

importante per l'analisi congiunturale in quanto consente di valutare la dinamica degli aggregati economici a prescindere dalla distribuzione in corso d'anno del numero di giorni lavorativi.

Come di consueto, inoltre, è stata prodotta ed inviata ad Eurostat una tavola semplificata del Conto delle amministrazioni pubbliche; entro settembre sono state trasmesse le serie storiche disaggregate per branca di attività economica ed entro il mese di dicembre sono state fornite le tavole relative alle imposte per categoria e tipo, alla spesa pubblica per funzione e ai conti non finanziari dei settori istituzionali. Inoltre, sono state aggiornate le stime del consumo individuale e del consumo collettivo, nonché degli investimenti per prodotto e branca proprietaria, dello stock di capitale e degli ammortamenti. Sono state, inoltre, elaborate e diffuse le serie degli occupati, delle posizioni e delle Ula regolari e irregolari a livello nazionale e regionale. Con riferimento alla componente irregolare va tenuto presente che la contabilità nazionale italiana, al pari di quella degli altri Paesi dell'Unione europea, segue gli schemi e le definizioni del Sec95 che impongono di contabilizzare nel Pil anche l'economia non direttamente osservata. L'Istituto statistico dell'Unione europea (Eurostat), vigila sul rispetto del Sec e sulla bontà delle metodologie adottate dagli Stati membri, accertandone e certificandone la validità, anche e soprattutto in relazione alla capacità di produrre stime esaustive del Pil. In questo contesto sono stati diffusi i dati che riguardano l'analisi delle attuali stime del Pil e dell'occupazione attribuibile alla parte di economia non osservata costituita dal sommerso economico, cioè derivante dall'attività di produzione di beni e servizi che, pur essendo legale, sfugge all'osservazione diretta in quanto connessa al fenomeno della frode fiscale e contributiva.

E' stato diffuso per la prima volta, il Conto economico trimestrale delle amministrazioni pubbliche (Ap), in anticipo rispetto a quanto stabilito in sede comunitaria. In base ai regolamenti approvati nel 2000 e nel 2002, tutti i paesi membri devono produrre, infatti, i Conti economici trimestrali delle Ap entro il 2005, anno in cui ha termine la fase sperimentale ed Eurostat renderà disponibili anche i conti degli altri paesi europei.

Per quanto concerne i "Conti e le analisi territoriali" sono state completate e rilasciate le nuove stime regionali relative agli anni 1995-2002, allineate con le stime annuali diffuse a marzo 2003. Il set completo dei dati relativi ai conti regionali ha compreso, oltre al conto delle risorse e degli impieghi e al conto della distribuzione del reddito, anche le analisi a 25 branche (derivate dalla classificazione Nace-Rev.1) su valore aggiunto, redditi da lavoro dipendente, retribuzioni lorde, contributi sociali effettivi e figurativi, investimenti fissi lordi, occupati interni (dipendenti e indipendenti) e corrispondenti unità di lavoro. I consumi delle famiglie sono disaggregati in 12 gruppi di beni e servizi; i consumi delle amministrazioni pubbliche in 10 funzioni di spesa. Sono state, inoltre, pubblicate le stime provinciali, per gli anni 1995-2001, relative agli occupati interni, alle unità di lavoro e al valore aggiunto ai prezzi base. Tali dati incorporano una revisione delle stesse, relativamente agli anni 1999 e 2000, resa necessaria, sia per il riallineamento con i dati nazionali (diffusi a marzo 2003) e regionali (diffusi a settembre 2003), sia per l'opportunità di utilizzare un'informazione statistica di base più completa. I dati sono stati trasmessi all'Eurostat in anticipo rispetto alle scadenze imposte dal Regolamento Sec95. Quest'ultimo prevede che, a distanza di 24 mesi dall'anno di riferimento delle stime, i paesi dell'Unione Europea provvedano al rilascio dei dati, analizzati in tre macrobranche (agricoltura, industria e servizi), al livello 3 della nomenclatura delle unità territoriali statistiche che, per l'Italia, corrisponde alla disaggregazione del territorio nazionale nelle 103 province.

