

Metodi e aspetti tecnologici di raccolta dei dati sulle imprese¹.

1. INTRODUZIONE

Nell'ambito della statistica ufficiale la funzione di acquisizione dati riveste particolare importanza in quanto con essa vengono alimentati i processi di produzione statistica. Con terminologia riconosciuta in ambito internazionale, ricordiamo che l'attività complessiva di acquisizione si articola in *data collection primario*, dovuto alle indagini dirette, che generano dati elementari grazie alla compilazione o all'autocompilazione di questionari di rilevazione, e *data collection secondario*, significativo per quei processi che richiedono l'uso o il riuso di informazione raccolta e/o prodotta da altre fonti.

L'acquisizione dei dati, così intesa, costituisce un'attività particolarmente delicata, in ragione di diversi fattori:

- relativamente ai processi di produzione la sua posizione a monte delle altre fasi operative la rende critica per l'avanzamento dei processi stessi;
- costituisce una delle maggiori componenti di costo dell'attività statistica;
- prevede la comunicazione con l'esterno pertanto è un momento di *esposizione* dell'Istituto;
- richiede che vengano considerate misure per la riservatezza e la sicurezza dei dati e dell'ambiente informatico.

Questo documento è dedicato al data collection primario realizzato per una specifica classe di rispondenti: le imprese che rispondono a indagini dirette in modalità self-interviewing (questionario autosomministrato) sulle caratteristiche economiche strutturali.

Prima di entrare nel cuore dell'argomento introdurremo, nel paragrafo 2, il tema dell'innovazione della fase di raccolta e descriveremo, nel paragrafo 3, la piattaforma web operativa per le soluzioni di cattura; il paragrafo 4 sarà dedicato alle specifiche soluzioni adottate per la cattura dei dati delle rilevazioni economiche strutturali; il paragrafo 5 presenterà le motivazioni della scelta del formato elettronico per il modello di rilevazione; il paragrafo 6 descriverà l'integrazione tra gli aspetti innovativi e quelli tradizionali e il trattamento successivo alla raccolta; infine, nel paragrafo 7 sarà dato cenno agli sviluppi futuri della piattaforma e dei processi descritti.

2. L'INNOVAZIONE DELLA FASE DI RACCOLTA: USO DEL WEB

Se si è interessati ad ottimizzare la tempestività, migliorare la qualità e contenere i costi nella raccolta dei dati, la risposta tecnologica per rendere più efficiente l'acquisizione è sicuramente quella dei canali telematici, che ci consentono di sostituire gradualmente il data collection

¹ I paragrafi 1, 2, 3, e 7.1 sono a cura di R. Balestrino, i paragrafi da 4 a 4.4 di F. Garritano, il paragrafo 5 di C. Cipriano, i paragrafi da 6 a 6.3 e 7.2 di L. Fanfoni

tradizionale con il data capturing. D'altra parte, è riconosciuto che il fattore tempestività sia anche componente della qualità stessa dell'informazione statistica e i Regolamenti europei impongono costantemente la riduzione del ritardo nella diffusione del prodotto statistico.

E' naturale quindi che tra le tecniche CASIC (Computer Assisted Survey Information Collection), che innovano la fase di raccolta, sia in crescente diffusione l'uso dei servizi Internet e in particolare il ricorso al WEB.

Introdurre l'innovazione richiede però, sia presso l'Amministrazione che gestisce la rilevazione sia presso gli interlocutori esterni, non solo un'adeguata dotazione tecnologica ma anche un'organizzazione del lavoro rivisitata per sostenere le nuove modalità operative. Inoltre, se i canali telematici entrano a tutto campo nei processi di produzione, è necessario che siano garantiti adeguati livelli di servizio, sia a livello di rete pubblica (politiche degli Internet Provider, capacità di banda, potenziamento delle infrastrutture) sia a livello di rete aziendale (monitoraggio accurato della intranet e di tutti i nodi interessati al completamento del processo di cattura dati, dal nodo di front-end per i rispondenti al server di back-end presso il settore di produzione dell'Istat che gestisce la rilevazione).

Parallelamente, consentire di trasmettere dati in rete e depositare "pacchetti" dall'esterno nel nostro ambiente operativo, espone l'Istituto a rischi di intrusione e corruzione dell'informazione e implica perciò l'adozione di adeguate misure di sicurezza per proteggere i dati stessi e l'ambiente di produzione. Si capisce pertanto con quanta attenzione vada affrontato l'impianto dei nuovi metodi e tecniche negli ambienti di produzione.

In generale non è possibile definire una soluzione tecnologica standard valida per qualunque operazione di raccolta in ambiente statistico; la struttura e la lunghezza del questionario, la complessità dell'informazione da rilevare certamente influenzano le scelte da operare circa un efficace utilizzo della rete. D'altra parte, oltre al questionario, che gioca un ruolo determinante nell'operazione di raccolta, l'altro fattore cruciale da considerare è il rispondente, al quale si richiede dotazione tecnologica adeguata, alfabetizzazione informatica, tempo e competenza sulla materia per la quale deve rispondere.

L'esperienza fin qui maturata ci ha portato a sviluppare due modelli di soluzione, entrambi basati sull'utilizzo del WEB e di un questionario elettronico, ma che differiscono per la modalità di compilazione del modello: la soluzioni online e offline.

Offrire al rispondente una compilazione online presuppone la disponibilità, presso l'amministrazione che conduce l'indagine, di un web server su cui vengono allestite pagine web contenenti il questionario in formato ipertestuale (HTML) o altro formato elettronico (per esempio PDF); il server accetta query da client esterni (browser di altri utenti web, nostri rispondenti) che possono così compilare il form. Con l'utilizzo di linguaggi ad hoc (p.es. JAVA, PHP, PERL...) è possibile realizzare l'editing dei dati contestualmente all'immissione e gestire i percorsi di somministrazione del questionario. In questo caso la compilazione si svolge online e i dati immessi dal rispondente sono contestualmente resi disponibili nell'ambiente di produzione che viene alimentato quindi con tempestività massima. L'architettura dell'applicazione è di tipo client-server, dove il carico elaborativo complessivo è suddiviso tra lato-client e lato-server: il traffico che si genera, in ragione del carattere dinamico che deve assicurare il colloquio e lo scambio di dati tra i due lati, influisce sulle prestazioni e condiziona il tempo di connessione totale, pertanto in generale è consigliabile ridurre l'attività di controllo sui dati immessi e puntare invece sulla tempestività nella disponibilità dei dati stessi.

Questa modalità è anche nota, tra le tecniche CASIC, come WEB survey. Va ricordato che le web survey al loro nascere hanno generato preoccupazioni di carattere metodologico, indotte dalla limitata diffusione dell'accesso alla rete Internet nelle popolazioni classiche oggetto di studio, condizione che può dare luogo a un fenomeno di autoselezione dei rispondenti, con conseguente

impossibilità di estrarre campioni casuali, e difficoltà di trattamento classico degli aspetti inferenziali nel riportare all'universo i risultati ottenuti osservando le unità campionarie.

Tuttavia questo vincolo, relativo alle indagini campionarie, non sussiste se l'intervento della rete Internet ha luogo a valle dell'estrazione campionaria: la rete non rappresenta lo strumento di estrazione delle unità campionarie ma unicamente il mezzo di raccolta. In questo caso, che è l'unico praticato nell'ambito della statistica ufficiale, più propriamente bisognerebbe parlare di WEB collection e non di WEB survey: la raccolta via WEB viene inclusa in un processo di raccolta mix-mode, o multi-tecniche. La raccolta mix-mode richiede che venga comunicato a tutte le unità campionarie la possibilità di utilizzare, in alternativa alla tecnica tradizionale, un questionario elettronico attraverso canali telematici.

In alternativa alla WEB survey, può essere talvolta più conveniente, per esempio per ridurre i tempi di connessione del rispondente o per la complessità del questionario stesso, che la compilazione avvenga offline. I servizi WEB sono essenziali anche in questo caso: il web server infatti deve consentire che il rispondente, accedendo al sito predisposto per questa funzione, possa scaricare sulla propria postazione client il questionario elettronico in formato eseguibile, e, in una successiva sessione, possa restituire il questionario, che avrà compilato con comodo nel proprio ambiente di elaborazione.

Questa modalità minimizza il tempo di connessione alla rete internet. Anche la compilazione offline, per ragioni assolutamente analoghe a quelle già esposte per la WEB survey, va somministrata in modalità mix-mode, associata cioè al canale tradizionale.

