utente può, cliccando su ciascun link, accedere direttamente alla voce d'interesse o scorrere l'intera pagina per visionare tutte le informazioni.

Il link alla voce "Referenti, contatti in Istat", presente nella versione per intranet, è sostituito con il riquadro "per informazioni" posto sulla parte destra della pagina. Sopra ad esso, per i software open source per i quali è possibile il download, è previsto il riquadro "download" contenente i link per l'accesso al/i manuale/i ed alla pagina di identificazione dell'utente a seguito della quale si ottiene la pagina contenente il link al file di download del software. Per i software SAS sviluppati in Istituto per i quali è possibile il download, è previsto il riquadro "download" contenente i link per l'accesso al/i manuale/i, al file di download ed alla procedura di identificazione dell'utente nella quale va inserito il codice per ottenere la password che permette la possibilità di sbloccare il sw..

Oltre al layout, vi sono altre differenze tra le due versioni riguardanti i manuali disponibili, l'area riguardante le metodologie e le tecniche, l'esistenza delle sezioni relative ai progetti ancora in fase di implementazione, ai nuovi rilasci software ed ai software open source.

Entrambe le versioni, presentano alcune limitazioni in merito alle procedure di download di software e di manuali d'uso, consentendole, con alcune eccezioni, solo per quelli sviluppati all'interno dell'Istituto. Allo stato attuale, ad esempio, nella versione OTS per intranet, sia per alcuni software concessi su licenza all'Istituto (ad es. ACTR) che per alcuni software prototipali (ad es. DIESIS), è indispensabile rivolgersi ai relativi referenti dell'Istat. Nel caso del software Blaise, ad esempio, per il download del software è necessario contattare il referente ISTAT, mentre per il manuale, disponibile sul sito del prodotto, è presente il link associato.

In entrambe le versioni, sono invece presenti, per alcuni software i link alle pagine dei siti web (ad es. per Argus vi è il link al progetto CASC).

Nella versione OTS per internet, non si menziona la sezione relativa a "Progetti in corso", individuabile in quella per intranet nell'ambito delle aree di interesse dei software disponibili. I software in essi presenti, transiteranno nella rispettiva area sia di intranet che di internet al momento del loro rilascio.

Allo stato attuale, nella versione OTS per internet non è presente né la sezione relativa a "Nuovi rilasci software" né quella riguardante il software open source presente, in quanto si tratta di una sezione che è stata inserita solo di recente.

Nella versione per internet, diversamente da quella per intranet, per ciascuna fase d'indagine per la quale esiste almeno un software, vi è un link descrittivo delle associate metodologie e tecniche.

In tale versione, come per quella presente all'interno dell'intranet, vi è un invito rivolto all'utenza limitato all'invio di informazioni relative ad inesattezze e documentazioni aggiuntive sui prodotti presenti.

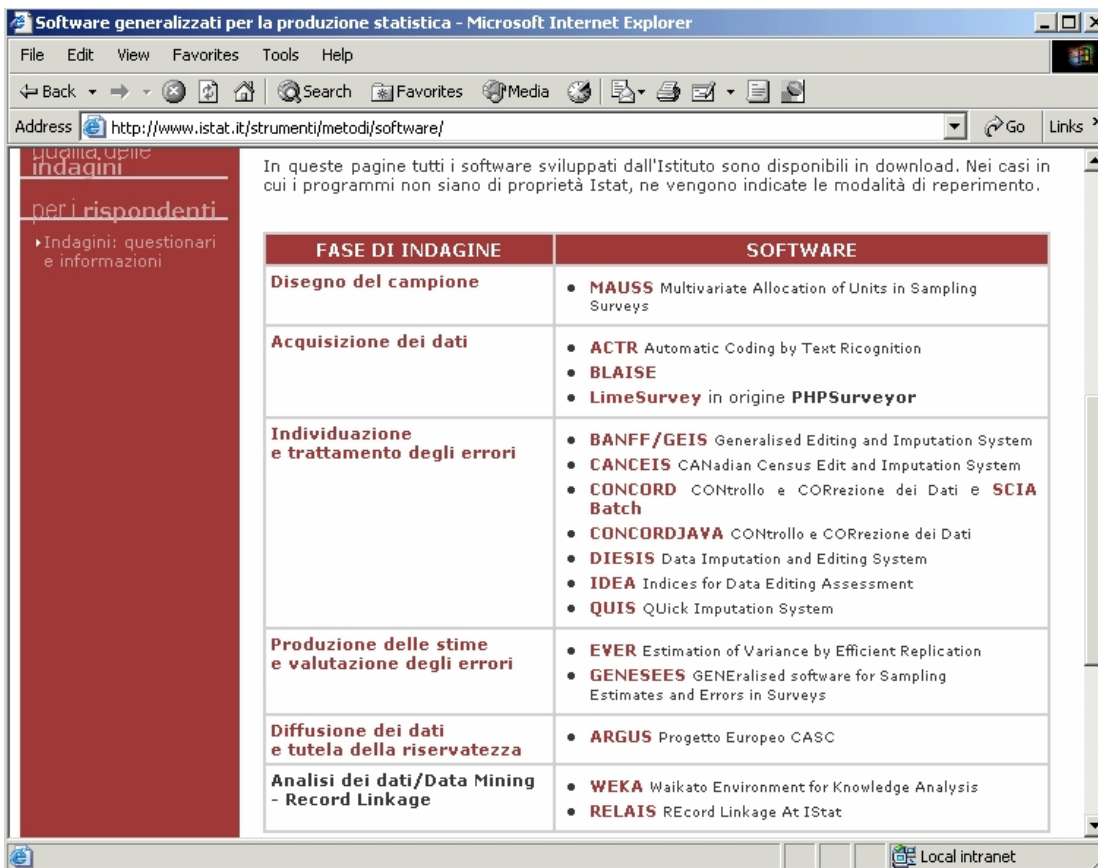


Figura 13: Tabella dei software e delle connesse fasi d'indagine

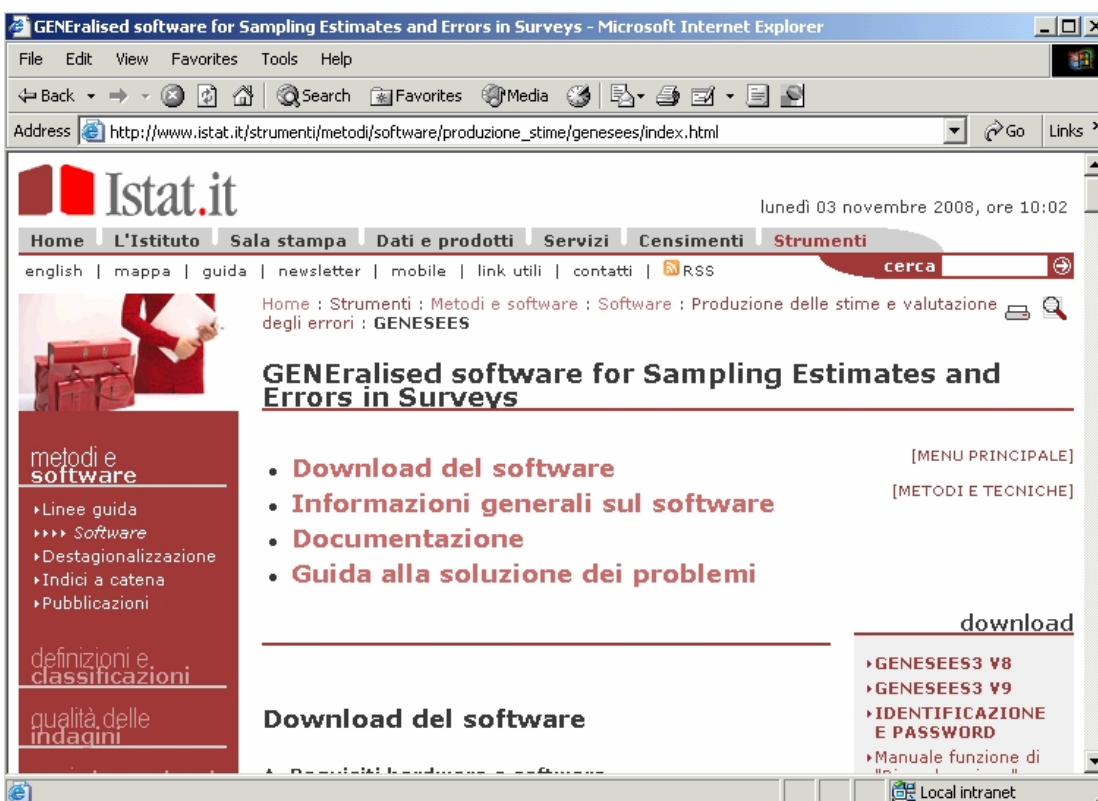


Figura 14: Pagina principale del software GENESEES

## 6. I Software dell'OTS nella configurazione attuale – Aggiornamento 2008

Nel paragrafo 6 vengono riportati i software disponibili nell'OTS – aggiornamento dicembre 2008.

In particolare nel paragrafo 6.1 è riportato l'elenco completo dei software; nei paragrafi 6.2 e successivi si evidenziano i nuovi rilasci avvenuti nel corso del 2008.

### 6.1 L'elenco completo dei software attualmente diffusi tramite l'OTS di Daniela Pagliuca

Come riportato nel paragrafo 3, le aree di interesse dominanti per operare nello sviluppo di nuovi strumenti e nella divulgazione di strumenti già esistenti, possono vedersi in relazione alle diverse fasi di una indagine statistiche.

Dall'attivazione dell'OTS ad oggi (dicembre 2008) i software che sono raccolti nell'OTS, suddivisi per le suddette fasi, sono i seguenti :

#### Fase: Disegno del campione

*MAUSS* (Multivariate Allocation of Units in Sampling Surveys) E' un software generalizzato per la determinazione dell'allocazione campionaria. Esiste la versione per PC e dal 2008 la versione per Server Linux (si veda paragrafo 6.5)

*MAUSS R* (Multivariate Allocation of Units in Sampling Surveys – versione R con interfaccia Java) E' un software generalizzato per la determinazione dell'allocazione campionaria *open source*, sviluppato infatti utilizzando il linguaggio R con interfaccia Java. E' distribuito nell'OTS a partire dal 2008 (si veda paragrafo 6.2).