Nel contesto della realizzazione del progetto "Informazione statistica territoriale e settoriale per le politiche strutturali 2001-2008" finanziato nell'ambito del Quadro comunitario di sostegno - Obiettivo 1-, in base al modello econometrico messo a punto in collaborazione con l'Università di Udine, è stata prodotta una stima anticipata della dinamica di alcuni aggregati economici nelle grandi ripartizioni geografiche che costituisce un'analisi territoriale preliminare delle serie di contabilità nazionale per l'anno 2002. Gli aggregati presi in considerazione sono: unità di lavoro, valore aggiunto, prodotto interno lordo (Pil), consumi finali interni e spesa delle famiglie per consumi finali. Inoltre, nello stesso contesto, l'Istat rende disponibili per la prima volta le stime, relative agli anni 1996-2000, degli occupati interni e del valore aggiunto nei Sistemi locali del lavoro (Sll), disaggregate per tre macro-branche di attività economica (agricoltura,

industria e servizi, secondo la definizione del Sec95). La dimensione territoriale scelta è quella dei 784 Sistemi locali del lavoro (Sll), costruiti dall'Istat sulla base dei flussi di pendolarismo per motivi di lavoro. Si tratta di aree svincolate dagli usuali criteri amministrativi e definite dall'organizzazione dei rapporti sociali ed economici, all'interno delle quali esiste la massima coincidenza tra domanda e offerta di lavoro. La realizzazione delle stime dei vari aggregati a questo livello territoriale è resa possibile grazie ai significativi miglioramenti intervenuti nei dati di base. In particolare, per ciascun anno considerato si è provveduto alla costruzione di un data set informativo, ovvero un archivio integrato realizzato attraverso l'uso di molteplici fonti statistiche ed amministrative, con informazioni disaggregate a livello di unità locale. Nella produzione di dati per Sll rientrano anche le stime del numero di "persone in cerca di occupazione" e di "occupati residenti", le cui basi informative provengono invece dall'indagine trimestrale sulle forze di lavoro.

E' continuata la costruzione del prototipo di una banca dati comunale che integra dati di origine amministrativa e di tipo censuario e che consente l'estrazione di tali dati secondo varie articolazioni territoriali (regioni, province, comuni, sistemi locali del lavoro, ecc.).

Infine, molto è stato fatto sul fronte della costruzione, verifica, implementazione e aggiornamento degli indicatori di dotazione e di performance delle infrastrutture che ammontano a oltre 200 indicatori, tutti costruiti su base provinciale e in serie storica dal 1996 al dato più recente disponibile.

Per quanto concerne i conti economici nazionali per settore istituzionale sono state presentate le stime, aggiornate al 2002, per gli anni 1990-2002. Essi illustrano in maniera sistematica e integrata i comportamenti dei diversi operatori nei momenti essenziali del processo economico: produzione; formazione; distribuzione; redistribuzione e utilizzazione del reddito; accumulazione finanziaria e non finanziaria.

E' stato diffuso il nuovo schema input-output previsto dal Sistema europeo dei conti 1995. In tale schema il quadro delle interdipendenze tra gli operatori economici è completamente rivisto e comprende due tipi di tavole principali: le tavole delle risorse e degli impieghi (Sut); e le tavole input output simmetriche (Siot).

Le tavole delle risorse e degli impieghi sono matrici per branca di attività economica e per prodotto che descrivono dettagliatamente tutti i processi di produzione interni e tutte le operazioni sui prodotti dell'economia nazionale. Nel nuovo sistema input-output la tavola simmetrica è soltanto una tavola derivata; una matrice cioè, prodotto per prodotto o branca per branca, che ricompone in una singola tavola, attraverso un opportuno algoritmo, le informazioni provenienti dalle tavole delle risorse e degli impieghi.

Riguardo ai conti ambientali è proseguita l'attività di analisi e approfondimento delle singole tematiche in sede nazionale ed internazionale. In particolare sono stati realizzati: nell'ambito dei Conti dei flussi di materia, un bilancio materiale per il 1997 e una serie storica 1980-98 di indicatori relativi agli input fisici dell'economia; nell'ambito della Namea, conti delle emissioni per gli anni 1990-94; nell'ambito dell'Epea, un conto pilota per il 1997. I lavori relativi alla prima e alla terza tematica sono stati co-finanziati dall'Eurostat.

## **8.5 Metodologie statistiche**

Nel corso del 2003 all'interno dell'Istituto nazionale di statistica è si proceduto ad una profonda ristrutturazione dei settori le cui attività riguardano la ricerca e lo sviluppo nel campo delle metodologie statistiche.

Prima del luglio 2003, più settori erano responsabili, all'interno dell'Istat, di tali attività: il Servizio Metodologia di base per la Produzione statistica (Mps), ed il Servizio Coordinamento, integrazione e qualità (Ciq), entrambi interni al Dipartimento Integrazione e standard tecnici (Dist); nonché l'Unità operativa Metodologia per le indagini sociali (Cin/E) del Dipartimento delle Statistiche sociali (Diss) e l'Unità operativa per il Coordinamento della ricerca metodologica (Dise/G) del Dipartimento delle statistiche economiche (Dise).