Entrambe le soluzioni basate sul WEB, oltre alla tempestività, garantiscono la drastica riduzione dei costi: i ruoli umani dell'intervistatore e dell'operatore di data-entry non sono necessari né si devono sostenere costi di riproduzione fisica del modello di rilevazione. Inoltre non generano esclusione di rispondenti, anzi li agevolano dando loro la possibilità di selezionare la modalità preferita; va evidenziato d'altra parte che, a fronte del generale abbattimento di costi indotto dall'uso della rete Internet, di cui si è già detto, un nuovo costo gestionale si genera, dovuto alla necessità di integrare e monitorare i diversi canali di acquisizione, innovativi e tradizionali, in un unico processo di produzione. Una valutazione specifica andrebbe fatta anche circa gli eventuali costi da sostenere per adeguare le competenze dei settori di produzione all'uso dei nuovi strumenti.

3. INDATA: UNA PIATTAFORMA WEB DEDICATA AL DATA CAPTURING

Per realizzare l'obiettivo della razionalizzazione e standardizzazione dell'ambiente dedicato alla cattura dati, alcuni strumenti applicativi sono già stati sviluppati e resi operativi a cura della struttura MTS/C, dedicata al data capturing in collaborazione, per quanto concerne il rispetto delle condizioni di sicurezza ambientale, con la struttura GSS/D, dedicata alla sicurezza informatica, e per le specifiche soluzioni di cattura, con le strutture locali delle Direzioni di produzione interessate a innovare le proprie indagini, principalmente DCSS e DCSC.

Lo strumento che descriviamo di seguito è il sito web ***indata.istat.it***, che concentra le attività di cattura dati via WEB relative a diverse rilevazioni statistiche.

Indata.istat.it è un sito attivo dalla fine degli '90, sviluppato con esclusive risorse interne dell'Istituto, attualmente configurato in ambiente AIX 4.3, con webserver APACHE 1.3.20 e prevalenza di linguaggio PHP 4.0.6 per lo sviluppo delle applicazioni.

Il parco software per l'allestimento del web server (APACHE, dbms MYSQL e PHP) appartiene all'ambito del software open source, cioè software "a codice sorgente aperto", che, attraverso la disponibilità totale del codice sorgente, ne consente la libera circolazione e autorizza di conseguenza l'attivazione di eventuali processi di modifica, evoluzione e riuso. In questo senso le

scelte operate sono anche in sintonia con la direttiva del 19 dicembre 2003 del ministro dell'innovazione e delle tecnologie che invitava le Pubbliche Amministrazioni ad aderire all'offerta open source.

In particolare l'adozione del linguaggio PHP, in sostituzione del precedente PERL utilizzato per i primi sviluppi prototipali del sito, si è rivelata una scelta efficace per lo sviluppo di applicazioni WEB dinamiche, sia dal punto di vista delle prestazioni sia per quanto concerne la creazione di competenze di analisi e programmazione a partire da skill di sviluppo tradizionale. Inoltre, come già detto, l'appartenenza di PHP al mondo open source ne garantisce un aggiornamento continuo, grazie al quale sono state rese disponibili in rete numerose librerie (PEAR) di moduli già sviluppati e il linguaggio, nell'ultima versione, supporta per la scrittura del codice, oltre alla modalità procedurale, anche quella orientata agli oggetti (sviluppo object-oriented).

Tutte le opzioni offerte sul sito *indata* sono ispirate ai seguenti requisiti/obiettivi:

- ❑ Presentare l'Istituto all'esterno con immagine e identità omogenea e stabile
- ❑ Garantire l'identità reciproca di chi trasmette e chi riceve i dati
- ❑ Garantire la riservatezza dei dati nel processo di acquisizione e la sicurezza complessiva dell'ambiente
- ❑ Minimizzare l'impatto sull'ambiente operativo dell'utente esterno (non è richiesta installazione di software sui client)
- ❑ Dare riscontro all'utente esterno dell'operazione da lui effettuata (e-mail di conferma)
- ❑ Favorire il monitoraggio delle attività di acquisizione
- ❑ Favorire la gestione interna degli ambienti operativi dedicati all'acquisizione e contenerne i costi

Il sito è esposto sulla rete Internet al di fuori del firewall di Istituto ed è accessibile ad utenti esterni autorizzati per:

- compilare online modelli elettronici;
- scaricare (download) modelli elettronici;
- accedere e stampare documentazione e istruzioni di indagine;
- caricare (upload) modelli elettronici compilati offline;
- trasmettere qualunque dataset in modalità sicura.

Come è naturale, date le attività istituzionali, la classe più numerosa degli utenti del sito è quella dei rispondenti a indagini dirette, ma *indata* viene utilizzato anche, per esempio, dalle amministrazioni regionali, che costituiscono organi intermedi di rilevazione per alcune indagini del settore Agricoltura, e alcune società di mercato, che trasferiscono microdati rilevati per conto dell'Istituto.

In generale, nell'articolazione delle funzioni e nella presentazione dei contenuti, il sito *indata* ha progredito gradualmente verso un modello standard che cerca di dare risposta alle aspettative sia del rispondente sia di chi gestisce l'indagine. Per qualunque indagine trattata e qualunque sia la modalità di cattura adottata, tutte le operazioni si svolgono in sicurezza. L'utente, al primo accesso, riceve un messaggio con cui un ente autorizzato alla certificazione accerta l'identità del sito Istat. Il rispondente stesso deve effettuare una tantum un'operazione di registrazione e fornire in seguito le proprie credenziali per accedere alle sessioni protette, cioè tutte le pagine web dinamiche attraverso cui client e server scambiano informazioni che si vogliono tenere riservate (per esempio user-id e password o i microdati di indagine).

Il modello adottato prevede anche una funzione generale di orientamento/help del rispondente che si esplica nella possibilità di:

- essere informato sulle finalità della rilevazione a cui partecipa,

- essere istruito sulle modalità di compilazione del modello e sulle caratteristiche operative del form elettronico,
- inviare senza uscire dalla sessione un messaggio di comunicazione a un indirizzo Istat.

A *indata.istat.it*, attraverso la intranet d'Istituto, accedono anche utenti interni responsabili delle rilevazioni, preventivamente autorizzati ad effettuare operazioni di monitoraggio dell'andamento della raccolta.

Riportiamo la homepage del sito e la successiva pagina menu delle indagini per avere un'idea delle realtà produttive ospitate.





4. INDATA: LE INDAGINI CHE OSSERVANO LE IMPRESE

Tra le classi di rispondenti tradizionali (famiglie, istituzioni, imprese) le imprese sono state le prime a cui è stata proposta la modalità telematica di risposta, a partire dalle esperienze di carattere sperimentale iniziate alla fine degli anni '90.

A questo punto si può ritenere superata la fase sperimentale: l'approccio multitecniche alla raccolta dei dati, già descritto al punto 2, è entrato nei processi di produzione delle indagini rivolte alle imprese che tradizionalmente svolgono autointerviste su modelli trasmessi via posta o fax e che, in misura sempre maggiore e tendente alla esaustività, sono dotate di attrezzature informatiche e accesso in rete.

Le prime applicazioni WEB per la compilazione online di un questionario elettronico furono sviluppate ad hoc su indata.istat.it con linguaggio PERL. Di queste è rimasta attiva solo la linea relativa alla Rilevazione trimestrale del fatturato e dell'occupazione delle imprese.

Successivamente, a cura del Laboratorio Data Capturing fu selezionato nel 2000 un sistema software di mercato, TELEFORM, che prevede l'acquisizione mista (approccio multitecniche) di moduli digitali, in formato HTML o PDF attraverso WEB, e moduli cartacei, per mezzo di scanner o fax-server con successivo riconoscimento automatico dei caratteri. A tutt'oggi con TELEFORM la compilazione del questionario elettronico avviene in modalità online.

Nella tabella che segue sono riportate le rilevazioni sulle imprese di industria e servizi trattate fino a questo momento e le rispettive modalità di cattura che si sono applicate.

	RILEVAZIONE	MODALITA' APPLICATIVA DI CATTURA
1	Rilevazione trimestrale del fatturato e ordinativi	Sviluppo ad hoc – linguaggio PERL – questionario HTML-compilazione online
2	Rilevazione trimestrale su posti vacanti e ore lavorate	TELEFORM – questionario PDF-compilazione online
3	Rilevazione periodica sull'attività alberghiera	TELEFORM – questionario PDF-compilazione online
4	Rilevazione mensile su occupazione, orari di lavoro e retribuzioni (attualmente su sito teleform.istat.it)	TELEFORM – questionario PDF-compilazione online
5	Rilevazione mensile sulle vendite al dettaglio (attualmente su sito teleform.istat.it)	TELEFORM – questionario PDF-compilazione online
6	Rilevazione annuale del sistema dei conti delle imprese	Sviluppo ad hoc- linguaggio PHP- questionario EXCEL- compilazione offline
7	Rilevazione annuale Stima provvisoria del valore aggiunto	Sviluppo ad hoc- linguaggio PHP- questionario EXCEL- compilazione offline
8	Rilevazione annuale della produzione industriale	Sviluppo ad hoc- linguaggio PHP- questionario EXCEL- compilazione offline
9	Rilevazione annuale sulla struttura del costo del lavoro	Sviluppo ad hoc- linguaggio PHP- questionario EXCEL- compilazione offline
10	Rilevazione annuale sulle telecomunicazioni	Sviluppo ad hoc- linguaggio PHP- questionario EXCEL- compilazione offline

Relativamente al gruppo delle statistiche economiche strutturali sulle imprese, lo sviluppo delle soluzioni di cattura ha seguito un modello cooperativo tra la struttura centrale del data capturing e le strutture locali di produzione (direzione DCSS, servizio SSI) che si è rivelato molto fruttuoso. Per queste cinque rilevazioni è stata adottata la modalità di compilazione offline di un modello in formato di foglio elettronico (file .xls), giudicata più agevole per il rispondente. Le motivazioni relative a questa scelta e la descrizione del formato e degli strumenti utilizzati saranno descritti nei paragrafi successivi.