#### Fase: Acquisizione dati

*ACTR* (Automatic Coding by Text Recognition) Sistema generalizzato per la codifica automatica di quesiti rilevati a testo libero.

*BLAISE* Sistema generalizzato per l'*acquisizione dati assistita* da computer

*LimeSurvey* è un tool Open Source per creare questionari per indagini pubblicate sul Web (e in parte per il data entry)

#### Fase: Controllo e correzione dati

*GEIS*, *BANFF* (Generalised Editing and Imputation System) sono due software generalizzati per il controllo e la correzione dei dati (imputazione) per le variabili numeriche, continue e non-negative.

*CONCORD* (CONtrollo e CORrezione dei Dati) è un software generalizzato integrato per il controllo e correzione dei dati. Viene distribuita la versione batch di SCIA (modulo interno di CONCORD)

*CONCOR JAVA* è la versione *open source* di CONCORD. E' distribuito nell'OTS a partire dal 2008 (si veda paragrafo 6.4).

*QUIS* (QUick Imputation System) è un software per l'imputazione delle mancate risposte parziali per variabili quantitative.

*CANCEIS* (CANadian Census Edit and Imputation System) è un software generalizzato per il controllo e la correzione di variabili qualitative e quantitative.

*DIESIS* (Data Imputation and Editing System - Italian Software) è un software per il controllo e la correzione di variabili qualitative e quantitative.

*IDEA* (Indices for Data Editing Assessment) è un software per la valutazione degli effetti di procedure di controllo e correzione su un insieme di dati grezzi a fini di documentazione, monitoraggio, miglioramento della procedura di controllo e correzione stessa.

*Fase: Produzione stime ed errori campionari*

*GENESEES* (GENERalised software for Sampling Estimates and Errors in Surveys) è un software per il calcolo delle stime e degli errori campionari. Esiste la versione per PC e la versione per e dal 2008 la versione per Server Linux (si veda paragrafo 6.5).

*EVER* (Estimation of Variance by Efficient Replication) è un software *open source* dedicato al calcolo delle stime e degli errori di campionamento in indagini complesse (si veda paragrafo 6.3)

*Fase: Diffusione dati e tutela della riservatezza*

*ARGUS* è un software per la tutela della riservatezza nella fase di rilascio dell'informazione statistica.

*Fase: Analisi dati e record Linkage*

*WEKA* (Waikato Environment for Knowledge Analysis) è uno strumento Open Source che consente di fare Data mining.

*RELAIS* (REcord Linkage At IStat) è toolkit per il record linkage (*RL*).

Nei prossimi paragrafi (dal 6.2 al 6.5) vengono in particolare evidenziati i nuovi rilasci avvenuti nel corso del 2008.

## 6.2 Nuovo software 2008: Mauss R (Multivariate Allocation of Units in Sampling Surveys - - versione open source)

### Fase: Disegno del campione

#### Informazioni generali

---

MAUSS-R (versione R con interfaccia Java) è il prototipo di un software generalizzato per la determinazione dell'allocazione campionaria, implementato nell'ambito della struttura che in Istat si occupa dei software generalizzati.

Il software disponibile su questo sito nella versione Beta, riguarda la determinazione dell'allocazione campionaria nel caso multivariato e per più domini di stima per le indagini ad uno stadio di campionamento.

La base metodologica a cui è riferito, costituisce un'estensione del metodo di allocazione di Neyman al caso di più variabili e adotta come metodo di risoluzione una generalizzazione della proposta di Bethel (1989), già implementata nel software MAUSS.

Funzioni

#### - *Definizione del progetto*

Un "progetto", per MAUSS-R, è individuato dalla cartella di lavoro, cioè dalla cartella che conterrà tutti i file generati dall'applicazione. Altre informazioni necessarie all'individuazione del progetto sono i nomi dei file di input che devono essere preparati dall'utente: il file contenente le informazioni sugli strati e sulle variabili di interesse (medie, scarto quadratico medio) e quello con i vincoli desiderati (coefficienti di variazione per le stime). Per la descrizione dei file si rimanda all'help interattivo

#### - *Definizione dei parametri*

Si possono modificare i seguenti parametri:

- \* Il numero minimo di unità per strato (default=2);
- \* Il numero massimo di iterazioni (default=25) della procedura. Queste iterazioni possono essere necessarie nel caso in cui ci siano strati in cui il numero di unità da allocare calcolato è maggiore o uguale alla popolazione. Lo strato è impostato come censuario e la procedura viene re-inizializzata;
- \* Il numero Massimo di iterazione nell'algoritmo di Chromy (default=200);
- \* Epsilon (default=1e-11): questo valore è usato per confrontare le differenze fra i risultati nelle diverse iterazioni; se la differenza è minore di epsilon la procedura termina.

#### - *Calcolo dell'allocazione*

Lancia il programma R che calcola l'allocazione campionaria nel caso multivariato per più domini di stima per le indagini ad uno stadio di campionamento.

#### - *Reports*

Visualizzazione delle stampe relative a informazioni generali sulla popolazione e ai risultati dell'allocazione.

#### Download

---

(Il software è disponibile su PC in ambiente Windows)

#### A. Requisiti hardware e software

Hardware

- 512 MB di memoria RAM
- 5 Mb su disco C:\

Software

- Java 2 Runtime Environment versione 6.0 o superiore
- Ambiente R versione 2.7.0 o superiore

## B. Download del software

*Per poter effettuare il download occorre compilare un brevissimo questionario identificativo. L'unico scopo della notifica è tenere traccia delle persone interessate a MAUSS-R in modo da poterle informare dei nuovi sviluppi del progetto.*

1. Download della versione Beta di MAUSS-R (file setup\_MaussR.exe)
2. Download delle istruzioni operative MAUSS-R *Operating Steps*
3. Istruzioni di installazione in ambiente Windows

MAUSS-R richiede l'installazione dell'ambiente JAVA (Java 2 Runtime Environment 6.0 o superiore) e dell'ambiente R.

E' necessario che la variabile d'ambiente PATH, che contiene il percorso di ricerca dei comandi, contenga il riferimento alle cartelle degli eseguibili R e Java (versione 6).

Per reimpostare la variabile di ambiente PATH in ambiente Windows :

*Start -> Impostazione -> Pannello di controllo -> Sistema -> Avanzate -> Variabili d'ambiente*

Qui reimpostare la variabile PATH selezionando la variabile PATH e cliccando sul bottone modifica. Si deve aggiungere, all'inizio della stringa, il percorso della cartella che contiene il file java.exe e quello che contiene R.exe separati da “;”.

Per esempio:

```
PATH=C:\Programmi\Java\jre1.6.0_03\bin;
C:\Programmi\R\R-7.1\bin;
C:\WINDOWS\system32;C:\WINDOWS;C:\WINDOWS\System32\Wbem;
```

Per installare il software si deve scaricare il file setup\_MaussR.exe sul proprio PC ed eseguirlo.

## Documentazione

---

- Bethel J. (1989) "Sample Allocation in Multivariate Survey" Survey Metodology, 15, pp. 47-57
- Chromy J. (1987) "Design Optimization with Multiple Objectives" Proceedings of the Survey Research Methods Section American Statistical Association, pp. 194-199
- Buglielli M.T. (2008) “MAUSS-R Operating Steps”

## Referenti presso l'Istat:

---

Unità Operativa MTS/E - “Software generalizzato per la produzione statistica”:

Maria Teresa Buglielli - bugliell@istat.it

Daniela Pagliuca - pagliuca@istat.it

\* Aspetti Metodologici:

De Vitiis Claudia - devitiis@istat.it

\* Informazioni generali:

Unità Operativa MTS/E - “Software generalizzato per la produzione statistica” - softgen@istat.it



## 6.3 Nuovo software 2008: Ever (Estimation of Variance by Efficient Replication)

**Fase: Produzione stime e valutazione errori campionari**

### Informazioni generali

---

*Che cos'è EVER*

EVER è un package R sviluppato e mantenuto, nell'ambito della struttura che in Istat si occupa dei software generalizzati, da Diego Zardetto..

Nell'attuale versione 0.9-6 il package EVER è disponibile corredato di manuale utente ed help on-line in lingua italiana. A partire dalla release 1.0 è prevista la pubblicazione del package nell'archivio CRAN (Comprehensive R Archive Network, <http://cran.r-project.org>) e la documentazione del package sarà distribuita in lingua inglese.

Il package EVER è interamente sviluppato in codice R interpretato: non contiene, cioè, codice compilato C o Fortran.

EVER può essere utilizzato sia in ambiente Windows che in ambienti UNIX-like.

*Che cosa fa EVER*

EVER è un software dedicato al calcolo delle stime e degli errori di campionamento in indagini complesse.

Nella versione attuale EVER rende disponibili le seguenti funzionalità principali:

- Replicazione del campione
- Calibrazione dei dati replicati
- Calcolo delle stime, degli errori standard e degli intervalli di confidenza di:
  - Totali
  - Medie
  - Distribuzione di frequenza assoluta
  - Tabelle di contingenza
  - Rapporti tra totali
  - Quantili
- Calcolo delle stime, degli errori standard e degli intervalli di confidenza di stimatori definiti dall'utente (arbitrari, anche *privi di una rappresentazione analitica*).
- Stime ed errori in sottopopolazioni

Nelle prossime versioni verranno inclusi in EVER altri stimatori (coefficienti di regressione, distribuzioni di frequenza cumulativa, ...). Versioni successive alla 1.0 renderanno disponibili funzionalità grafiche.

*Stima della varianza campionaria in EVER*

La tecnica di stima della varianza campionaria implementata nel package EVER si basa sul metodo DAGJK (*Delete-A-Group Jackknife*) esteso proposto da Kott.