La ristrutturazione ha portato alla creazione di due nuovi servizi, all'interno della Direzione centrale per le metodologie statistiche e le tecnologie informatiche (Dcmt):

- il Servizio Mts (Metodologie, tecnologie e software per la produzione dell'informazione statistica);

- il Servizio Psm (Progettazione e supporto metodologico nei processi di produzione statistica. Questi nuovi servizi hanno costituito il contenitore nel quale sono transitate tutte le strutture metodologiche precedentemente esistenti, con l'eccezione della struttura Dcmt/C (Metodologie e sistemi di supporto all'integrazione), che corrisponde alla vecchia unità Ciq/B.

Il Servizio Mts unifica alcuni settori di ricerca e sviluppo metodologico (essenzialmente, quelli relativi alla qualità dei dati ed alle fasi di acquisizione e trattamento), con settori di ricerca e sviluppo concernenti software generalizzato, sistemi informativi statistici e tematiche legate agli aspetti tecnologici informatici dei processi di produzione. Gli obiettivi prioritari di questo servizio, nel campo metodologico, sono quelli di proseguire l'attività di ricerca e sviluppo e di assicurare la definizione e la condivisione di standard e pratiche raccomandate all'interno dell'Istituto e, in prospettiva, per l'intero Sistan.

Per quanto riguarda i *Sistemi per la codifica automatica e l'acquisizione assistita* dei dati (struttura Mts/H), è proseguita l'attività di ricerca e sviluppo su entrambi i fronti.

Le attività inerenti l'acquisizione dei dati assistita da computer hanno riguardato principalmente tre linee progettuali: l'estensione della cosiddetta strategia *'in-house'*, già adottata per l'indagine sulle nascite, ad altre rilevazioni Cati, l'internalizzazione del sistema Capi dell'indagine sulle forze di lavoro, il supporto ad utenti interni ed esterni all'Istat per le attività inerenti l'acquisizione dati in modalità Cati/Capi/Cadi<sup>1</sup>. Relativamente alla strategia *'in-house'*, che prevede di affidarsi ad una ditta esterna soltanto per la messa a disposizione del *call center* e la realizzazione delle interviste, ma di progettare e sviluppare con risorse interne tutto il software per la fase di acquisizione, sono iniziate nel 2003 le attività inerenti la progettazione e l'analisi dei sistemi software per le indagini su 'Inserimento professionale dei laureati' e su 'Percorsi di studio e di lavoro dei diplomati'.

Per quanto attiene l'internalizzazione del sistema Capi dell'indagine sulle forze di lavoro, nel corso dell'anno è stato progettato e sviluppato un prototipo circoscritto ai moduli funzionali all'internalizzazione del sistema Capi *client* (sistema che integra l'agenda dei rilevatori, i contatti ed il questionario elettronico).

Infine è stato fornito supporto agli utenti interni, relativamente alle fasi di progettazione, sviluppo e messa in produzione di un'applicazione Cadi per l'Indagine sulla Struttura e sulle produzioni delle aziende agricole 2003. L'adozione della registrazione controllata è stata introdotta al fine di elevare la qualità dei dati registrati e di garantirne un livello omogeneo rispetto a tutti i soggetti delegati a svolgere questa funzione. L'applicazione è stata infatti fornita all'esterno, sia alle regioni che effettuano autonomamente la registrazione dei dati, che alla ditta esterna che completa l'acquisizione per le restanti regioni.

E' stata inoltre fornita assistenza a due soggetti esterni: la Regione Toscana e gli uffici di statistica della Bosnia Erzegovina.

Il supporto dato alla Regione Toscana ha riguardato la progettazione dei questionari di indagine in funzione della tecnica di rilevazione Cati ed il sistema software Blaise, che la Regione Toscana adotta per le proprie rilevazioni telefoniche.

Per la Bosnia Erzegovina, invece, si è lavorato nell'ambito di un progetto di cooperazione al fine di progettare, sviluppare e fornire un'applicazione software (e trasmettere il relativo know-how) per la registrazione controllata (Cadi) dei dati dell'indagine sui consumi delle famiglie.

Relativamente alla tematica inerente la codifica di variabili rilevate a testo libero, le attività hanno riguardato due aspetti: la codifica *automatica* (da eseguirsi in modalità *batch*) e la codifica *assistita* (quale supporto al codificatore manuale durante o dopo la rilevazione).

Per quanto concerne la codifica automatica, per cui viene adottato il sistema software Actr, avendo messo in produzione le applicazioni di codifica automatica delle variabili 'Professione', 'Attività economica', 'Titolo di studio', 'Stato estero' e 'Comune', è stata fornita assistenza nell'ambito del progetto di codifica di queste variabili testuali rilevate nell'ambito delle Convivenze nell'ambito del Censimento della popolazione e delle abitazioni. Tale supporto, oltre all'aspetto informatico, ha riguardato l'aggiornamento delle basi informative in funzione degli spunti informativi derivanti dalle risposte fornite dagli intervistati.