In questo paragrafo saranno invece presentate le funzionalità predisposte sul sito INDATA e relative all'interfaccia “front-end” con le imprese rispondenti. Il flusso procedurale prevede alcuni passi standard che descriviamo di seguito:

- Le imprese selezionate nel campione di una indagine vengono informate, per canale tradizionale (postale), dell'innovazione che l'Istituto propone relativamente alla raccolta dei dati. Questo avviene generalmente nella lettera d'invito a partecipare alla rilevazione dove si comunica anche l'indirizzo URL del sito web a cui connettersi e la coppia (user-id, password) da utilizzare per accedere allo spazio dei rispondenti.
- I rispondenti che accedono al sito, dopo avere selezionato la rilevazione di proprio interesse sono guidati da un menu funzionale standard che consente di attivare le seguenti funzionalità:
 - visualizzazione delle informazioni sull'indagine, delle avvertenze e della guida alla compilazione del questionario (anche in formato PDF), dei contatti relativi al personale dell'indagine per il supporto e l'assistenza al rispondente nella fase di raccolta dei dati;
 - registrazione “una tantum” in qualità di rispondente all'indagine; per questa funzionalità si utilizza la coppia (user-id, password provvisoria) comunicata dall'Istituto;
 - modifica della password, in ottemperanza alle normative vigenti sulla riservatezza;
 - prelievo (download) del questionario elettronico;

- invio (upload) del questionario elettronico compilato.
- Il rispondente preleva il modello salvandolo sul proprio computer. A questo punto può compilare, salvare e modificare il questionario in qualsiasi momento, come un documento qualsiasi.
- Una volta completata la compilazione si collega di nuovo sulla pagina dell'indagine e, usando la user-id e password personale, invia il modello elettronico. L'invio consiste in un upload in modalità protetta del file Excel verso il server INDATA.

Queste in sintesi le fasi di risposta delle indagini elettroniche implementate per SSI sul sito INDATA.

Tutte le applicazioni sono state realizzate in linguaggio PHP, con dati relativi agli utenti (le imprese rispondenti) residenti su db MySQL. Di seguito vengono descritte in dettaglio le funzionalità principali di registrazione, prelievo e invio dei modelli, e il sistema di gestione degli accessi per le varie indagini:

4.1 Registrazione del rispondente

La prima operazione che il rispondente deve effettuare è la registrazione. Sulla lettera che ha ricevuto sono evidenziati il codice impresa e la password attuale (o provvisoria): con queste informazioni può procedere con la registrazione, indicando inoltre la propria password personale. Con questa password personale potrà accedere alle pagine per la ricezione e invio del modello telematico. Inoltre potrà cambiare la password in qualsiasi momento, tramite l'apposita funzione, presente sul sito INDATA. Di seguito viene presentata la pagina di registrazione dell'indagine TLC.

Le password provvisorie (o attuali) vengono generate automaticamente in maniera casuale da una procedura, sviluppata dal gruppo MTS-C, che ha come input la lista di spedizione predisposta dalle strutture SSI. Il personale SSI utilizza la lista di spedizione, con l'aggiunta del campo di password iniziali, per la definizione della lettera di avviso personalizzata della rilevazione

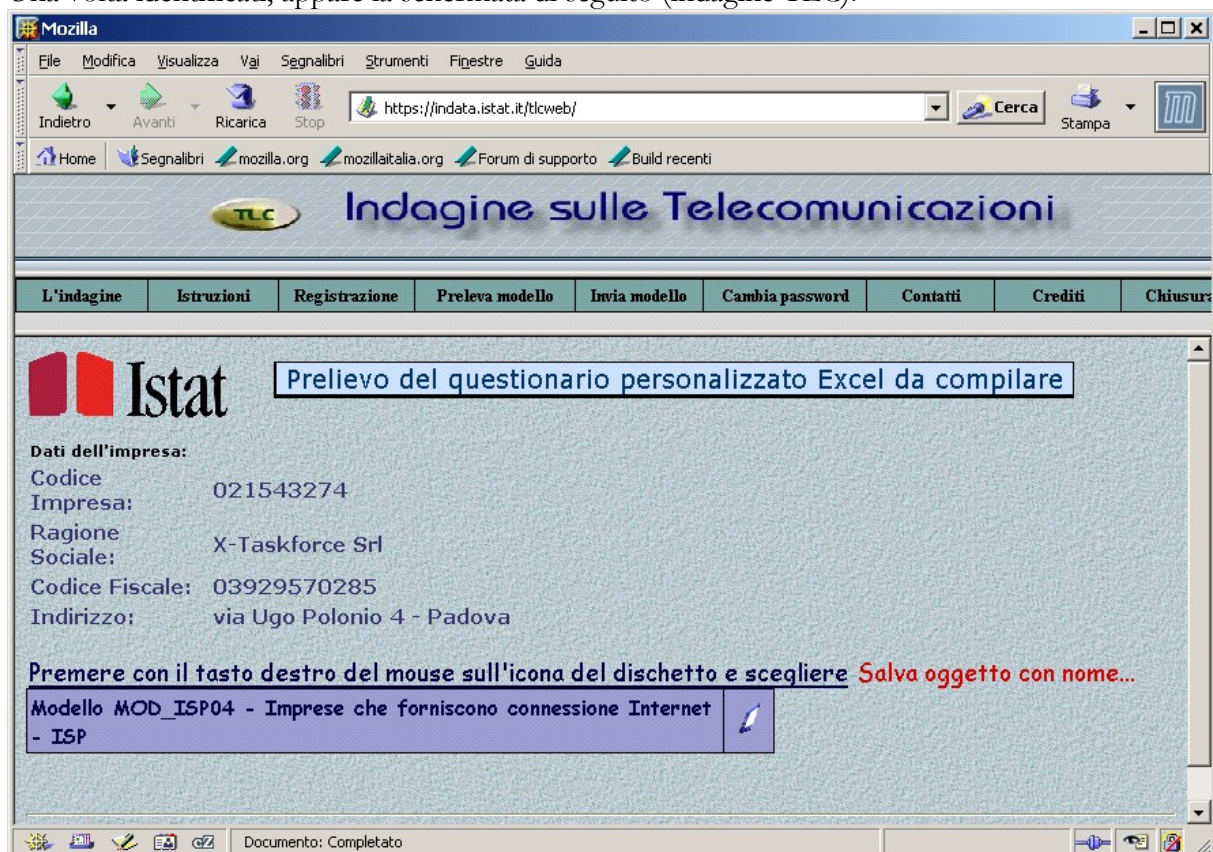
elettronica. In questo modo ogni impresa riceve una lettera con la propria user-name e password iniziale.

Per ragioni di sicurezza, il sistema è stato concepito in modo tale che il personale ISTAT non può conoscere le password personali inserite o modificate dal rispondente. Nell'eventualità che il rispondente dimentichi la password, il personale dell'indagine può soltanto ripristinare la password iniziale (vedi la gestione delle operazioni di raccolta). Quando accade ciò il rispondente deve registrarsi di nuovo e definire una nuova password iniziale.

4.2 Prelievo dei modelli elettronici

Per accedere a questa funzione, occorre essere un rispondente già registrato. Nell'ambito della stessa sessione di registrazione, non viene richiesta la password personale (in quanto già inserita). Se invece la sessione è diversa, per prima cosa viene richiesta l'identificazione dell'impresa, con password personale.

Una volta identificati, appare la schermata di seguito (indagine TLC):



Nella pagina vengono presentate le informazioni principali dell'impresa, e indicata la modalità di prelievo del modello personalizzato. Il rispondente, seguendo le istruzioni, salva il modello personalizzato su una cartella a scelta sul proprio computer, esattamente come per un qualsiasi documento di lavoro.