Il metodo DAGJK può essere visto come una variante computazionalmente efficiente del tradizionale metodo *jackknife stratificato*. La necessità di costruire una replica dei pesi originali per ogni PSU inclusa nel campione rende, di fatto, irrealistico il ricorso al metodo jackknife tradizionale per indagini "complesse e grandi" (decine di migliaia di PSU in strati numerosi e di dimensione molto variabile). L'utilizzabilità pratica del metodo DAGJK poggia, al contrario, sulla capacità del metodo di costruire – per una vasta gamma di stimatori e di disegni di campionamento – stime degli errori standard (quasi non distorte anche con un piccolo numero (qualche decina) di repliche). In aggiunta alla sua peculiare efficienza computazionale, il metodo DAGJK gode dei principali vantaggi comuni ai più diffusi metodi di replicazione del campione.

L'idea base di tutti i metodi di replicazione del campione consiste nello stimare la varianza campionaria di uno stimatore arbitrario mediante una adeguata misura della variabilità delle sue stime su repliche opportunamente costruite di un campione originario. Si tratta, dunque, di metodi intrinsecamente versatili, in grado, cioè, di fornire stime della varianza campionaria senza fare ricorso ad ipotesi restrittive sulla distribuzione dei dati della popolazione e/o sulla forma funzionale degli stimatori. Poiché tutto quello di cui necessitano è (i) la definizione della tecnica di replicazione e (ii) la definizione del metodo di calcolo dello stimatore su un campione, i metodi di replicazione si prestano, fra l'altro, a stimare la varianza di *stimatori privi di una rappresentazione analitica* (non esprimibili, cioè, come funzioni di valori direttamente osservabili sulle unità statistiche).

EVER è concepito per sfruttare appieno la versatilità del metodo di replicazione DAGJK: oltre a coprire gli stimatori di uso più comune nelle indagini campionarie su vasta scala, il package fornisce, infatti, all'utente uno strumento amichevole per calcolare stime, errori standard ed intervalli di confidenza di stimatori arbitrari, definiti dall'utente medesimo. Questa funzionalità rende il package EVER particolarmente attraente in tutti i casi in cui il metodo di linearizzazione di Taylor per la stima della varianza campionaria sia applicabile solo al prezzo di forti approssimazioni (il problema della stima della povertà relativa è un possibile esempio).

### Requisiti software

Il package EVER richiede l'installazione di **R** versione **2.5.1 o superiore**.

### Download

---

*Per poter effettuare il download occorre compilare un brevissimo questionario identificativo. L'unico scopo della notifica è tenere traccia delle persone interessate ad EVER in modo da poterle informare dei nuovi sviluppi del progetto.*

- Sistemi Windows
  - Windows binary: EVER\_0.9-6.zip
- Sistemi Unix-like
  - Package source: EVER\_0.9-6.tar.gz
- Reference manual: EVER.pdf

### *Installazione e caricamento del package*

- Istruzioni per sistemi Windows
  1. Salvare il file EVER\_(versione).zip in una directory arbitraria
  2. Lanciare una sessione interattiva R e selezionare, dal menu "Pacchetti", l'opzione "Installa pacchetti da file zip locali..."
  3. R richiederà di indicare la directory in cui risiede il file EVER\_(versione).zip e di selezionare il file
  4. Caricare il package EVER digitando l'istruzione `library(EVER)` oppure `require(EVER)`. In alternativa è possibile selezionare, dal menu "Pacchetti", l'opzione "Carica pacchetto..." e scegliere EVER dal menu a tendina
- Istruzioni per sistemi Unix-like
  1. Salvare il file EVER\_(versione).tar.gz in una directory arbitraria
  2. Posizionarsi nella directory che contiene il file EVER\_(versione).tar.gz e digitare dal prompt di shell l'istruzione:  
R CMD INSTALL EVER\_(versione).tar.gz

3. Lanciare una sessione R e caricare il package EVER digitando l'istruzione `library(EVER)` oppure `require(EVER)`

Per ulteriori informazioni è possibile consultare il capitolo 6 "Add-on packages" del manuale "R Installation and Administration".

## Referenti presso I'istat:

---

Unità Operativa DPTS/DCMT/MTS-E  
"Software generalizzato per la produzione statistica"

Diego Zardetto – [zardetto@istat.it](mailto:zardetto@istat.it)

## Documentazione

---

### 1. Generalità sul calcolo delle stime e degli errori di campionamento

- Wilkinson, G.N., Rogers, C.E. - (1973) "Symbolic Description of Factorial Models for Analysis of Variance" Journal of the Royal Statistical Society, series C (Applied Statistics), Vol. 22, pp. 181-191.
- Särndal, C.E., Swensson, B., Wretman, J. - (1992) "Model Assisted Survey Sampling" Springer Verlag.
- Deville, J.C., Särndal, C.E. - (1992) "Calibration Estimators in Survey Sampling" Journal of the American Statistical Association, Vol. 87, No. 418, pp. 376-382.
- Estevao, V., Hidiroglou, M. A., Särndal, C. E - (1995) "Methodological principles for a generalized estimation system at Statistics Canada" Journal of Official Statistics, 11, n.2, pp. 181-204.
- Singh, A.C., Mohl, C.A. - (1996) "Understanding calibration estimators in survey sampling" Survey Methodology, 22, pp. 107-115.
- Vanderhoeft, C. - (2001) "Generalized Calibration at Statistic Belgium" Statistics Belgium Working Paper n. 3 [http://www.statbel.fgov.be/studies/paper03\\_en.asp](http://www.statbel.fgov.be/studies/paper03_en.asp).
- Fuller, W.A. - (2002) "Regression estimation for survey samples" Survey Methodology, 28, pp. 5-23.
- Lumley, T. - (2006) "survey: analysis of complex survey samples" R package version 3.6-5. <http://cran.at.r-project.org/web/packages/survey/index.html>.
- Scannapieco, M., Zardetto, D., Barcaroli, G. - (2007) "La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESEES/SAS" Contributi Istat n. 4. [http://www.istat.it/dati/pubbsci/contributi/Contributi/contr\\_2007/2007\\_4.pdf](http://www.istat.it/dati/pubbsci/contributi/Contributi/contr_2007/2007_4.pdf).

### 2. Stima della varianza campionaria

- Woodruff, R. S. - (1971) "A Simple Method for Approximating the Variance of a Complicated Estimate" Journal of the American Statistical Association, Vol. 66, n. 334, pp. 411-414.
- Kalton, G. - (1979) "Ultimate cluster sampling" Journal of the Royal Statistical Society, series A (General), Vol. 142, Part 2, pp. 210-222.
- Krewski, D., Rao, J. N. K. - (1981) "Inference From Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Methods" The Annals of Statistics, Vol. 9, No. 5, pp. 1010-1019.
- Binder, D. A. - (1983) "On the variances of asymptotically normal estimators from complex surveys" International Statistical Review, 51, pp. 279-292.
- Rust, K. - (1985) "Variance Estimation for Complex Estimators in Sample Surveys" Journal of Official Statistics, 1, pp. 381-397.
- Bellhouse, DR. - (1985) "Computing Methods for Variance Estimation in Complex Surveys" Journal of Official Statistics, Vol.1, No.3, pp. 323-329.
- Dorfman, A., Valliant, R. - (1993) "Quantile variance estimators in complex surveys" Proceedings of the ASA Survey Research Methods Section, pp. 866-871.

- Rao, J. N. K., Lohr, S. L. - (1999) "Some Current Trends in Sample Survey Theory and Methods" Sankhya: The Indian Journal of Statistics, Special issue on Sample Surveys, Volume 61, Series B, Pt. 1, pp. 1-57.
- Duchesne, P. - (2000) "A Note on Jackknife Variance Estimation for the General Regression Estimator" Journal of Official Statistics, Vol.16, No.2, pp. 133-138.
- Valliant, R. - (2000) "Variance estimation for the general regression estimator" Survey Methodology, 28, pp. 103-114.
- Rao, J. N. K., Lohr, S. L. - (2004) "Sample Survey Methods: Recent Developments and Applications" two-day workshop slides, Joint Statistical Meetings, Toronto.

### *3. Metodo Delete-A-Group jackknife*

- Kott, Phillip S. - (1998) "Using the Delete-A-Group Jackknife Variance Estimator in NASS Surveys" RD Research Report No. RD-98-01, USDA, NASS: Washington, DC.
- Kott, Phillip S. - (1999) "The Extended Delete-A-Group Jackknife" Bulletin of the International Statistical Institute. 52nd Session. Contributed Papers. Book 2, pp. 167-168.
- Kott, Phillip S. - (2001) "The Delete-A-Group Jackknife" Journal of Official Statistics, Vol.17, No.4, pp. 521-526.
- Kott, Phillip S. - (2008) "Building a Better Delete-a-Group Jackknife for a Calibration Estimator" NASS Research Report, NASS: Washington, DC.

## 6.4 Nuovo software 2008: Concord-Java (CONtrollo e CORrezione dei Dati - versione *open source*)

### Fase: Individuazione e trattamento degli errori

#### Informazioni

---

CONCORDJAVA è la versione *Open Source* di CONCORD, un software generalizzato integrato per il controllo e correzione dei dati. L'applicazione richiama i software precedentemente sviluppati ed utilizzati in Istat:

- SCIA sviluppato da E. Riccini Margarucci, F. Silvestri e P. Floris. Nella sezione Download di CONCORD è disponibile la versione batch di SCIA
- RIDA sviluppato da T. Buglielli, G. Massimini e R. Colosi

L'applicazione, attualmente rilasciata in versione Beta, è disponibile per il download nella versione in lingua italiana ed in inglese.

I diversi metodi residenti nel software sono implementati in moduli distinti:

- SCIA - (*Sistema di Controllo e Imputazione Automatica*).

Esegue il controllo e la correzione di variabili qualitative applicando integralmente la metodologia di Fellegi-Holt. Per ciascun record errato il sistema prima individua il minimo numero di variabili da imputare e poi effettua l'imputazione garantendo azioni di imputazione coerenti con un predefinito insieme di regole di controllo;

- RIDA - (*Ricostruzione dell'Informazione con Donazione Automatica*).

Esegue l'imputazione di variabile qualitative e quantitative mediante donatore di minima distanza. Operazioni propedeutiche sono:

- la classificazione delle unità in esatte ed errate;
- la loro registrazione in due file distinti;
- la identificazione dei valori da imputare mediante un predefinito carattere (di errore).

Propedeutica ai vari passi è la fase di definizione delle variabili, cioè dei campi del record da sottoporre a controllo, e degli edit o regole di controllo sia formali che sostanziali individuabili a partire dal questionario e dalla conoscenza relativa ai fenomeni indagati.