---

<sup>1</sup> CATI: Computer Assisted Telephone Interviewing; CAPI: Computer Assisted Personal Interviewing; CADI: Computer Assisted Data Input

Riguardo la codifica assistita, è stato predisposto l'ambiente applicativo in Blaise per la codifica della professione, finalizzato soprattutto all'utilizzo di questa funzione nel corso di rilevazioni Cati. E' stato effettuato un test per valutare i risultati non soltanto dal punto di vista qualitativo, ma anche in funzione della utilizzabilità di questo modulo nel corso di un'intervista telefonica, dipendente soprattutto dal suo impatto sulla durata dell'intervista. Il test ha dato risultati positivi sotto entrambi i punti di vista.

Per quanto riguarda i *Metodi per il trattamento degli errori non campionari* (struttura Mts/I) sono state effettuate le attività descritte di seguito.

Nell'ambito del progetto *Nuovi strumenti per il controllo e l'imputazione degli errori e delle mancate risposte parziali*, le attività hanno riguardato: 1) lo studio di nuove metodologie per il controllo e l'imputazione degli errori e delle mancate risposte parziali, 2) nel caso non fosse già disponibile il relativo software, la loro implementazione in strumenti che rendessero possibile sia le attività di test e valutazione, sia la diffusione dei metodi all'interno dell'istituto.

Diverse metodologie sono state oggetto di studio e valutazione. Sono state applicate sperimentalmente tecniche di imputazione delle mancate risposte parziali basate su modelli di regressione, inclusi i modelli per l'imputazione multipla disponibili nel software *IVEware* per dati di tipo misto e il metodo noto come *Predictive Mean Matching* per variabili di tipo quantitativo. Sono state sottoposte a valutazione le tecniche di controllo e imputazione disponibili nel software *Banff*<sup>2</sup> per l'individuazione di errori casuali in variabili numeriche continue. In questo ambito, è stata completata un'attività di stage per la valutazione comparativa sperimentale dell'accuratezza di diverse tecniche di imputazione attualmente disponibili in Istat, tra cui lo stesso *Banff*. Sono state studiate e sperimentate tecniche di individuazione di errori sistematici in dati quantitativi basate sull'uso di *modelli mistura*. Sono state studiate e valutate sperimentalmente tecniche di imputazione supportate dall'uso degli algoritmi per la costruzione di alberi di regressione disponibili nel software *Waid*. E' proseguita la valutazione delle tecniche di imputazione delle mancate risposte basate sull'uso di reti bayesiane per dati di tipo qualitativo. E' stata condotta un'attività di tutoraggio per un tirocinio Mami-Istat (Mami e' un master in applicazioni della matematica nell'Industria e nei servizi che si svolge presso l'Universita' di Milano Bicocca). Titolo del tutoraggio: "Uso di tecniche Bayesiane nella statistica ufficiale". Per quanto riguarda lo sviluppo di strumenti software, è stato realizzato il software *Quis* per l'imputazione di valori mancanti in dati quantitativi utilizzando diversi metodi: (i) regressione multivariata basata sull'uso dell'algoritmo (*Expectation-Maximization*), (ii) *predictive mean matching*, (iii) *donatore di distanza minima*, (iv) *imputazione multipla*. E' stato inoltre fornito supporto metodologico alla realizzazione di un software generalizzato per l'imputazione di mancate risposte parziali in variabili miste (categoriche e numeriche continue) con donatore di minima distanza.

Per quanto riguarda attività di ricerca svolte nell'ambito della collaborazione a progetti europei, sono state completate le attività di valutazione e documentazione nell'ambito del progetto *Euredit*, che si è chiuso nel 2003, con la produzione dei deliverables finali sulle sperimentazioni svolte dai componenti della struttura. E' proseguita la collaborazione alle attività del gruppo internazionale *Un/EceE Working Group on Statistical Data Editing*, che si è concretizzata nella presentazione di una relazione invitata e di un contributo alla Conferenza svoltasi nel 2003 e nella predisposizione di un lavoro per l'*Un Statistical Journal*.