La funzione di personalizzazione, realizzata interamente in PHP, con accesso ai dati anagrafici dell'impresa residenti su database MySQL, esegue i seguenti passi:

- 1) acquisizione delle informazioni anagrafiche di personalizzazione dell'impresa individuata;
- 2) apertura, in formato binario, di un modello "template" del questionario di rilevazione in formato excel. Questo è il file excel di base, unico per tutti i rispondenti, che contiene,

oltre alle sezioni del questionario, le informazioni sulle coordinate dei campi di personalizzazione;

- 3) scrittura dei campi di personalizzazione sul modello di base e salvataggio con nome parametrizzato all'interno di una area temporanea propedeutica al download.

Per alcune indagini è stato necessario articolare questa funzione di creazione “on-line” del modello personalizzato; l'implementazione più complessa ha riguardato l'indagine PRODCOM. Per essa, infatti, occorre consentire l'accesso a livello di impresa, come per le altre, ma ogni impresa doveva poter prelevare tanti modelli per quante sono le Unità Locali di Produzione ad essa riferita (in numero variabile tra 1 e 40). In questo caso la funzione acquisisce il numero di modelli da personalizzare per la singola impresa e in ciclo ripete le procedure di personalizzazione presentando nella pagina di prelievo dei modelli tante righe con indicazioni di salvataggio per quanti sono i modelli relativi alle diverse unità locali di produzione.

Inoltre, dal momento che ogni singola unità locale di produzione fa riferimento a una determinata tipologia produttiva, era necessario definire all'interno del modello il repertorio merceologico (l'elenco dei prodotti) pertinente a tale tipologia. A tale scopo sono stati realizzati 16 modelli elettronici di base per la personalizzazione, tutti uguali tra loro tranne che per il repertorio merceologico contenuto. Per ogni singola unità locale di produzione, viene dapprima acquisita la tipologia, quindi la procedura apre il modello di base riferito a una delle 16 tipologie.

4.3 Invio dei modelli elettronici

Il rispondente compila il questionario elettronico sul proprio computer, salvando di volta in volta le modifiche effettuate. Una volta completata la compilazione, il rispondente si collega alla pagina dell'indagine sul sito INDATA, e aziona la funzione Invia Modello, che presenta la seguente pagina di richiesta identificazione:



Il rispondente inserisce user-name, password personale e indirizzo e-mail, nel caso desideri avere una ricevuta di ritorno dell'avvenuta ricezione del modello da parte dell'ISTAT.

Una volta identificata l'impresa rispondente, viene presentata la schermata di invio, con le informazioni anagrafiche principali, il campo del percorso completo del file e i bottoni: SFOGLIA e INVIA.

Il rispondente può scrivere direttamente nel campo il percorso e nome del file relativo al modello elettronico compilato e residente sul proprio computer. In alternativa può facilmente selezionare il file con la finestra di navigazione delle risorse del proprio computer che appare quando si preme il pulsante SFOGLIA. Una volta indicato il file, si procede con la spedizione (pulsante INVIA). Di seguito al schermata di spedizione:



L'operazione di invio consiste in un "upload" sicuro (protocollo https) dal computer locale del rispondente, al server INDATA.

Il sistema di ricezione effettua una serie di controlli sulla tipologia di file ricevuto e, per più recenti rilevazioni elettroniche, anche sulla presenza dell'identificativo di questionario. Questo controllo è necessario per verificare che il modello inviato sia relativo alla replicazione di indagine di riferimento. Per esempio, dalla pagina presentata in precedenza, è possibile inviare soltanto modelli dell'indagine TLC, anno 2004.

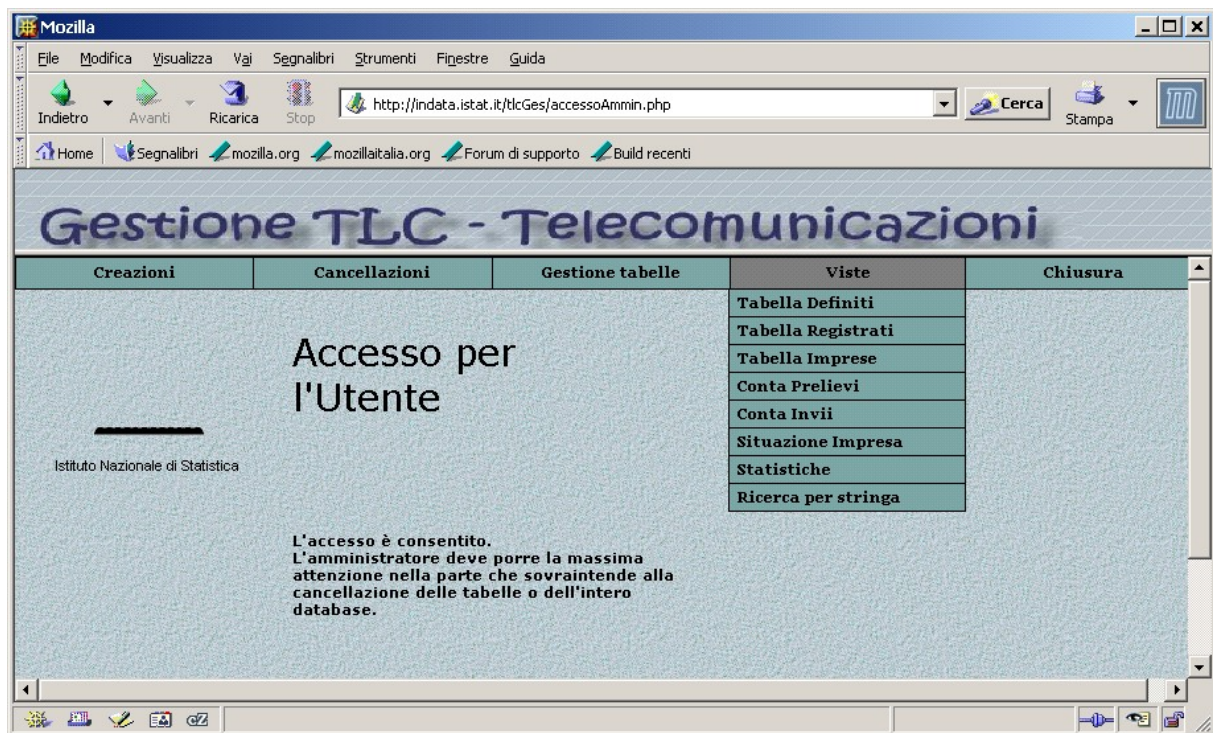
L'esito della spedizione viene segnalato da pagina con messaggio. Se richiesto, viene spedita automaticamente una mail all'indirizzo indicato dal rispondente.

Anche per questa funzione sono state necessarie alcune articolazioni "ad-hoc". Per esempio, per l'indagine PRODOCOM, si è stabilito di agevolare l'invio dei modelli elettronici per le imprese pluri-localizzate, quelle cioè con un numero di modelli maggiore di uno. Per queste è stata definita una pagina di spedizione con la possibilità di selezionare più file, e di inviare in un'unica soluzione i modelli individuati (upload multiplo).

4.4 Gestione e monitoraggio delle operazioni di raccolta

La gestione e il controllo degli accessi dei rispondenti è a cura del personale di supporto alla produzione dell'indagine, per il quale sono state realizzate applicazioni (in linguaggio PHP) che accedono ai database degli utenti di INDATA.

Di seguito viene presentata la schermata principale dell'applicazione di gestione e monitoraggio della raccolta dati su INDATA, realizzata per l'indagine TLC:



Le principali operazioni che vengono effettuate dal personale SSI presso le indagini sono le seguenti:

- Visualizzazione della lista delle imprese rispondenti, con le informazioni anagrafiche e le password provvisorie, con possibilità di esportazione in formato elaborabile;
- Visualizzazione delle informazioni sugli accessi delle imprese rispondenti al sito INDATA, con dettaglio della data e dell'operazione effettuata (registrazione, prelievo, invio), e con possibilità di analisi specifiche sulle operazioni di raccolta dati;
- Ripristino della password iniziale, per le imprese che lo richiedono. Una volta effettuata questa operazione, il rispondente deve effettuare ex-novo la registrazione;
- Modifica delle informazioni relative alle imprese rispondenti, con possibilità di inserimento e cancellazione di imprese.

5. IL FOGLIO ELETTRONICO COME QUESTIONARIO DI RILEVAZIONE PER LE INDAGINI SSI

L'implementazione di una piattaforma di acquisizione elettronica dei dati per le indagini del servizio SSI è scaturita dall'esigenza di migliorare e ottimizzare alcune fasi del processo di produzione, e dalla richiesta, da parte di molte imprese rispondenti, di utilizzare le attuali tecnologie informatiche per la compilazione e trasmissione dei modelli di rilevazione.