CONCORDJAVA

#### Download

---

##### A. Requisiti hardware e software per il download

###### Hardware

- 256 Mb memoria RAM
- 30 Mb su disco C:/

###### Software

- Java 2 Runtime Environment 6.0 o superiore

##### B. Download

*Per poter effettuare il download occorre compilare un brevissimo questionario identificativo. L'unico scopo della notifica è tenere traccia delle persone interessate a CONCORDJAVA in modo da poterle informare dei nuovi sviluppi del progetto.*

- Download di CONCORDJAVA 1.0 in italiano (versione Beta) CONCORDJAVA (file *setup\_ConcordJava.exe*)
- Download di CONCORDJAVAEN 1.0 in inglese (versione Beta) CONCORDJAVAEN (file *setup\_ConcordJava.exe*)
- Istruzioni di installazione in ambiente Windows

ConcordJava richiede l'installazione dell'ambiente JAVA (Java 2 Runtime Environment 6.0 o superiore).

E' anche necessario che la variabile d'ambiente PATH, che contiene il percorso di ricerca dei comandi, contenga il riferimento alla cartella JavaPer reimpostare la variabile di ambiente PATH in ambiente Windows :*Start -> Impostazione -> Pannello di controllo -> Sistema -> Avanzate -> Variabili d'ambiente*

Qui reimpostare la variabile PATH aggiungendo il percorso della cartella che contiene il file *java.exe*.

Per esempio:

```
PATH=C:\Programmi\Java\jre1.6.0_03\bin;C:\WINDOWS\system32;C:\WINDOWS;C:\WINDOWS\System32\Wbem;
```

Per installare il software si deve scaricare il file *setup\_ConcordJava.exe* sul proprio PC ed eseguirlo.

L'applicazione richiama l'editor di testo notepad.exe. Se si preferisce un altro editor di testo e' possibile modificare, nel file *concord.bat*, il settaggio della variabile EDITOR. Per esempio: set EDITOR=C:\Programmi\Notepad++\notepad++.exe

## **Documentazione** (equivalente a Concord versione SAS)

---

-Abbate C., Bove G., Crescenzi F. - (1992) *"Metodi statistici multivariati per la ricostruzione dell'informazione mancante"* Avanzamenti metodologici e statistiche ufficiali. Atti delle prime giornate di studio SIS-ISTAT Roma 13-14 dicembre 1992

-Abbate C., Giommi A. - (1993) *"Metodi di ponderazione e di correzione di dati elementari"* Atti del Convegno "La qualità dell'informazione statistica e la qualità industriale" SIS-ISTAT-AICQ, Roma 10 maggio 1991

-Abbate C., Schivano R. - (1993) *"Efficacia dell'imputazione da donatore con distanza minima"* Atti del Convegno SIS, Sanremo 1993

-Abbate C. - (1996) *"La completezza delle informazioni e l'imputazione da donatore con distanza mista minima"* Quaderni di Ricerca, ISTAT 1996

-Alboni, F. - (1994) *"Il controllo e la revisione dei dati"* L'utilizzazione della rete di contabilità agraria in Emilia-Romagna (Cap.7) a cura di F.Alvisi e C.Filippucci, Calderini Editore

-Barcaroli G. - (1992) *"An integrated system for edit and imputation of data in the Italian Statistical Institute"* Survey and Statistical Computing, pp.167-177

-Barcaroli G. - (1993) *"Un approccio logico formale al problema del controllo e della correzione dei dati statistici"* Quaderni di ricerca ISTAT n.9/1993

-Barcaroli G., Ceccarelli C., Luzi O. - (1995) *"An edit and imputation system of quantitative variables based on macroediting techniques"* Proceedings of the International Conference on Survey Measurement and Process Quality Bristol (UK), 1-4 Aprile 1995, pp.12-17

-Barcaroli G., Ceccarelli C., Luzi O., Manzari A., Riccini Margarucci E., Silvestri F. - (1995) *"The Methodology of Editing and Imputation by Qualitative Variables implemented in SCLA"* Documento interno ISTAT.

-Barcaroli G., Luzi O. - (1995) *"Sistema generalizzato per l'editing e l'imputazione di variabili quantitative (GEIS)"* Quaderni di ricerca ISTAT n.1/1995

-Barcaroli G. - (1998) *"La correzione probabilistica dei dati: il trattamento congiunto degli errori di rilevazione casuali e sistematici stocastici mediante l'applicazione del teorema di Bayes alla metodologia Fellegi-Holt"* Statistica Applicata vol. 10 n.2/1998

- Barcaroli G., D'Aurizio L., Luzi O., Manzari A., Pallara A. - (1999) *"Metodi e software per il controllo e la correzione dei dati"* Documenti ISTAT, n.1/1999
- Cotton C. - (1991) *"Functional description of the generalized edit and imputation system"* Business Survey Methods Division
- Davila H.E. - (1992) *"The Hidiroglou-Berthelot Method"* Statistical Data Editing Methods and Techniques United Nations, Vol. I, February, 1992
- Engstrom P., Angsved C. - (1994) *"A description of a geographical macro-editing application"* Statistical Commission and Economic Commission for Europe-Conference of European Statisticians Cork, Ireland, 17-20 October 1994
- Fellegi I.P., Holt D. - (1976) *"A systematic approach to edit and imputation"* Journal of the American Statistical Association, vol.71, pp.17-35
- Ford B.L. - (1983) *"An overview of hot-deck procedures"* Incomplete data in sample survey vol.1, pg. 191, Academic Press, New York
- Garcia Rubio E., Villan Criado I. - (1988) *"Sistema DLA, Sistema de deteccion y imputacion automatica de errores para datos cualitativos"* Instituto Nacional de Estadistica, Madrid, 1988
- Granquist L. Statistical Data Editing Methods and Techniques United Nations, Vol. I, February, 1992
- a - *"A Review of methods for rationalizing the editing of survey data"*
- b - *"The Aggregate Method"*
- c - *"The Top-Down Method"*
- d - *"On the need for generalized numeric and imputation system"*
- Granquist L. - (1995) *"Improving the traditional editing process"* Business Survey Methods, John Wiley and sons
- Granquist L. - (1995) *"An overview of methods of evaluating editing processes"* Conference of European Statisticians (Athens, Greece, 6-9 November), Working Paper n. 3
- Grenless J.S., Reece W.S., Zieschang K.D. - (1982) *"Imputation of Missing Values when the Probability of Response Depends on the Variable Being Imputed"* Journal of the American Statistical Association, 77, pp 251-261
- Hawkins D.M. - (1974) *"The Detection of Errors in Multivariate Data Using Principal Components"* Journal of the American Statistical Association, Vol. 69. No 346
- Hidiroglou M.A., Berthelot J.M. - (1986) *"Statistical Editing and Imputation for Periodic Business Surveys"* Survey Methodology, June 1986, vol.12, N.1, pp.73-83
- Jackson J.E. - (1959) *"Quality control methods for several related variables"* Technometrics, vol. 1, n. 1
- Kalton G., Kasprzik D. - (1986) *"The treatment of missing survey data"* Survey methodology, 12, 1, Statistics Canada
- Kleijnen J., Van Grodendaal W. - (1992) *"Simulation. A Statistical Perspective"* John Wiley, New York
- Kovar J.G., MacMillian J.H., Whitridge P. - (1988) *"Overview and strategy for the generalized edit and imputation system"* Statistics Canada, Methodology Branch, April 1988 (updated February 1991)
- Kovar J.G., Whitridge P. - (1995) *"Imputation of business survey data"* Business Survey Methods, John Wiley and sons
- Latouche M., Berthelot J.M. - (1992) *"Use of Score Function to Prioritize and Limit Recontacts in Editing Business Surveys"* Journal of Official Statistics, Vol.8, No.3, Part II.
- Lee H., Ghangurde P.D., Mach L., Yung W. - (1992) *"Outliers in sample survey"* Statistics Canada Methodology Branch
- Little R.J.A., Smith J. - (1983) *"Multivariate Edit and Imputation for Economic Data"* American Statistical Association Proceedings of the Survey Research Methods Section
- Little R.J.A., Smith J. - (1987) *"Editing and Imputation for Quantitative Survey Data"* Journal of the American Statistical Association Vol. 82, N. 397, Applications Section
- Lindstrom K. - (1992) *"A macroediting application developed in PC-SAS"* Statistical Data Editing Methods and Techniques United Nations, Vol. I, February, 1992
- Luzi O., Ceccarelli C. - (1997) *"Le componenti principali nello studio dell'editing multivariato"* Atti della XXXV Riunione Scientifica della Società Italiana di Economia, Demografia e Statistica
- Luzi O. - (1996) *"Applicabilità ed impatto potenziale dei metodi per l'editing di dati quantitativi basati sugli approcci del Macroediting e dell'Editing Selettivo"* Contributi ISTAT

- Masselli M., Signore M., Panizopn F. - (1992) *“Il sistema di controllo della qualità dei dati”* Manuale di tecniche di indagine, Vol.6 ISTAT
- Winkler W.E. - (1994) *“SPEER Edit System”* Computer system and unpublished documentation Statistical Research Division U.S. Bureau of the Census, Washington D.C., USA.
- Riccini E. - (2004) *“CONCORD versione V.1.0 - Controllo e correzione dati. Manuale utente e aspetti metodologici”* Tecniche e strumenti n. 1 - ISTAT



## 6.5 Una nuova release dei software Mauss e Genesees versione SAS su Server Linux

### a) MAUSS – Fase Disegno del Campione

Il software MAUSS è disponibile anche su PC in ambiente Windows .

Su Linux:

Per la sua utilizzazione occorre:

- Connettersi al server NORVEGIA
- Posizionarsi sulla directory /softgen/mauss/lancio
- Lanciare la procedura immettendo il comando ./mauss.sh

Dopo aver eseguito ogni elaborazione, la procedura registrerà le directory e i file di input ed output nel dataset sas progetti.sas7bdat, che verrà scritto nella home dell'utente, dentro la directory progetti/mauss.