Un primo sottoprodotto dell'attività di ricerca e sviluppo di strumenti è rappresentato dalle attività di supporto alle indagini per la progettazione e la realizzazione di procedure di controllo e correzione, ed in particolare per l'integrazione nei processi di correnti di trattamento dei dati di nuovi strumenti o metodi già sottoposti a sperimentazione e valutazione. In particolare, è stata completata la procedura di controllo e correzione dei dati per il questionario sulle Reti di distribuzione del Sistema delle indagini sulle acque. E' stato dato supporto metodologico e operativo alla realizzazione della procedura di controllo e correzione della Rilevazione integrativa delle scuole secondarie di II grado. Sono state avviate le attività di supporto al disegno, alla realizzazione e al test della procedura di controllo e correzione per l'indagine Eu-

<sup>2</sup> Il software *Banff* è una versione di *Geis* (Generalised Editing and Imputation System) di Statistics Canada, utilizzabile in ambiente *Sas*

Silc su reddito e condizioni di vita. Sono state avviate le attività di supporto al disegno e alla realizzazione della procedura di controllo e correzione per l'indagine su Struttura e produzione delle aziende agricole.

E' infine proseguita l'attività di diffusione delle conoscenze acquisite attraverso l'attività di ricerca all'interno dell'Istat. In particolare, è stato effettuato un corso di formazione interno sui metodi e le funzionalità del software Banff.

Per quanto riguarda il progetto sulla *valutazione dei metodi di controllo e imputazione degli errori e delle mancate risposte parziali*, le attività hanno riguardato l'integrazione nel software Idea (*Indexes for Data Editing Assessment*) di nuovi indicatori e di nuove funzionalità (soprattutto per la parte Sidi). Il software Idea consente per l'appunto il calcolo di una serie di indicatori per la valutazione degli effetti dell'applicazione di procedure di controllo e correzione dati, inclusi gli indicatori di qualità previsti in Sidi. E' stata inoltre aggiornata la relativa documentazione.

Per quanto riguarda le attività svolte nell'ambito del progetto *Integrazione di dati da diverse fonti*, è proseguito lo studio delle tecniche di *matching statistico* per la ricostruzione di informazioni mancanti mediante integrazione di dati da diverse fonti, statistiche o amministrative. I risultati di tale studio sono stati illustrati in una relazione invitata alla Conferenza internazionale Isi 2003. Dal punto di vista applicativo, in collaborazione con altri settori di ricerca dell'Istituto, tecniche di *matching statistico* sono state applicate per la costruzione della Sam (*Social Accounts Matrix*). I risultati di tale attività sono in corso di pubblicazione. Sempre in questo ambito, è stata effettuata una collaborazione con l'Università degli studi di Roma La Sapienza (Prof.ssa G. Jona Lasinio) per la realizzazione di una tesi di laurea dal titolo "Confronto tra metodi di integrazione statistica basati sull'uso di informazione ausiliaria e sull'ipotesi di indipendenza condizionata".

Per quanto riguarda il *Software generalizzato per la produzione statistica* (struttura Mts/F), si è provveduto ad effettuare: 1) il rilascio del software Genesees v3.0 per il calcolo dei pesi, delle stime, degli errori campionari e per la loro presentazione sintetica; 2) il rilascio della versione beta del software Mauss per l'allocatione ad uno stadio di campionamento e l'implementazione di un modulo software, a livello prototipale, che implementa la metodologia per l'allocatione a due stadi e che potrebbe rappresentare una nuova funzione da inserire in Mauss.

La versione 3 di Genesees (acronimo per *Generalised software for Sampling Estimates and Errors in Surveys*) permette di supportare il processo di produzione statistica attuato nelle indagini campionarie per la realizzazione delle importanti fasi di attività che riguardano il calcolo dei pesi, delle stime e degli errori campionari e la loro rappresentazione sintetica. Con tale versione l'utente ha anche a disposizione uno strumento di analisi esplorativa che consente di scegliere il modello più adeguato per la pubblicazione degli errori campionari in modo sintetico.

Genesees v3.0 è infatti composta dalle tre funzioni:

La funzione di *Riponderazione*, applicabile in tutti i casi in cui esistano informazioni ausiliarie, espresse in termini di totali noti di variabili legate a quelle di interesse. Essa è finalizzata al calcolo dei pesi finali da attribuire alle unità campionarie, sulla base di totali noti delle variabili ausiliarie e dei valori assunti da queste nel campione estratto. Il contesto metodologico nel quale la funzione è stata concepita è quello degli stimatori di calibrazione (*calibration estimators*).

La funzione di *Stime ed errori campionari* finalizzata al calcolo delle stime e degli errori di campionamento e che produce, per ciascuna sottopopolazione di interesse: (i) le stime oggetto di indagine e i corrispondenti errori di campionamento assoluti, relativi, e gli intervalli di confidenza; (ii) le principali statistiche che forniscono informazioni sull'efficienza della strategia di campionamento utilizzata (effetto del disegno ed effetto dello stimatore); (iii) i modelli di regressione per la presentazione sintetica degli errori di campionamento (anche tale funzione fa riferimento alla teoria degli stimatori di calibrazione).