Molte e diversificate potevano essere le soluzioni tecnologiche e le modalità di acquisizione elettronica dei dati. La scelta adottata di utilizzare documenti del tipo Foglio Elettronico (Excel) come modelli di rilevazione scaricabili dal sito di data-capturing INDATA, si è basata sulle seguenti considerazioni:

- i modelli di rilevazione utilizzati constano in diversi casi di molte pagine (si pensi al Sistema dei Conti delle Imprese e alla Rilevazione del Costo del Lavoro), e, in generale, difficilmente possono essere compilati in un'unica sessione. Risulta quindi indispensabile

un formato informatico che permetta il salvataggio, da parte del rispondente, delle informazioni compilate anche in forma parziale;

- non è raro il caso in cui per una stessa impresa, ci siano più persone incaricate di compilare le parti del questionario di loro competenza; in questo caso il formato del modello elettronico deve poter essere “trasportabile” da una postazione all’altra;
- i modelli di rilevazione delle statistiche strutturali presentano nella gran parte dei casi variabili numeriche, sulle quali gravano controlli di congruenza (uguaglianza con altre variabili, controlli di soglia, quadrature sulle sommatorie). Tradizionalmente questi controlli vengono effettuati dai revisori dell’UO tramite il software di controllo e correzione (sviluppato in Oracle Forms) a partire dal dato proveniente dalla registrazione esterna. La possibilità di inserire strumenti di controllo e di verifica all’interno del modello elettronico di rilevazione migliora notevolmente la qualità del dato;
- l’utilizzo di maschere per l’inserimento dei dati a carico del rispondente, in ambiente web, è per l’SSI una soluzione poco praticabile, in quanto:
 - la progettazione e creazione delle maschere “on-line” deve essere fatta da personale informatico specializzato (linguaggi Php, Java, etc...) e comporta la creazione di software ad-hoc per la singola indagine, con manutenzione evolutiva per le replicazioni annuali di indagine. La soluzione del foglio elettronico Excel permette di evitare l’impiego di informatici specializzati nello sviluppo di software: il modello in Excel non utilizza funzioni macro, ma solo funzionalità e controlli standard di Excel, che possono essere implementati da personale non informatico;
 - il salvataggio delle compilazioni parziali dei modelli nella modalità on-line, comporterebbe il salvataggio di dati riservati dell’azienda su database di server esposto dell’ISTAT; questa opzione aumenta la complessità dei sistemi di sicurezza definiti per la piattaforma di data-capturing;
- Il foglio elettronico è sicuramente uno degli strumenti più diffusi presso le medie e grandi imprese, e molto diffusa è anche la conoscenza di questo tra i professionisti del campo economico, contabile e amministrativo.

Queste considerazioni hanno portato a utilizzare per le rilevazioni elettroniche in SSI il foglio elettronico come supporto per la compilazione dei modelli da parte delle imprese.

Per le rilevazioni elettroniche di SSI attualmente su INDATA, è stato definito un file Excel con un layout in tutto simile al questionario cartaceo. Ogni sezione del modello è rappresentata da un singolo foglio elettronico, e anche i campi di personalizzazione sono posizionati nella stessa posizione della versione cartacea. Tutte le celle del foglio excel sono bloccate, non possono cioè essere digitate, ad eccezione delle celle dove vanno inserite le informazioni. Su queste celle sono stati inseriti controlli tramite funzioni di convalida dei dati, regole di formattazione condizionale e funzioni di segnalazione delle incongruenze. Inoltre in una cella nascosta e bloccata è stato inserito un identificativo dell’indagine in modo da essere certi che il rispondente stia inviando il questionario Excel che ha scaricato dal sito INDATA relativo all’indagine di pertinenza.

Questi strumenti “leggeri” fanno sì che il rispondente sia agevolato nell’inserimento e modifica dei dati, senza pesanti vincoli di risposta e senza l’implementazione di macro Excel. Sulla base di precedenti esperienze si è riscontrato che la presenza di macro all’interno questionario elettronico in formato Excel complica notevolmente la rilevazione elettronica: per default, quando in Excel si tenta di aprire un file con macro, viene visualizzato un avviso di protezione del sistema, in quanto le macro di Excel possono contenere dei virus. Questo scoraggia notevolmente il rispondente (anche se sufficientemente informato dell’assenza di virus nel software) dall’aprire il foglio elettronico e rispondere alla rilevazione. Sulla base delle esperienze effettuate, possiamo affermare che i controlli “non-macro” implementati nei modelli elettronici SSI consentono di ottenere dati privi di errori derivanti da digitazioni errate e congrui con le regole imposte.

Infine, nei modelli Excel di SSI sono stati attribuiti nomi di variabili su tutte le celle dove vanno inserite le informazioni. Questo semplifica l'implementazione delle procedure di acquisizione dei modelli elettronici, in quanto queste effettuano la lettura delle sole celle dichiarate come "variabili da acquisire".

6. TRATTAMENTO DEI DATI A VALLE DELLA RACCOLTA

Il trattamento dei modelli elettronici delle indagini SSI viene effettuato a cura del personale informatico del Servizio. Questa fase ha usualmente inizio nel momento in cui i modelli elettronici vengono trasmessi dalla piattaforma INDATA verso i server di sede del Servizio SSI. Viste le complessità e le peculiarità dei processi di produzione delle diverse indagini, sono state implementate applicazioni e procedure di gestione dei modelli elettronici pervenuti, con ampio riutilizzo del software e sviluppo di procedure generalizzate. In questo paragrafo saranno descritti in dettaglio le tre sotto-fasi in cui è viene suddiviso il trattamento dei modelli elettronici.

6.1 Trasferimento e archiviazione dei modelli elettronici sui server di lavorazione

L'operazione di spedizione del modello elettronico da parte delle imprese è in realtà una procedura di upload dal computer locale del rispondente verso il server che ospita la piattaforma di data-capturing INDATA.

Ogni ora vengono eseguiti gli script di trasferimento automatico dei modelli dal web-server INDATA verso uno degli application-server di sede Tuscolana (NEPAL). In questo modo INDATA, che è comunque un web-server esposto, non viene "appesantito" dai modelli spediti dalle imprese: questi infatti rimangono su INDATA al massimo per un'ora.

I modelli di tutte le indagini elettroniche SSI vengono archiviati sul server NEPAL. Al momento le fasi di lavorazione (revisione preliminare e acquisizione) dei modelli elettronici sono effettuati su piattaforma Windows, quindi si è reso necessario il trasferimento ulteriore dei modelli su pc-server del servizio. Questo trasferimento viene effettuato tramite applicazione in ambiente web, scritta in PHP e residente su un pc-server del servizio SSI, shinano140.

Tramite il proprio browser, l'utente (il gestore della rilevazione elettronica), richiama una pagina web stabilita per la propria indagine (vedi es. Indagine RCL) e può monitorare l'arrivo dei questionari sul server NEPAL e decidere, nel caso, di scaricarli per le lavorazioni successive.

Qualora l'utente decidesse di scaricare i questionari, cliccando sul pulsante "Download" automaticamente lo script PHP crea, se non esiste già, sul PC server una cartella con un nome parametrizzato con anno-mese-giorno dove verranno trasferiti i questionari scaricati. Sul server NEPAL, in una area di lavoro dedicata all'indagine, i questionari scaricati vengono rinominati e rimangono archiviati come backup.