Tale dataset è utilizzato dal software per permettere all'utente di ripetere una elaborazione già effettuata, variando eventualmente i parametri di elaborazione ma non directory o file di input-output (analogo uso della versione per PC/Windows). Il file di log e i file di output vengono invece scritti nella directory di output scelta dall'utente.

Si consiglia di eliminare periodicamente quanto non necessario dalla propria home, onde evitare che lo spazio assegnato alla stessa si esaurisca e non sia quindi possibile registrare il dataset progetti.sas7bdat.

Per una corretta visualizzazione delle schermate presentate dalla procedura, si consiglia di impostare la risoluzione dello schermo del proprio PC a 1280 x 1024 pixel.

### b) GENESEES – Fase Produzione stime ed errori campinari

Il software GENESEES è disponibile anche su PC in ambiente Windows .

Su Linux:

Per la sua utilizzazione occorre:

- Connettersi al server NORVEGIA
- Posizionarsi sulla directory /softgen/genesees/lancio
- Lanciare la procedura immettendo il comando ./genesees.sh

Dopo aver eseguito ogni elaborazione, la procedura registrerà le directory e i file di input ed output nel dataset sas progetti.sas7bdat, che verrà scritto nella home dell'utente, dentro la directory progetti/genesees.

Tale dataset è utilizzato dal software per permettere all'utente di ripetere una elaborazione già effettuata, variando eventualmente i parametri di elaborazione ma non directory o file di input-output (analogo uso della versione per PC/Windows).

Il file di log e i file di output vengono invece scritti nella directory di output scelta dall'utente. Si consiglia di eliminare periodicamente quanto non necessario dalla propria home, onde evitare che lo spazio assegnato alla stessa si esaurisca e non sia quindi possibile registrare il dataset progetti.sas7bdat.

Per una corretta visualizzazione delle schermate presentate dalla procedura, si consiglia di impostare la risoluzione dello schermo del proprio PC a 1280 x 1024 pixel. Informazioni generali

## 7. Procedura di notifica della registrazione utenti e rilascio della password – Aggiornamento Dicembre 2008

Nelle procedure istituzionali stabilite per il rilascio tramite l'OTS del software sviluppato in Istat è prevista, al momento della installazione, la registrazione da parte della utenza che deve compilare un brevissimo questionario identificativo. Ovviamente la registrazione è a servizio dell'utenza stessa, poiché ogni utilizzatore viene inserito in una lista da adoperare (se lo autorizza) per i successivi contatti per eventuali aggiornamenti e/o per messaggi informativi e/o patch.

Come accennato nei precedenti paragrafi 3 e 4, per i software open source è anche prevista l'accettazione della licenza EUPL e, nel caso di alcuni software non open source, è prevista anche la richiesta di una password di utilizzo, che serve a registrare le singole installazioni; in questo secondo caso, una volta installato il software e alla sua prima esecuzione, viene infatti richiesta all'utente una password di utilizzo, mostrando un codice che, inserito nel questionario identificativo, permetterà l'invio della password in automatico via email all'utente.

Nei casi di software open source, per quanto sopra detto, la registrazione non prevede ovviamente alcun codice da inserire e rilascio di password.

In questo paragrafo viene illustrata la procedura informatica che è stata sviluppata per la registrazione della utenza e il rilascio della password.

La realizzazione di questa procedura informatica è nata dall'esigenza iniziale di fornire in tempo reale l'invio della password, indispensabile all'utilizzo dei software generalizzati quali CONCORD, GENESEES e MAUSS sviluppati con interfacce SAS ed utilizzando come canale di comunicazione il portale dell'Istituto.

A seguito della compilazione di una form di registrazione, con alcuni dati identificativi e dell'indirizzo di posta elettronica dell'utente, che rispettano la tutela della privacy secondo quanto previsto dal decreto legislativo n. 196 del 30 giugno 2003, il sistema prevede l'invio in automatico di due mail: la prima all'utente con la password di riferimento ed una seconda di notifica ad un prefissato indirizzo di posta interna che colleziona le informazioni degli utenti che utilizzano tali prodotti.

Tale processo informativo veniva fino ad ora gestito manualmente tramite l'invio di una mail da parte dell'utente finale a *softgen@istat.it*, ed a carico della struttura MTS/E veniva calcolata la password mediante uno specifico algoritmo, inviando successivamente una mail di ritorno con la password a fronte del codice fornito nell'ambito della prima esecuzione del software generalizzato.

Nell'implementazione di tale applicazione, si è ritenuto utile, tener traccia anche di altre tipologie di software che non implicano una password per il loro utilizzo: CONCORD JAVA e RELAIS. Entrambi i software generalizzati in questione forniscono il download al termine della registrazione di un file eseguibile di setup che permette mediante passi guidati di tipo wizard l'installazione sulla macchina utente dell'applicazione finale.

Di diversa natura è il package R EVER, che rappresenta l'implementazione di funzionalità aggiuntive a quelle già presenti all'interno del kernel del software statistico R, dove il funzionamento avviene effettuando un download di un file di tipo .zip per sistemi operativi Windows o di un file di tipo .tar.gz per sistemi operativi di tipo Unix-like. Il setup del package avviene all'interno del software R e non mediante l'esecuzione di un file di setup.

L'intera procedura è stata realizzata in HTML e PHP, seguendo le direttive e gli standard dell'Istituto in ottica Open Source.

La procedura gestisce in maniera integrata le due diverse filosofie applicative di utilizzo dei software generalizzati, entrambe offrono vantaggi e svantaggi. Quelli sviluppati con interfaccia SAS che possono essere scaricati senza registrazione dall'utente, devono essere dapprima installati sulla macchina utente, alla prima esecuzione viene fornito un codice per ottenere la password, effettuata la registrazione fornendo tale codice alla procedura, verrà fornita via mail la password di sblocco per l'esecuzione del software. I secondi essendo realizzati in linguaggi Open Source non richiedono nessun tipo di password, ma per tener traccia degli utilizzatori per effettuare il download è richiesta anticipatamente la registrazione dell'utente.

La prima realtà applicativa offre come vantaggio la certezza della registrazione degli utilizzatori, in contrapposizione lo svantaggio è dato dal doppio accesso all'OTS per gli utenti.

Nella seconda modalità operativa, il vantaggio è l'unico accesso all'OTS che permette contestualmente la registrazione e il successivo download del software, ma come svantaggio comporta il mancato controllo sulla veridicità dell'indirizzo email (in quanto l'applicazione prevede solo un controllo formale dell'indirizzo digitato dall'utente) e che l'utente registrato sia il solo utilizzatore del software scaricato.

Si è ipotizzato di eliminare la componente della password da quei software descritti precedentemente con interfacce in linguaggi proprietari SAS, ma nell'ottica di minima invasività alle modalità operative già in produzione si è preferito gestire nella procedura di registrazione i due diversi scenari operativi. Se in futuro si decidesse di intervenire sull'eliminazione delle password, con un minimo impatto implementativo sul codice sorgente PHP, si potranno gestire tutti i software generalizzati con la procedura definita per quelli di tipo "Open Source".

La procedura viene invocata dalle pagine dell'OTS Intranet e Internet. Per i soli software con password, la richiesta di password può essere invocata direttamente dall'applicazione SAS, in quanto tali interfacce permettono di catturare il "codice per password" in una pagina HTML e richiamare direttamente la form di registrazione, senza alcuna digitazione, in quanto una errata trascrizione dei 15 caratteri numerici, provocherebbe una password di ritorno errata, con il conseguente obbligo di registrarsi nuovamente.

La procedura prevede inoltre alcuni controlli prettamente formali riguardanti l'inserimento di un nominativo dell'utente, del suo indirizzo di posta elettronica. Inoltre per i soli software che richiedono la password per l'utilizzo verifica la corretta la lunghezza del codice per password di 15 caratteri e che in esso non esistano caratteri non strettamente numerici.

Il contesto architetturale dell'Istituto, prevede che la procedura in questione risieda unicamente sull'application server SAMU residente sul dominio "istat.it". L'applicazione verrà richiamata direttamente dalle pagine dell'OTS (Osservatorio Tecnologico dei Software generalizzati) nato appunto con lo scopo di rendere disponibili i software generalizzati per la produzione statistica e le relative informazioni agli utenti interessati. La fruizione di tali servizi riguarda sia gli utenti esterni che accedono via Web, sia a coloro che hanno la possibilità di accedere dalla rete aziendale Intranet. In entrambi i casi l'invio delle mail sarà totalmente a carico del Mail -Server SMTP dell'ISTAT.

REGISTRAZIONE SOFTWARE GENERALIZZATO

Informazioni Identificative

Nominativo Utente (\*) :  
Marco Broccoli

Ente/Istituto/Società di appartenenza :  
Istituto Nazionale di Statistica

Email (\*) : broccoli@istat.it

Città : Roma

Stato : Italia

Informazioni aggiuntive

Vuoi ricevere email di info su aggiornamenti (patch e/o nuove versioni)? Si

Invia Cancella (\*) Campi obbligatori

Per problemi di registrazione contattare softgen@istat.it

**Figura 15:** Concessione Password

**REGISTRAZIONE SOFTWARE GENERALIZZATO**

**Informazioni per la password**  
 Codice per Password (\*): B 159873326547466

**Informazioni Identificative**  
 Nominativo Utente (\*): Marco Broccoli  
 Ente/Istituto/Società di appartenenza: Istituto Nazionale di Statistica  
 Email (\*): broccoli@istat.it  
 Città: Roma  
 Stato: Italia

**Informazioni aggiuntive**  
 Vuoi ricevere email di info su aggiornamenti (patch e/o nuove versioni)?  Sì

(\*) Campi obbligatori

Per problemi di registrazione contattare [softgen@istat.it](mailto:softgen@istat.it)

**Figura 16:** Registrazione Download

**PROCEDURA TERMINATA**

GRAZIE!  
 La registrazione è terminata.

**Riceverà una email da [softgen@istat.it](mailto:softgen@istat.it) all'indirizzo definito nella procedura contenente la password da utilizzare per l'abilitazione del software.**

Per problemi di registrazione contattare [softgen@istat.it](mailto:softgen@istat.it)

**Figura 17:** Avviso procedura terminata

**PROCEDURA TERMINATA**

GRAZIE!  
 La registrazione è terminata.