La funzione *Analisi dei modelli* permette all'utente di determinare la migliore rappresentazione sintetica degli errori campionari. Tale funzione, infatti, oltre a produrre i modelli per la presentazione sintetica degli errori di campionamento, come già era previsto nella versione 2.0

di Genesee, permette in aggiunta di analizzarne la validità in modo semplice ed interattivo, grazie al supporto di alcune funzionalità grafiche. L'utente viene in tal modo agevolato nell'individuazione di alcuni valori giudicati estremi rispetto al modello scelto e può procedere alla determinazione di un nuovo modello, che non tenga in considerazione tali valori estremi.

Mauss (acronimo per *Multivariate Allocation of Units in Sampling Surveys*) è un software generalizzato per determinare l'allocazione campionaria multivariata per disegni ad unico stadio di campionamento. Il software è dotato di una interfaccia *user-friendly* e di opportuni controlli logici e assicura una buona flessibilità. Allo stato attuale è divulgata la versione beta del software.

Mauss nasce da un progetto in Istat, svoltosi nel corso del 2003, per lo sviluppo di uno strumento software prototipale utile alla determinazione dell'allocazione campionaria nel caso multivariato e per più domini di stima, applicabile alle indagini con disegni ad uno e a due stadi di campionamento. Il progetto nasce dalla esigenza di potenziare un prototipo sviluppato in precedenza in Istat per determinare l'allocazione delle unità in campioni ad un unico stadio di campionamento ed è stato articolato in due fasi:

- la prima fase ha riguardato la realizzazione del software generalizzato Mauss per la determinazione dell'allocazione campionaria multivariata per disegni ad unico stadio di campionamento: si è partiti dal prototipo implementato in precedenza, che è stato rivisto, corretto ed esteso nelle funzionalità. Le potenzialità garantite dal prototipo di partenza sono state ampliate, inserendo alcune opzioni aggiuntive per agevolare la valutazione di possibili alternative;
- la seconda fase è stata programmata per rispondere alle esigenze degli utenti che lavorano con disegni campionari a due stadi, utilizzati principalmente nelle indagini Istat che si occupano delle famiglie. Attualmente è stato sviluppato un modulo, che rappresenta un software prototipale, per implementare la metodologia dell'allocazione a due stadi. Tale modulo potrebbe essere inserito nel software Mauss.

I progetti relativi alle *Metodologie e standard per la valutazione della qualità*, e, più in generale, per il miglioramento della qualità dell'informazione statistica (struttura Mts/G) hanno interessato diverse linee di attività.

Relativamente alla prima linea di attività, "*Sviluppo di strumenti per la standardizzazione, il monitoraggio e la documentazione dei processi produttivi*", i progetti realizzati hanno riguardato in modo particolare il consolidamento all'interno dell'Istat del Sistema informativo di documentazione delle Indagini, Sidi, che gestisce metadati e indicatori di qualità. In particolare, l'attività è stata concentrata sul popolamento degli indicatori standard per monitorare i processi di produzione. A tal fine è stata costituita, all'interno dell'Istituto, una rete di "referenti per la qualità e la documentazione" che hanno seguito un apposito corso di formazione e ricevuto un incarico nominativo. Per facilitare il calcolo degli indicatori standard gestiti in Sidi è stata avviata sia un'attività di produzione di strumenti generalizzati, sia un'attività di integrazione con altri database Oracle, gestiti presso i servizi di produzione, che contengono informazioni utilizzabili per le finalità suddette. Sulla base di esigenze espresse dagli utenti interni del Sistema, sono state estese le funzionalità di Sidi per gestire un insieme più ampio di indicatori standard di qualità.

Il sistema informativo Asimet, integrato con Sidi, sviluppato per la produzione delle note metodologiche dell'Annuario statistico italiano (Asi), a partire dalla documentazione validata gestita in Sidi, è stato esteso per consentire la documentazione delle elaborazioni. Ciò ha implicato anche un'estensione della documentazione gestita in Sidi (metadati relativi alle elaborazioni).

E' stato effettuato uno studio di fattibilità per lo sviluppo di un sistema informativo di diffusione di metadati di indagine e tavole statistiche, denominato Simet. Obiettivo del sistema è quello di rendere accessibile agli utenti esterni, attraverso Internet, informazioni sulle indagini gestite nei sistemi informativi Sidi e Asimet, collegandole alle tavole pubblicate dell'Asi.