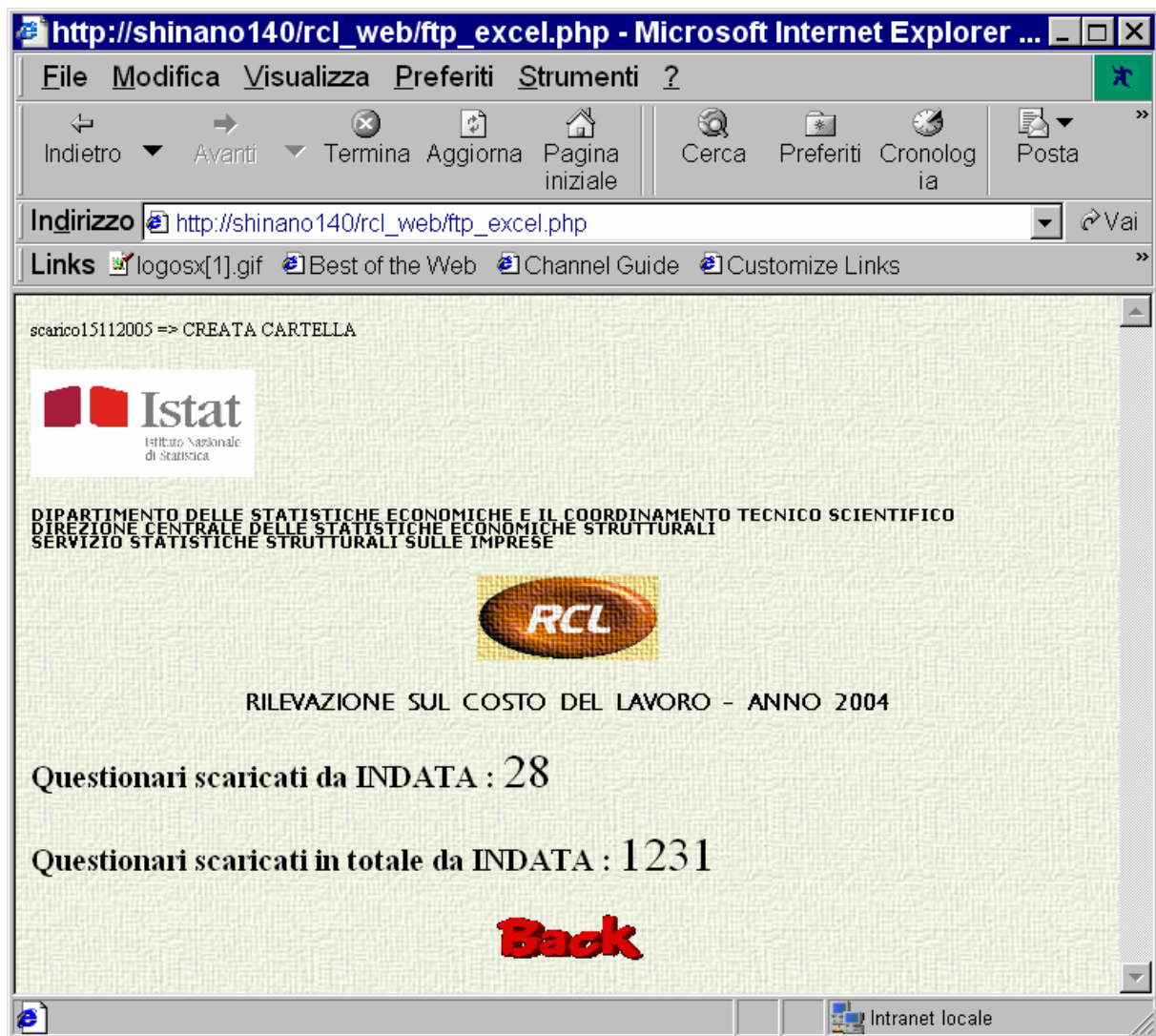
Tutto ciò evita all'utente l'uso di client FTP con tutti i relativi parametri (server, userid, password, tipo di trasferimento, area di partenza e area di arrivo) ed inoltre gli permette di sapere in qualsiasi momento e da qualsiasi postazione della rete ISTAT la situazione degli arrivi. Di seguito vengono presentate le schermate principali di questa applicazione:

Microsoft Internet Explorer window showing a web page from http://shinano140/rcl_web/scarica_ftp.php.

The page content includes:

- Istat logo (Istituto Nazionale di Statistica)
- DIPARTIMENTO DELLE STATISTICHE ECONOMICHE E IL COORDINAMENTO TECNICO SCIENTIFICO
DIREZIONE CENTRALE DELLE STATISTICHE ECONOMICHE STRUTTURALI
SERVIZIO STATISTICHE STRUTTURALI SULLE IMPRESE
- RCL logo
- RILEVAZIONE SUL COSTO DEL LAVORO - ANNO 2004
- Questionari arrivati su INDATA : 28
- Scarica i questionari excel :
- Back button

The status bar at the bottom indicates "Intranet locale".



6.2 Controllo preliminare e revisione interattiva dei modelli elettronici

Le diverse indagini SSI che hanno adottato la modalità elettronica seguono da sempre processi differenziati di lavorazione, con alcune caratteristiche comuni. Data l'intenzione di apportare gradualmente e parallelamente l'innovazione della rilevazione elettronica, si è stabilito di lasciare inalterati i processi esistenti, per garantire la continuità della produzione. In questo senso le procedure informatiche sviluppate per il controllo e acquisizione dei modelli elettronici hanno "riprodotto" le fasi di produzione esistenti, in molti casi differenti tra indagine e indagine.

La prima fase del processo tradizionale di produzione per le indagini che utilizzano il modello cartaceo è la revisione preliminare dei modelli compilati dalle imprese e inviati alle strutture di produzione. Il personale di produzione verifica ogni singolo modello pervenuto. La qualità della verifica varia da indagine a indagine: a volte si verificano solo le informazioni anagrafiche e le variabili più importanti, altre volte si fanno controlli sui modelli dell'anno precedente, in altri casi ancora si effettuano controlli di congruenza (manuali) e si apportano sul modello note e suggerimenti per agevolare le operazioni di data-entry dei modelli.

Il fattore comune tra tutte le indagini in questa prima fase riguarda l'utilizzo di un sistema centralizzato di gestione del monitoraggio del pervenuto, realizzato e gestito da personale SSI in ambiente client-server Oracle. I revisori utilizzano le maschere di questo sistema per assegnare un codice al questionario, corrispondente a una modalità di risposta, per esempio: questionario

utilizzabile, impresa trasferita, fusa, cessata, fuori campo di osservazione, etc. etc. Esiste attualmente anche un codice ad-hoc per la risposta valida elettronica

Queste informazioni inserite sono utilizzate in vari modi: monitorare i modelli pervenuti, generare liste di sollecito, effettuare statistiche su rispondenti e non rispondenti.

Una volta assegnati i codici di risposta a un insieme consistente di modelli, questi vengono assemblati per "pacco". Più pacchi di modelli costituiscono un "invio" alla ditta di registrazione esterna, che provvede all'imputazione elettronica dei modelli cartacei.

Di seguito viene illustrata l'applicazione della modalità elettronica in questa fase di produzione, per le diverse indagini realizzate:

- Sistema dei Conti delle Imprese:

I modelli elettronici pervenuti sul pc-server di Servizio (Shinano140) vengono sottoposti ad una prima acquisizione, con una applicazione sviluppata in VBA Access. Questa apre tutti i modelli elettronici pervenuti, verifica le informazioni anagrafiche e acquisisce i dati significativi per un controllo preliminare dei modelli. Questa procedura automatica assegna a ogni modello un codice di risposta e predispose i modelli per la verifica "interattiva", attraverso la quale è possibile verificare e eventualmente correggere i modelli. La procedura di verifica interattiva è stata implementata con maschere Access che accedono ai dati acquisiti (sempre su DB Access) e che consentono l'apertura dei modelli Excel in lavorazione. Il personale dell'indagine effettua la verifica per i soli modelli che non sono risultati validi al termine del controllo preliminare automatico. Sempre con procedura VBA Access, vengono generati periodicamente i codici risposta assegnati a ogni modello, che vengono poi vengono trasmessi e caricati sul database del Sistema di Monitoraggio del Pervenuto. In questo modo vengono mantenute omogenee le informazioni relative al materiale pervenuto, sia esso cartaceo o elettronico.

- Rilevazione Produzione Industriale (PRODCOM):

Questa indagine tradizionalmente cura molta attenzione a questa prima fase: i modelli vengono dapprima smistati a seconda della tipologia produttiva dell'unità locale (a cui è assegnato un repertorio merceologico), quindi si procede con una prima revisione accurata del modello, con controlli anche sugli anni precedenti e molti contatti telefonici di chiarimento con le imprese.

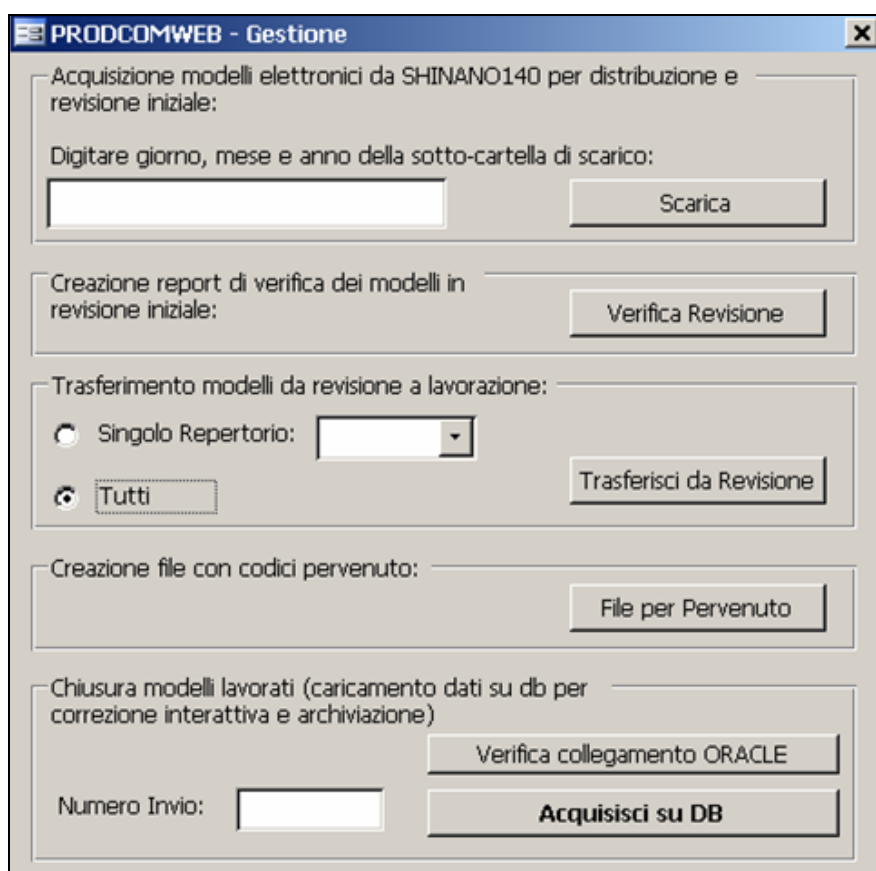
L'applicazione di gestione, sempre implementata in VBA Access, effettua primo trasferimento automatico dal pc-server di Servizio (shinano140), verso le aree revisione iniziale (su pc-server della struttura di produzione). La procedura di trasferimento legge uno per uno i modelli elettronici pervenuti. Se la lettura ha esito negativo (errore nel file), il modello viene trasferito su una cartella che ospita i modelli errati e verificato manualmente dai gestori del sistema. Se la lettura non dà errori, il modello viene trasferito nell'area di Revisione Iniziale, all'interno della sottocartella relativa al repertorio merceologico a cui fa riferimento il modello. In questa area sono definite 16 cartelle, tante per quanti sono i repertori merceologici (tipologie di produzione) dell'indagine PRODCOM.