**A. Requisiti software**

RELAIS 1.0 richiede l'installazione degli ambienti Java e R. Con riferimento all'ambiente Java si richiede l'installazione di Java 2 Runtime Environment 6.0 o superiore; con riferimento all'ambiente R e' necessaria l'installazione della versione 2.5.1 o superiore, e' infine necessario installare il package R lpSolve versione 5.5 o superiore.

**B. Download**

Download del software [RELAIS](#)

Per problemi di registrazione contattare [softgen@istat.it](mailto:softgen@istat.it)

**Figura 18:** Download

## **8. Lo sviluppo dell'Osservatorio tecnologico e delle attività legate ai software generalizzati**

In questo paragrafo vengono presentati alcuni punti che sono alla base delle attività da svolgersi nel prossimo futuro per quanto concerne l'OTS.

### Assistenza utenza

E' in corso la valutazione di una applicazione software, OIL, Online Interactive heLpdesk, sviluppata presso l'Ufficio Sistemi Informativi del CNR. Lo scopo è stabilire se attivare o meno un progetto di riutilizzo in Istat per usufruire del software OIL, utile ad agevolare lo svolgimento delle attività di assistenza agli utenti dei software generalizzati.

Il software OIL è oggi adoperato all'interno del CNR stesso, che lo ha sviluppato, per gestire l'assistenza di software di tipo gestionale e potrebbe essere riutilizzato in Istat.

In alternativa occorrerà valutare ulteriori analoghi strumenti per l'help desk

OIL verrebbe ad essere integrato con la procedura già attiva in Istat di registrazione delle utenze (si veda paragrafo 7); in tal modo - in automatico - l'utenza dei software generalizzati corrisponderebbe all'utenza di OIL, ovvero a quella supportata per l'assistenza nei problemi riscontrati nell'uso del software stesso e grazie a ciò le soluzioni adottate, a fronte di problematiche che spesso si ripetono, rimarrebbero memorizzate e disponibile.

### Registrazione informazioni utenza su archivio

Attualmente la registrazione della utenza dei software generalizzati scaricati tramite l'OTS - internet o intranet - avviene tramite l'invio automatico di una email che, al momento del download del software, è spedita alla casella elettronica [softgen@istat.it](mailto:softgen@istat.it). Lo scopo è quello di garantire eventuali successivi contatti.

Tale registrazione non è però memorizzata in alcun archivio: tramite il messaggio contenente l'email dell'utente, manualmente, occorre inserire l'indirizzo in una specifica rubrica associata al software, così da avere a disposizione la lista delle utenze del software stesso.

Si sta attualmente prendendo in considerazione la possibilità di registrare tale informazioni in un archivio fisico che automaticamente poi inserisca l'email dell'utente in una apposita rubrica. Ciò sarebbe comunque necessario nel caso si proceda nell'utilizzo di OIL in Istat (vedi punto precedente)

### Lista collettiva delle utenze interessate

Come generalizzazione del punto di cui sopra, si sta valutando la possibilità di soddisfare quella utenza che, pur non avendo scaricato alcun software, potrebbe essere interessata a ricevere informazioni sui software generalizzati. Si potrebbe considerare una lista dei contatti atta a ricevere informazioni generali o notizie connesse, anche specifiche delle nuove versioni dei software sviluppati prescindendo dal singolo software. Per la circolazione delle informazioni la soluzione migliore potrebbe essere quella di creare una lista dinamica sulla base delle diverse rubriche dei singoli software e di una lista generale.

### Portali ed iniziative internazionali

Nell'ambito delle attività dell'Eurostat e della Comunità Europea, con particolare riferimento ad ESSnet (European Statistical System network) e OSOR (Open Source Observatory and Repository) occorrerà stabilire le eventuali connessioni dell'OTS in particolare con il repository OROS (per maggior dettaglio si veda il paragrafo 9.)

## 9. Informazioni utili per ricognizioni internazionali

A livello internazionale esistono vari siti consultabili per raccogliere informazioni utili alla ricerca in campo statistico informatico, molti riguardanti in particolare i software generalizzati. Nel seguito vengono evidenziate alcune associazioni ritenute particolarmente interessanti

Un istituzione attiva è la ISI - International Statistical Institute (<http://isi.cbs.nl/>).

Fondata nel 1885 ed è una delle associazioni scientifiche più antiche. A questa appartengono più di 2000 membri eletti e riconosciuti come leader di campi statistici e rappresenta più di 133 paesi diversi. Oltre a questo gruppo di esperti ci sono anche circa 3000 membri aggiunti che appartengono a sezioni specializzate dell'Istituto

Tra le sezioni della ISI una è particolarmente interessante per il tema qui trattato.

Si tratta della IASC - International Association For Statistical Computing (<http://isi.cbs.nl/iasc.htm>).

La IASC è stata fondata nel 1977 (durante la 41-ma sessione dell'ISI) allo scopo di stimolare l'interesse nello scambio di conoscenze specificatamente tecnico-computazionali, attraverso contatti e meeting tra statistici, esperti computazionali, professionisti, così come relazioni tra organizzazioni, istituzioni.

I meeting vengono spesso organizzano in contemporanea alle stesse sessioni dell' ISI ma lo IASC organizza anche sue proprie conferenze.

Tra le conferenze la più nota e seguita è quella biennale dell'ISI ; lo IASC invece organizza stabilmente (tramite la European Regional Section - ERS) il Compstat biennale e (tramite the Asian Regional Section - ARS) la Statcomp triennale

Per maggiori dettagli si può leggere l'Appendice 1 (lista conferenze internazionali ISI/IASC).

E' anche in stretto contatto con la Interface Foundation of North America e con sta cercando di estendere i contatti con Middle East and South America.

L'ISI ha diverse pubblicazioni, di riviste, libri, newsletter etc. Tra questi è da evidenziare il Computational Statistics & Data Analysis,

([http://www.elsevier.com/wps/find/journaldescription.cws\\_home/505539/description/](http://www.elsevier.com/wps/find/journaldescription.cws_home/505539/description/)), che è la rivista ufficiale dello IASC, dedicata alla diffusione di ricerca e applicazioni statistiche che comprendendo sia gli aspetti computazionali che di analisi dei dati.

Le informazioni sullo IASC si possono avere anche tramite la ISI Newsletter. (<http://isi.cbs.nl/Newsltr.htm>)

Sempre in campo internazionale si può evidenziare la ASC - Association for Survey Computing - <http://www.asc.org.uk/>.

E' una associazione Inglese, ma alla quale appartengono anche membri internazionali. E' una associazione composta di molti membri a stretto contatto con istituzioni pubbliche e private (affiliata con la British Computer Society e con lo IASC, in contatto con la Royal Statistical Society, la Social Research Association and la Market Research Society).

Nel tempo ha esteso i suoi obiettivi ed ha deciso di divenire una fonte informativa circa gli ultimi pacchetti statistici e le più aggiornate tecniche statistiche disponibili.

Organizza una serie di conferenze e workshop più specialistici e dal 1999 ha deciso di usare il web come strumento principe per la pubblicazione: anche la sua principale pubblicazione, che è il *Register of Software for Statistical and Survey Analysis* (<http://www.asc.org.uk/Register/>) è divenuto un documento disponibile , è ora disponibile su web. In effetti l'ASC cura il Registro catalogando in dettaglio i pacchetti più noti, ritenuti di interesse per l'Associazione.

La pubblicazione avviene anche su carta, ma è disponibile solo per i membri (o a pagamento), mentre chiunque interessato può usufruire di queste informazioni online.  
Per maggior dettaglio si può leggere l'Appendice 2 "lista dei software del Register Asc".

Interessante è anche l'associazione ASA - American Statistical Association –  
<http://www.amstat.org/index.cfm>.

Anche questa associazione americana è molto antica, fondata nel 1839 con lo scopo di promuovere attività varie collegate alla applicazione della scienza statistica. Pubblica ad esempio il "Journal of Computational and Graphical Statistics", dove si cerca di stimolare l'uso dei metodi computazionali e grafici

Una specifica sessione dell'ASA è la SRMS - Survey Research Methods Section (<http://www.amstat.org/sections/SRMS/index.html>) che in particolare vuole stimolare proprio l'applicazione statistica, sia sotto l'aspetto teorico che applicativo e diffonderne le informazioni.

In particolare collabora in progetti utili per la ricerca, quali ad esempio, la creazione del "Summary of Survey Analysis Software" ( sito ritrovabile in quello della Harvard Medical school di Boston <http://www.hcp.med.harvard.edu/statistics/survey-soft/>) elenco eccellente, in cui anche qui si raccolgono molte informazioni inerenti i software statistici (si veda Appendice 3).

Anche qualche Istituto Nazionale di Statistica si può ritenere particolarmente attivo per quanto concerne lo sviluppo degli strumenti generalizzati.  
Nel seguito se ne evidenziano alcuni.

Statistics Canada (sito <http://www.statcan.gc.ca/>) che ha sviluppato software molto utilizzati quali:

- ACTR Automated Coding by Text Recognition è Sistema generalizzato per la codifica automatica di quesiti rilevati a testo libero.
- BANFF - software generalizzati per il controllo e la correzione dei dati (imputazione) per le variabili numeriche, continue e non-negative (versione SAS).
- CANCEIS Canadian Census Edit and Imputation System – Software per il controllo e correzione dati

Statistics Netherlands (il cui sito, già citato, è [www.cbs.nl](http://www.cbs.nl)) che ha sviluppato software altrettanto noti quali :

- BLAISE
- software Argus (moduli Mu-Argus, per i dati elementari, e Tau-Argus per le tabelle)

LO U.S. Census Bureau (<http://www.census.gov/>) che permette di effettuare il download di alcuni software di dominio pubblico (da Software to Download in <http://www.census.gov/main/www/access.html>) quali:

- VPLX
- CSPro
- IMPS

Anche l'Eurostat organizza interessanti convegni quali l'NTTS "New Techniques and Technologies for Statistics", che è una conferenza internazionale che vuole cogliere l'impatto delle nuove tecnologie inerenti l'acquisizione, produzione e diffusione dei dati.

Vuole stimolare e facilitare la collaborazione di progetti di cooperazione.