Per quanto riguarda la seconda linea di attività, "*Sviluppo e applicazione di metodologie per stimare l'accuratezza dei dati*", è stata consolidata l'attività di ricerca relativamente alla stima dell'errore di risposta in presenza di dati di reintervista accoppiati con i dati originali mediante una procedura di Record linkage. In considerazione dell'obiettivo di analisi dell'errore di

risposta sui dati del Censimento della popolazione del 2001, utilizzando l'indagine di copertura come reintervista, si è iniziato a predisporre le procedure di stima per i domini di interesse. A tal fine sono stati utilizzati i dati sulle prime sezioni di censimento lavorate e quindi messe a disposizione da parte della struttura che ha in carico l'indagine di copertura.

Con riferimento allo studio delle metodologie statistiche per la valutazione della qualità dei dati, l'attività ha riguardato: i) l'approfondimento dei modelli a classi latenti nella stima dell'errore di risposta, con le diverse potenzialità che la metodologia offre per la stima delle diverse componenti dell'errore; ii) lo studio e l'applicazione di metodologie per l'integrazione di dati provenienti da fonti diverse con particolare riguardo agli aspetti di qualità inerenti il matching statistico, soffermandosi in modo specifico sugli approcci metodologici alla valutazione della precisione delle stime ottenute dopo l'uso di una procedura di matching.

Nell'ambito della terza linea di attività, *“Armonizzazione a livello europeo delle attività per il miglioramento della qualità”*, è proseguita l'attività svolta dal Gruppo di lavoro Eurostat *“Leg Implementation Group”*, costituito con il compito di seguire lo stato di implementazione delle raccomandazioni del Leg sulla qualità (*Leadership Group*) tra gli Istituti nazionali di statistica e di favorire la costituzione di progetti comuni finalizzati a sviluppare attività legate a specifiche raccomandazioni. La ricognizione presso gli Istituti di statistica è stata effettuata somministrando la seconda edizione di uno specifico questionario ed i risultati sono stati presentati al Cps (Comitato per il programma statistico). Si sono conclusi i seguenti progetti, svolti in collaborazione tra più istituti, che hanno riguardato: “individuazione e analisi delle variabili chiave di processo e redazione di un manuale”; “predisposizione di una check-list per l'autovalutazione”; “stato dell'arte sulle indagini per valutare la soddisfazione degli utenti”. Sono in fase di avanzato sviluppo i progetti relativi a “metodi per ridurre il peso statistico sui rispondenti” e “stato dell'arte sui metodi di auditing”.

E' stata inoltre costituita, all'interno delle attività del Gruppo di lavoro Eurostat *“Valutazione della qualità delle statistiche”*, una task force con il compito di definire un set di indicatori standard a livello europeo per ciascuna componente della qualità, secondo la scomposizione ufficiale definita da Eurostat. L'attività della task force ha finora riguardato la definizione di un set minimo di indicatori orientati al produttore. Nel corso del prossimo anno è prevista la definizione di indicatori rivolti all'utente dell'informazione statistica.

E' stata svolta attività di formazione in ambito Sistan e di cooperazione internazionale su tematiche inerenti la qualità.

Il servizio Psm è nato con le seguenti finalità:

- *Progettazione* dei principali aspetti metodologici dell'impianto di rilevazione delle più importanti indagini dell'Istituto;
- *Supporto* relativo alla revisione di particolari aspetti metodologici delle indagini dell'Istat (e di alcuni rilevanti Enti del Sistan) sulle famiglie e sulle imprese, istituzioni e aziende agricole;
- *Sviluppo della ricerca* su temi metodologici strategici per il miglioramento della qualità dell'informazione statistica prodotta e i cui risultati devono essere strettamente connessi alla progettazione delle strategie di rilevazione;
- *Formazione e comunicazione*;
- *Le risorse umane sono prevalentemente dedicate (per circa un 70%) alle attività di progettazione e supporto.*

Coerentemente con le finalità sopra elencate, il servizio nei primi mesi della sua attività ha avviato i seguenti progetti che hanno coinvolto trasversalmente le risorse di tutte le unità operative del servizio:

- *Progettazione e reingegnerizzazione della rilevazione sui prezzi.* Tale attività è sviluppata nel quadro di una *Direttiva strategica* assegnata dal Consiglio dell'Istat alle Direzioni Dcmt e Dcpc. Nell'ambito di questa attività sono stati elaborati alcuni documenti relativi alla documentazione dell'attuale processo di indagine e alle proposte relative alla riprogettazione dell'attuale impianto;
- *Studio di metodi per la produzione di stime anticipate e la revisione di indicatori congiunturali.* Tale attività è sviluppata nel quadro di una *Direttiva strategica* assegnata dal Consiglio dell'Istat alle Direzioni Dcmt e Dcsc. Nell'ambito di questa attività è stato

rilasciato un primo rapporto metodologico ed è stato elaborato un metodo che permette il rilascio di stime anticipate per l'indagine mensile sull'attività edilizia;

- *Calcolo di metodi indiretti per ottenere stime per domini sub-regionali* dell'indagine Multiscopo sulle famiglie sul ricorso ai servizi sanitari; inoltre è stato approfondito lo studio del disegno campionario per la nuova indagine per la nuova dell'anno 2004 in cui è previsto un finanziamento da parte del Ministero della salute di circa 36.000 famiglie campione per produrre stime attendibili a livello sub-regionali.