I modelli distribuiti per tipologia produttiva vengono revisionati tutti con apertura del file in Excel: le 16 sotto-cartelle della revisione iniziale sono condivise verso i clienti dei revisori (un revisore per tipologia produttiva dell'unità locale). I revisori aprono dal loro pc i modelli nella sottocartella di loro competenza e effettuano la revisione iniziale. Al termine della revisione, assegnano al modello un codice relativo alla modalità di risposta. Questa assegnazione avviene tramite casella combinata posta all'interno del foglio excel, in un campo contrassegnato come riservato ISTAT. Fino a quando questa assegnazione non viene effettuata, il modello è considerato non lavorato.

Periodicamente viene attivata una procedura che provvede al trasferimento dei modelli lavorati in prima revisione verso l'area di acquisizione vera e propria. Per ogni repertorio, vengono letti i modelli in revisione, in particolare viene letta la casella con il codice risposta. Se il codice non è presente, il modello è considerato ancora in lavorazione, quindi non viene trasferito. Se invece il codice è stato assegnato, la procedura trasferisce il modello nell'area di acquisizione. A seconda del codice di risposta assegnato, la procedura trasferisce il modello nella sotto-cartella dei modelli Utilizzabili (che poi saranno caricati sul db di produzione) o nella sotto-cartella dei Non Utilizzabili.

Una ulteriore procedura automatica effettua il caricamento dei codici di risposta dei modelli elettronici verso il database del Sistema Informativo di Monitoraggio del Pervenuto. Per questa ultima fase valgono le considerazioni fatte in precedenza.

Di seguito viene presentata la maschera utilizzata dai gestori della rilevazione elettronica PRODCOM per avviare le procedure automatiche descritte:



The screenshot shows a web application window titled "PRODCOMWEB - Gestione". It contains several sections for automated tasks:

- Acquisizione modelli elettronici da SHINANO140 per distribuzione e revisione iniziale:** Includes a text input field for "Digitare giorno, mese e anno della sotto-cartella di scarico:" and a "Scarica" button.
- Creazione report di verifica dei modelli in revisione iniziale:** Includes a "Verifica Revisione" button.
- Trasferimento modelli da revisione a lavorazione:** Includes radio buttons for "Singolo Repertorio:" (with a dropdown menu) and "Tutti", and a "Trasferisci da Revisione" button.
- Creazione file con codici pervenuto:** Includes a "File per Pervenuto" button.
- Chiusura modelli lavorati (caricamento dati su db per correzione interattiva e archiviazione):** Includes a "Verifica collegamento ORACLE" button and a "Numero Invio:" input field with an "Acquisisci su DB" button.

- Stima Provvisoria del Valore Aggiunto:

Questa è una delle poche indagini SSI (e l'unica di questo gruppo) per cui non si ricorre alla registrazione esterna dei modelli di rilevazione, dal momento che le variabili osservate sono in numero contenuto (il modello consta di una sola pagina) e che l'indagine ha vincoli di tempestività molto forti. Il processo di produzione per questa fase ripercorre quello delle altre indagini, con l'unica differenza che i modelli che superano la prima revisione vengono raggruppati e assegnati a personale interno ISTAT, per l'inserimento controllato dei dati (applicazione client-server Oracle).

La fase di revisione iniziale dei modelli elettronici è a cura del personale dell'indagine: i modelli pervenuti vengono visualizzati e controllati uno ad uno. Anche in questo caso è stato previsto all'interno del modello elettronico una casella combinata definita all'interno

di in un campo contrassegnato come riservato ISTAT. Questa consente l'assegnazione del codice relativo alla modalità di risposta. Fino a quando il revisore non assegna un codice di risposta, il modello è considerato non lavorato.

Periodicamente il gestore della rilevazione elettronica (una risorsa dell'indagine) avvia la procedura di acquisizione, sviluppata anch'essa in linguaggio VBA (maschere Access). Questa effettua la lettura dei modelli lavorati, genera il file con i codici di risposta e inserisce i dati dei modelli su tabelle Oracle transitorie. Una volta allineate le risposte con quelle già pervenute (controllo sul Sistema di Monitoraggio del Pervenuto), si procede con il caricamento dei dati sul DB del sistema di inserimento, controllo e correzione.

- Rilevazione del Costo del Lavoro:

Questa indagine è l'unica non annuale tra quelle per cui è stata predisposta la modalità elettronica. La fase di prima revisione per questa rilevazione quadriennale è del tutto simile alle altre indagini annuali. Anche per questa si ricorre alla registrazione esterna dei modelli, a differenza però delle altre indagini, non si effettua il controllo e la correzione interattiva dei dati, ma si utilizzano procedure batch (generalizzate e ad-hoc) per il controllo e la correzione automatica.

Relativamente all'acquisizione elettronica, anche per questa indagine la revisione iniziale dei modelli è a cura del personale dell'indagine, con apertura e controllo in Excel di tutti i modelli pervenuti. In questo caso è stata realizzata una applicazione VBA in Excel. Questa presenta una barra degli strumenti supplementare contenente i bottoni utili per assegnare il file Excel a una delle seguenti categorie: modello valido, incompleto, da sollecitare, fuori indagine. Il revisore sceglie uno di questi pulsanti al termine della revisione iniziale interattiva, con il file Excel attivo, così si attiva la procedura di trasferimento del modello dall'area di prima revisione all'area di acquisizione.

Periodicamente viene attivata la procedura di acquisizione dei modelli revisionati (sviluppata in VBA Access). Questa legge i modelli pervenuti e revisionati e genera un file sequenziale con i dati del modello, secondo un tracciato record messo a punto dal personale dell'indagine, molto simile al tracciato dei file provenienti dalla registrazione esterna.

I file sequenziali di dati vengono trasmessi ai responsabili delle elaborazioni dell'indagine, che provvede, tramite procedure SAS, al controllo della congruenza dei dati, alla generazione dei codici risposta per il Sistema di Monitoraggio del Pervenuto, all'accodamento dei dati sul dataset principale dell'indagine.

- Rilevazione delle Imprese di Telecomunicazioni:

Questa indagine, che osserva circa 350 imprese di telefonia fissa, mobile e ISP, ha utilizzato, per l'edizione in corso, la modalità elettronica esclusiva: alle imprese non è stato inviato il modello cartaceo di rilevazione, bensì la sola lettera di comunicazione dell'avvio dell'indagine, con le indicazioni per il download, la compilazione e la spedizione del modello elettronico.

Data l'esiguità delle risposte ricevute, non esistono procedure interattive di controllo e correzione dei modelli, né gestione del materiale pervenuto. Il personale dell'indagine, fino alla scorsa edizione, si occupava della stampa, imbustamento e spedizione dei modelli, poi, per la ricezione, veniva effettuato un data-entry dei dati con maschere Access.

Con la modalità elettronica saranno eliminate tutte le operazioni di spedizione cartacea e di inserimento dati dai modelli cartacei: con applicazione VBA saranno letti i modelli ritenuti validi e generati file di output con i dati secondo il formato stabilito dal responsabile dell'indagine.

6.3 Acquisizione dei dati e caricamento sui database di produzione delle indagini

Le indagini SSI che effettuano la fase di controllo e correzione interattiva dei dati, utilizzano applicazioni client-server, con maschere di visualizzazione, modifica e inserimento dei dati realizzate in Oracle Forms e residenti sui pc del personale di revisione; i dati sono archiviati su DB Oracle, residente sui server Linux della sede Tuscolana.