Le conferenze NTTS sono state organizzate nel 1992, 1995 and 1998 e l'ultima è stata congiunta con l'ETK (Exchange of Technology and Know-how) nel 2001 a Creta (*ETK 2001 – NTTS 2001 Exchange of Technology and Know-How, New Techniques and Technologies for Statistics*) in modo da congiungere conferenze diverse riguardanti le tecnologie: l'ETK infatti aveva lo scopo di creare un network che

identificasse e risolvesse i vari problemi che emergono tra i produttori delle statistiche ufficiali e i fornitori dei metodi e tool.

Tra le iniziative di cooperazione internazionale lanciate dall'Eurostat, si possono evidenziare i progetti che rientrano nell'ambito dell'ESSnet (European Statistical System network): con l'obiettivo di condividere le applicazioni software, si sta cercando di creare una architettura di riferimento comune ai paesi europei, prendendo in considerazione strumenti di diffusione internazionali quali il portale OSOR (Open Source Observatory and Repository), di attualissima implementazione e che ancora è in sviluppo cooperativo ed è nato nell'ambito del programma IDABC<sup>16</sup>: OSOR è una piattaforma per lo scambio di informazioni ma anche di esperienze, che si prefigge di divenire un portale attivo a livello europeo, includendo notizie su eventi, progetti, nuove applicazioni, così come informazioni circa raccomandazioni e standard da adottare per l'uso dell'OSS.

L'idea è quella di fungere da strumento che agisca come repository ma che serva anche a dare visibilità agli stessi progetti europei OS e alle iniziative delle diverse pubbliche amministrazioni (e Istituti Nazionali di Statistica) e che motivi le collaborazioni e i progetti di riuso.

---

<sup>16</sup> Il programma IDABC (Interoperable Delivery of Pan-European eGovernment Services to Public Administrations, Business and Citizens) mira a fornire servizi amministrativi paneuropei on-line alle amministrazioni pubbliche, alle imprese e ai cittadini. L'obiettivo è di migliorare l'efficienza delle amministrazioni pubbliche europee e la collaborazione fra di loro.



## APPENDICE 1: Lista conferenze internazionali ISI / IASC

- IASC -

<http://www.iasc-isi.org/Activities.html>

IASC provides international contacts among statisticians, computing professionals, organizations, institutions, governments, and the general public world-wide and fosters cooperation and interchange between them. Particularly, this is done by organizing as well as sponsoring Conferences, by establishing Regional Sections, by organizing Summer Schools and by maintaining a Mailing List.

Currently, two regional sections are established: the European Regional Section (ERS), which organises the biennial COMPSTAT meeting, and the Asian Regional Section (ARS), which organises the Asian Conference every three years. IASC seeks to extend its regional presences in all other parts of the world. There are close contacts to the Interface Foundation of North America and IASC is regularly sponsoring a session at the annual INTERFACE meeting. IASC particularly works on its outreach to the Middle East and South America.

### Forthcoming Events

57th ISI Session            2009    Durban (South Africa)

### Previous IASC World Conferences

1st IASC World Conference    1987    Japan

2nd IASC World Conference    1997    Pasadena (CA, USA)

3rd IASC World Conference    2005    Cyprus (Grece)

4th IASC World Conference & 6th IASC-ARS Conference    December 5-8, 2008    Yokohama (Japan)  
(ASC - Asian Regional Section of the IASC )

### Previous COMPSTAT Conferences

I COMPSTAT            1974    Vienna (Austria)

II COMPSTAT            1976    West-Berlin (Germany)

III COMPSTAT            1978    Leiden (The Netherlands)

IV COMPSTAT            1980    Edimbourgh (UK)

V COMPSTAT            1982    Toulouse (France)

VI COMPSTAT            1984    Prague (Czechoslovakia)

VII COMPSTAT            1986    Rome (Italy)

VIII COMPSTAT            1988    Copenhagen (Denmark)

IX COMPSTAT            1990    Dubrovnik (Yugoslavia)

X COMPSTAT            1992    Neuchâtel (Switzerland)

XI COMPSTAT            1994    Vienna (Austria)

XII COMPSTAT            1996    Barcelona (Spain)

XIII COMPSTAT            1998    Bristol (UK)

XIV COMPSTAT            2000    Utrecht

XV COMPSTAT            2002    Berlin (Germany)

XVI COMPSTAT            2004    Prague (Czech Republic)

XVII COMPSTAT\        2006    Rome (Italy)

XVIII COMPSTAT        August 24-29, 2008    Porto (Portugal)

### Previous ARS Conferences on Statistical Computing - (ARS - Asian Regional Section of the IASC)

1st ARS Conference    October 1993    Beijing (China)

2nd ARS Conference    August 1996    Sydney (Australia)

3rd ARS Conference    December 1998    Manila (Philippines)

4th ARS Conference    December 2002    Busan (Korea)

Previous ERS Summer Schools (ERS - European Regional Section of the IASC)

1st Summer School	1991	Prague (Czechoslovakia)
2nd Summer School	1993	Perugia (Italy)
3rd Summer School	1995	Izmir (Turkey)
4th Summer School	1997	Padua (Italy)
5th Summer School	1999	Spetses (Greece)
6th Summer School	2001	Capri (Italy)
7th Summer School	2003	Lisbon (Portugal)
8th Summer School	2005	Barcelona (Spain)
9th Summer School	2007	Island of Procida (Italy)

## **APPENDICE 2: Lista del software del register ASC - <http://www.asc.org.uk/Register/index.htm>**

### List of Packages

This is an alphabetical list of all the packages in the Register. Click on a package name to go to the details for that package, then use your 'Back' button to get back here.

#### Group Packages

##### A (14)

AccessPoint, ACE, adenquire.net, Alpha+, AMOS, Analyse-it for Microsoft Excel, Answertree, Ascribe, 2ask, Askia, Asteroid, Atlas Planning Suite, auditHOST, AYN eSurvey Server

##### B (13)

[B/D], Bellview® CAPI , Bellview® CATI, Bellview® Fusion, Bellview® Scan, Bellview® Web, Beyond 20/20, Beyond Question, 3BI Monitoring, Blaise, Blue/Feedback, BMDP, Brand2hand

##### C (15)

CallWeb, Ccount, Clementine, CleverForm, Clicktools, Code Cogs, Collect Data Now!, Community Express, Complete Report Automation, Conformat, Converso, CRM Q+, CRM Qi+, CSPro, CustomView360

##### D (8)

DataFriend, DataSet V, DatStat Illume, 360 Degree Feedback, Demographix, Digivey, Dimensions - Survey Research Platform, Dub Interviewer

##### E (11)

EasyFit, Easyresearch - online surveys, Entryware, EquivTest, ESBPdf Analysis, ESBSStats, eStore, E-Tabs Enterprise, E-Tabs Interactive, E-Tabs Reader - Professional / Lite, E-Tabs Writer

##### F (5)

FieldCenter, Fitter, ForexastX Wizard, ForSurveys, FUSE Enhanced Survey Solutions

##### G (5)

GenStat, gllamm, Globalpark, Grader, GS+ Geostatistics for the Environmental Sciences

##### H (2)

Halogen eSurveyor, Hosted Survey

##### I (16)

IDAMS, IdeaMap.NET, Inquisite, InSite, Instant Intelligence, Interviewer CAPI, Interviewer CATI, Interviewer VCC, Interviewer Web, Ioxphere, iQuest, IRQuest, ISSA, IT, IYCQ Manager, IYSS

##### J (1)

Jambo

##### K (4)

KeyPlan, Keypoint 4.0, Key Survey, KnowledgeWorx

##### L (1)

LimeSurvey

##### M (21)

MaceTech Options Analysis, MaceTech Sabre Analysis, MarketSight®, MARSC, Mentor, Merlin, MERLIN FASTAB, MERLINPLUS, MERLIN TOOLKIT, MERLIN to PowerPoint (MTPP),

Microtab, MIM, Mobile Researcher, MRDCL, mrInterview, mrPaper, mrScan, mrStream, mrTables, mrTranslate, mTAB

N (8)  
 NCSS, Nesstar, NetEpi, NetTeam.Survey, NIPO Interview System, nQuery Advisor, NSDstat, NVivo 8

O (8)  
 OfficeReports, OnePoint, Online Surveys, OnTraq, Opinio, Opinion-8, Opinion Taker, Origin Survey Manager

P (15)  
 PASS, Patient Profiles, Patient Satisfaction Survey Database, PerceptionStream, polliscope, PowerCATI system, powerTABLE, Preport, Print & Scan, Procevo, Prospex, P-Stat, Pulsar, Pulsar Web, PulseEFM

Q (9)  
 QDA Miner, QDS, QEDML Designer, QEDML Web Server, QPSMR, Qualrus, QuestionPro, Questor, Qure DMP

R (12)  
 Raosoft EZSurvey, Raosoft EZSurvey SMS, REDATAM, REDATAM WEBSERVER, Remark Office OMR, Remark Web Survey, Report Direct, Resampling stats, Research Reporter, Rogator G4, Rostock Survey Tool, Ruby

S (51)  
 SAS/STAT, SC, Self Coder, SENATE-Survey, SensorPro, Simstat, SIR, Snap Campus, SNAP Internet module, snap Keystation Edition, SNAP Network Interviewing module, snap PDA Interviewer Module, snap Professional Edition, snap ProNet Edition, SNAP Results Edition, snap Scanning Module, snap SurveyPaks, Snap Survey Software, SOLAS, Sphinx, S-Plus, SPSS Base and Modules, SPSS Text Analysis for Surveys, SPSS WebApp, STAR®, STARS - Survey Tabulation And Reporting System, Stata, Statistica, STATISTICA Data Miner, STATISTICA QC Miner, Statistix, StatPac for Windows, StatXP, StudyResult, StudySize, SumQuest, SuperSTAR, Survent, Survey Budgeting System, surveyclip.com, Survey Explorer, Survey Galaxy, SurveyGold Survey Software, Survey Manager - Encuestas Online, Surveyor, SurveyPro 3.0, Survey Quick, SurveySolutions, Survey Solutions, SYSTAT, SySurvey

T (11)  
 Tables Direct, Talent360, Team360, TeleForm, TextGrab, TEXTPACK, TextQuest, TPL Tables, TricTrac, TSP 4.5, TSS - The Survey System

U (1)  
 Unistat

V (7)  
 Vector, VentureFeedback.com, Video Survey Software, ViewsCount, ViewSuite360, Visual QSL, Vitalnet

W (6)  
 WARP-IT, WebMine, WebSTATISTICA, WebSurveyor, Winsteps, Winyaps

X (5)  
 XLMiner, XLStatistics, XpertSuite, XploRe, XSight

## APPENDICE 3: Summary of survey analysis software -

<http://www.hcp.med.harvard.edu/statistics/survey-soft/>

### Summary of Survey Analysis Software



This page is a summary of available software for the analysis of surveys with complex sample designs. Specifically, it includes software that can do variance estimation with such survey data. It does not attempt coverage of the many excellent packages which do not have the particular features required to analyse such data. Information can be viewed by package or comparatively by topic.