Per quanto riguarda le attività relative alla *Strategia campionaria e tecnica di rilevazione* (struttura Psm/A), il lavoro svolto dall' U. O. può essere sintetizzato nei seguenti punti:

- la definizione della strategia di pubblicazione delle stime di povertà regionali basate sui dati dell'indagine sui consumi delle famiglie; ciò ha comportato (i) lo studio di stimatori alternativi di tipo *diretto* e di tipo *indiretto* detti anche stimatori per piccole aree che sfruttano informazioni ausiliarie legate al fenomeno di interesse; (ii) lo studio della metodologia per la stima degli errori di campionamento;
- lo svolgimento della fase finale del progetto europeo Eurarea - con Statfin, Statsweden, StatNorway, Ons (leader del progetto) Ine University of Southampton - che ha condotto alla definizione finale (a) delle nuove metodologie proposte nell'ambito degli stimatori per piccole aree; (b) simulazioni dei nuovi metodi e dei metodi standard sui dati censuari del 1991 per la stima occupati e persone in cerca a livello provinciale e dei Sistemi locali del lavoro e l'analisi dei risultati sperimentali prodotti; (c) messa a punto di software generalizzati che implementano le nuove metodologie proposte;
- lo studio del disegno campionario dell'indagine Eusilc e la sua selezione;
- lo studio progettuale finalizzato a definire una riduzione della numerosità campionaria dell'attuale indagine sulle forze di lavoro. Lo studio è sostanzialmente terminato e si sta redigendo un rapporto con i principali risultati ottenuti;
- lo studio progettuale finalizzato a definire una strategia campionaria per l'indagine sull'uso delle acque.

Per quanto riguarda le attività relative ai *Processi di elaborazione e modelli di stima* (struttura Psm/B), è in fase conclusiva la redazione del volume sulle indagini di qualità (relative alla coperture e all'errore di misura) del Censimento dell'agricoltura.

E' in fase di conclusione uno studio empirico finalizzato a un uso intensivo dei dati amministrativi di fonte fiscale e di bilancio. I risultati di tale studio, permetteranno di diminuire le numerosità campionarie richieste dalle indagine sui conti economici delle imprese. Si è fornito supporto metodologico finalizzato alla progettazione delle strategie di campionamento delle indagini Istat su famiglie ed individui. Le indagini su cui si è lavorato sono in particolare: (i) l'indagine annuale multiscopo; (ii) l'indagine sui consumi delle famiglie; (iii) l'indagine sulla sicurezza dei cittadini; (iv) l'indagine viaggi e vacanze; (v) l'indagine sull'uso del tempo; (vi) l'indagine sull'inserimento professionale dei laureati; (vii) l'indagine sui percorsi di studio e di lavoro dei diplomati; (viii) l'indagine campionaria sulle nascite.

Relativamente al *Supporto metodologico per il rilascio di informazioni statistiche* (struttura Psm/C), le linee di attività possono essere schematizzate nei punti seguenti: (a) gestione del Laboratorio per l'Analisi di dati elementari, (b) integrazione di nuove funzionalità nel software Argus, (c) sviluppo di metodologie per la protezione di dati elementari e (d) formazione e diffusione delle nuove tecniche e procedure in merito alla tutela della riservatezza.

(a) Nel corso del 2003 sono proseguite le attività di gestione e promozione del Laboratorio di analisi di dati elementari Adele. E' stato creato un *data base* che raccoglie informazioni sull'utenza, la tipologia di analisi effettuate e i tipi di dati maggiormente richiesti. E' stata mantenuta la collaborazione con il gruppo che si occupa dell'archivio dei microdati (Armida) per far conoscere capillarmente l'esperienza del Laboratorio Adele all'interno dell'Istituto sfruttando gli incontri programmati per questa attività.

(b) Per quanto riguarda il *software* per la valutazione del rischio di violazione della riservatezza, e per la sua protezione, sono state integrate nuove funzionalità nel *software* Argus. Il *software* Argus viene sviluppato nell'ambito del progetto europeo Casc (*Computational Aspects of Statistical Confidentiality*) per la produzione di *file* di dati elementari (modulo  $\mu$ -Argus) e tabelle (modulo  $\tau$ -Argus) che rispettino il vincolo del segreto statistico. Tale software rappresenta un