This project has been undertaken with the encouragement of the Section on Survey Research Methods, American Statistical Association, but the Section has no responsibility for the content. The descriptions on these pages are based on information provided by the vendors/developers of this software or extracted from their manuals and home pages. No mention or information on these pages should be interpreted as endorsement of any software by the ASA or by the site editor. (DISCLAIMER)

Contributions to this page are welcome: corrections, information on other software, comparative review articles, and additions to the bibliography.

This page includes links to

- \* a list of analysis software packages;
- \* the same information organized by features;
- \* comparative reviews: on-line text and references to reviews in print;
- \* a UCLA site with examples of code for analysis of survey data using four packages;
- \* the PEAS (Practical Exemplars on the Analysis of Surveys) site at Napier University, with examples and reviews for R, SAS, SPSS, and Stata, and some text on theory.
- \* Other software references of interest for survey analysts, including software for
  - o hierarchical modeling,
  - o imputation of missing data,
  - o disclosure control, and
  - o survey administration.

#### Packages

- \* AM Software from American Institutes for Research.
- \* Bascula from Statistics Netherlands.
- \* CENVAR from U.S. Bureau of the Census.
- \* CLUSTERS from University of Essex.
- \* Epi Info from Centers for Disease Control.
- \* Generalized Estimation System (GES) from Statistics Canada.
- \* IVEware from University of Michigan.
- \* PCCARP from Iowa State University.
- \* R survey package from the R Project.  
(See also the pps package for some calculations for PPS designs.)
- \* SAS/STAT from SAS Institute.
- \* SPSS Complex Samples from SPSS Inc.
- \* Stata from Stata Corporation.

- \* SUDAAN from Research Triangle Institute.
- \* VPLX from U.S. Bureau of the Census.
- \* WesVar from Westat, Inc.

#### Comparative summary by topic

- \* Sampling designs for which analyses are made
- \* Estimands and types of analyses
- \* Methods for variance estimation
- \* "Feel", interface style
- \* Platforms for which available
- \* Availability and pricing
- \* Contact information

#### Review articles: on-line

\* "Sampling Error Software for Personal Computers" by Jim Lepkowski and Judy Bowles, reprinted from *The Survey Statistician: a general review of the reasons why special software is needed to analyse survey data, and a description of eight packages for the PC.*

\* "An Evaluation of Alternative PC-Based Packages for the Analysis of Complex Survey Data," by Steven B. Cohen (1997), *The American Statistician*, 51, 285-292. (Postscript file, Adobe PDF file - note that pagination is close to but not precisely identical with published version)

\* "Comparison of Variance Estimation Software and Methods", a report prepared by a consortium of the UK Office for National Statistics, Statistics Sweden, the University of Southampton and the University of Bath, under the sponsorship of Eurostat. (See title page for full author list.) PDF file, 41 pages, released 1999. General discussion of principles, and comparison of 5 packages with detailed description of capabilities.

\* "Software for Statistical Analysis of Sample Survey Data" by Barbara Lepidus Carlson, reprinted with permission from *Encyclopedia of Biostatistics*, Wiley, 1998. A brief discussion of variance estimation in surveys and references for a number of packages.

\* "Pitfalls of Using Standard Statistical Software Packages for Sample Survey Data" by Donna Brogan, reprinted with permission from *Encyclopedia of Biostatistics*, Wiley, 1998. An example comparing analyses using standard software to that using software that recognizes special features of survey design.

\* Introduction to Survey Analysis from UCLA Academic Technology Services defines some key concepts of survey analysis and gives examples from several packages. See also their Survey Analysis Portal for additional links.

\* Analysis of Survey Data from Household Surveys in Developing and Transition Countries from the United Nations. See Chapter 21 (by Donna Brogan) for software examples and reviews.

#### Review articles: in print (since 1989) and not specific to a single package.

\* Rao, Kumar and Roberts (1989), "Analysis of sample survey data involving categorical response variables: Methods and software," *Survey Methodology*, 15, 161-186.

\* Rowe, Westlake, and Rose (1990), "Software for statistical and social survey analysis 1989-90," *Computational Statistics and Data Analysis*, 9, 317-340.

\* Bausch and Bankhofer (1992), "Statistical software packages for PCs - A market survey," *Statistischen Hefte*, 33, 283-306.

\* Chell (1992), "A comparison of some software packages for survey analysis," in Westlake, Banks, Payne, and Orchard, *Survey and Statistical Computing*, Elsevier/North-Holland, 385-394.

*(Please send comments and suggestions regarding this page to Alan Zaslavsky, [zaslavsk@hcp.med.harvard.edu](mailto:zaslavsk@hcp.med.harvard.edu))*

## **APPENDICE 4: Link utili per il sito dell'OTS e per informazioni sulla licenza EUPL**

*OTS link accessibile da internet (sito dell'Istituto)*

<http://www.istat.it/strumenti/metodi/software/osservatorio.html>

*OTS link accessibile dalla intranet (interna all'Istat)*

[http://intranet.istat.it/doutdes/SW\\_statistico\\_1/repository/index](http://intranet.istat.it/doutdes/SW_statistico_1/repository/index)

*Link informazioni licenza EUPL*

<http://ec.europa.eu/idabc/eupl>

<http://www.osor.eu/eupl>

[http://labos.diee.unica.it/labos/resources/cms/documents/Facilitating\\_EUPL\\_\\_DIGIT\\_Garcia\\_Moran.pdf](http://labos.diee.unica.it/labos/resources/cms/documents/Facilitating_EUPL__DIGIT_Garcia_Moran.pdf)

[http://labos.diee.unica.it/labos/resources/cms/documents/Legal\\_context\\_EUPL\\_\\_dusollier.pdf](http://labos.diee.unica.it/labos/resources/cms/documents/Legal_context_EUPL__dusollier.pdf)

[http://www.publicsectoross.info/images/resources/7\\_312\\_file.pdf](http://www.publicsectoross.info/images/resources/7_312_file.pdf)

[http://labos.diee.unica.it/labos/resources/cms/documents/key\\_features\\_and\\_evolution\\_of\\_EUPLv1\\_\\_Laurant.pdf](http://labos.diee.unica.it/labos/resources/cms/documents/key_features_and_evolution_of_EUPLv1__Laurant.pdf)

<http://ec.europa.eu/idabc/servlets/Doc?id=27472>





## Documenti ISTAT(\*)

- 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 Franciscantonio, 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 Informatico 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: 1 - Analisi, valutazione e proposte in merito ad atti di normazione e finanziamento*
- 6/2007 – Silvia Bruzzone, Atonia 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*
- 9/2007 – Damiano Abbatini, Lorenzo Cassata, Fabrizio Martire, Alessandra Reale, Giuseppina Ruocco e Donatella Zindato – *La progettazione dei Censimenti generali 2010-2011 2 - Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani*
- 10/2007 – Marco Fortini, Gerardo Gallo, Evelina Paluzzi, Alessandra Reale e Angela Silvestrini – *La progettazione dei censimenti generali 2010 – 2011 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*
- 11/2007 – Domenico Adamo, Damiana Cardoni, Valeria Greco, Silvia Montecolle, Sante Orsini, Alessandro Ortensi e Miria Savioli – *Strategie di correzione del questionario sulla qualità della vita dell'infanzia e dell'adolescenza. Indagine multiscope sulle famiglie. Aspetti della vita quotidiana 2005*
- 12/2007 – Carlo Nappi – *Manuale per la preparazione di originali "ready to print"*
- 1/2008 – Franco Lorenzini – *Indagine sulle unità locali delle imprese: la flessibilità organizzativa e il ruolo degli uffici regionali come strategia per la riduzione del disturbo statistico e il raggiungimento di elevati tassi di risposta*
- 2/2008 – Elisa Berntsen, Simone De Angelis, Simona Mastroluca – *La progettazione dei Censimenti generali 2010-2011 4-L'uso dei dati censuari del 2000-2001: alcune evidenze empiriche*
- 3/2008 – Marina Peci – *Progetto SCQ -Scuola Conoscenza Qualità-Statistica e Studenti*
- 4/2008 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zel – *Messa a regime dell'uso dei dati fiscali (Modelli UNICO) per l'integrazione delle mancate risposte e la riduzione del numero delle unità campione della rilevazione PMI*
- 5/2008 – Giovanni Seri e Maurizio Lucarelli – *A.D.ELE. Il laboratorio per l'Analisi dei Dati ELEmentari. Monitoraggio dell'attività Anni 2004-2007*
- 6/2008 – Francesco Altarocca – *Strumenti informatici innovativi nella conduzione di indagini statistiche*
- 1/2009 – Silvia Dardanelli, Simona Mastroluca, Alessandro Sasso e Mariangela Verrascina – *La progettazione dei censimenti generali 2010 – 2011 5 - Novità di regolamentazione internazionale per il 15° Censimento generale della popolazione e delle abitazioni*
- 2/2009 – Rossana Balestrino e Alberto Gaucci – *Tecniche di cattura dati nei processi di produzione statistica*
- 3/2009 – Barbara Fiocco – *Le "misure" dell'Italia nell'Annuario Statistico Italiano*
- 4/2009 – Daniela Pagliuca, Raffaella Cianchetta, Marco Broccoli, Teresa Buglielli, Roberto Di Giuseppe e Diego Zardetto – *L'Osservatorio Tecnologico per i Software generalizzati (OTS) nel 2008*