La maggior parte delle indagini fanno ricorso a registrazione esterna dei modelli, quindi questi sistemi di produzione acquisiscono i file sequenziali provenienti dalla registrazione esterna e li presentano nelle maschere per la visualizzazione e correzione assistita. Nel caso invece dell'indagine SPVA, il sistema consente anche l'inserimento assistito dei dati da parte del personale dell'indagine.

Le applicazioni realizzate per le singole indagini condividono dati e procedure, e il sotto-sistema del Monitoraggio del Pervenuto; in questo senso l'insieme delle applicazioni sviluppate per le differenti indagini si colloca all'interno del Sistema Informativo Integrato di Produzione per le Indagini SSI.

L'adozione della modalità elettronica per queste indagini rendeva vincolante la compatibilità con i processi esistenti e la necessità di non modificare i sistemi informativi di produzione esistenti. Di seguito viene illustrata l'applicazione della modalità elettronica in questa fase di produzione, per le diverse indagini realizzate:

- **Sistema dei Conti delle Imprese:**

Il modello di rilevazione di questa indagine contiene alcune sezioni specifiche che non vengono elaborate dal personale dell'indagine, ma dalle strutture di competenza. In particolare nel questionario SCI, oltre ai Conti Economici, vengono rilevate variabili relative alla Ricerca e Sviluppo, alla Responsabilità Sociale delle Imprese, alle Spese per Protezione dell'Ambiente, al Controllo dell'Impresa. Tradizionalmente i modelli SCI venivano inviati alla ditta di registrazione, la quale restituiva, per ogni singolo invio, 5 file sequenziali: uno per l'indagine SCI, e 4 per le sezioni di altra competenza.

Con l'introduzione della modalità elettronica, si è stabilito di evitare la modifica delle procedure esistenti di acquisizione dei dati provenienti dalla registrazione esterna. Per questo motivo la procedura di acquisizione dei modelli elettronici (sviluppata in VBA) produce come output file sequenziali con lo stesso tracciato record dei file provenienti dalla ditta di registrazione e stabiliti nel piano di registrazione dell'indagine.

Per le variabili "proprietarie" di SCI, la procedura legge i modelli che hanno superato la prima revisione e produce in output un file sequenziale. Il sistema di produzione dell'indagine SCI (Oracle) acquisisce questo file esattamente come se fosse un file proveniente dalla ditta di registrazione: il dato viene quindi sottoposto al tradizionale controllo di congruenza preliminare, prima di essere revisionato dal personale, con le maschere di controllo e correzione interattiva.

Gli altri 4 file vengono generati e trasmessi alle strutture di competenza (SIP, SSI/B, SSI/D, FATS) che ne curano l'elaborazione. Anche per questi non è stato necessario richiedere una modifica delle procedure di caricamento, in quanto il tracciato è lo stesso dei file provenienti dalla registrazione esterna.

Un ulteriore file viene generato dalla procedura di acquisizione: quello relativo alle informazioni anagrafiche, contenente le variazioni e le note compilate dalle imprese contenute nei modelli elettronici. Questo file viene trasmesso al personale dell'archivio Imprese (ASIA), che cura la verifica e l'aggiornamento delle informazioni principali sulle Imprese.

- Rilevazione Produzione Industriale (PRODCOM):
Per l'indagine PRODCOM si è stabilito di far confluire i dati provenienti dai modelli elettronici direttamente nelle tabelle del DB Oracle del sistema di produzione. I modelli elettronici revisionati vengono acquisiti tramite una procedura automatica (sviluppata in VBA) che legge, per ogni modello, le celle compilate e scrive i dati letti e elaborati su tabelle transitorie del DB di produzione. Una volta verificati dal gestore del database PRODCOM, questi dati vanno ad alimentare le due tabelle principali su cui confluiscono i dati (da modello elettronico e da modello cartaceo): la tabella QUESTIONARIO, con le informazioni di sintesi relative all'unità locale di produzione, e la tabella PRODOTTI, con le informazioni oggetto della rilevazione. Una volta sul database di produzione, i dati provenienti dai modelli elettronici seguono l'usuale iter di controllo e correzione interattiva e di elaborazione per la produzione delle stime.
- Stima Provvisoria del Valore Aggiunto:
Dato che l'indagine SPVA effettua una registrazione interna dei dati, si è stabilito di inserire i dati provenienti dai modelli elettronici direttamente nelle tabelle del DB Oracle del sistema di produzione. I modelli elettronici vengono acquisiti tramite una procedura automatica (sviluppata in VBA) che legge, per ogni modello, le celle compilate e scrive i dati letti e elaborati su tabelle transitorie del DB Oracle. Una volta verificati dal gestore del database SPVA, questi dati vengono inseriti nella tabella principale del DB di produzione. Secondo questa procedura, il personale dell'indagine ha a disposizione l'usuale maschera per l'inserimento dei dati già riempita con i dati del modello elettronico. In questo modo i compiti del personale si limitano al controllo e alla eventuale correzione del dato, senza alcuna operazione di inserimento dati.

7. EVOLUZIONI E SVILUPPI FUTURI

Di seguito vengono evidenziate le prospettive di potenziamento delle architetture esistenti e le evoluzioni applicabili alle procedure esistenti, allo scopo di migliorare, potenziare e ottimizzare i processi di gestione e acquisizione dei dati relativi alle rilevazioni che osservano le imprese.

7.1 Evoluzione della piattaforma indata

Il server indata.istat.it, dedicato all'acquisizione dei dati in ambiente AIX sarà sostituito, nel primo semestre 2006 con HW di capacità e potenza molto più elevate dell'attuale: 2 server in architettura ridondata, e in alta affidabilità, ognuno dotato di 4 processori, 8 GB di RAM. I server saranno connessi in fibra ottica (schede Fiber Channel) verso una SAN(Storage Area Network) esterna a velocità di trasferimento di 2Gb/sec.

Il sistema operativo sarà LINUX REDHAT Enterprise Server, in accordo con le strategie recentemente messe a punto in ambito DCMT. Relativamente al parco software sarà sostanzialmente confermato la piattaforma LINUX, APACHE WEBserver, MYSQL, PHP. La versione di PHP sarà aggiornata da 4 a 5. Questa combinazione, anche nota come LAMP, è sempre più diffusa nell'ambito dei siti WEB e garantisce vantaggi prestazionali e disponibilità di funzioni applicative già sviluppate.

Con il rilascio del nuovo HW, il prossimo indata dovrà ospitare con certezza le applicazioni WEB sviluppate prima del 2002 con il sistema Teleform e al momento attive sul sito teleform.istat.it

(vedi tabella al punto 4), in ambiente WINDOWS. Le applicazioni saranno migrate da linguaggio ASP a PHP; il sito teleform sarà dismesso.

Il nuovo sito *indata* si candiderà inoltre ad ospitare le future applicazioni WEB di acquisizione, relative a qualunque rilevazione dell'Istituto, realizzate secondo i modelli di sviluppo più diffusi. Per questo sarà svolta un'attività di analisi al fine di verificare la possibilità di mettere a disposizione sullo stesso server *indata*, oltre all'ambiente di elaborazione PHP5 anche un ambiente idoneo ad eseguire applicativi sviluppati con linguaggi e strumenti del mondo JAVA.

Nell'ambito delle soluzioni web per la compilazione off-line, sarà presto introdotta su *INDATA* la possibilità di compilare of-line anche i modelli elettronici in formato PDF. Questa caratteristica sarà integrata con la soluzione *TELEFORM*, e amplierà notevolmente le funzioni attuali, consentendo ai client rispondenti, senza onere per loro, di salvare il modello PDF che trovano sul sito, compilarlo offline con salvataggio delle informazioni parziali, e inviarlo in maniera automatica.

7.2 Sviluppi futuri per le indagini che osservano le imprese

Il risultato ottenuto con le rilevazioni elettroniche introdotte nell'ultimo anno ha portato a una sempre maggiore richiesta da parte delle indagini del settore economico strutturale, sia in termini di estensione della modalità elettronica verso indagini che ancora utilizzano la modalità tradizionale di raccolta dei dati, che in termini di miglioramento e ottimizzazione dei processi elettronici di raccolta e trattamento dei dati.

Le indagini che osservano le imprese per cui sarà introdotta la modalità elettronica su piattaforma *INDATA* nei prossimi mesi sono le seguenti: Rilevazione sulle Tecnologie dell'Informazione e della Comunicazione, Rilevazione sulla Ricerca e Sviluppo, Rilevazione sulla Innovazione tecnologica, Rilevazione sulle Piccole e Medie Imprese.

Per la prima rilevazione citata è stato pianificato lo sviluppo della piattaforma *TELEFORM* (modello in formato PDF) con soluzione di tipo web "offline".

Per incentivare le imprese rispondenti e ridurre il fastidio statistico, si propone la definizione di strumenti e servizi informativi integrati. Tra questi si citano a titolo esemplificativo:

- realizzazione di un portale delle imprese rispondenti, con servizi personalizzati a seconda dell'impresa, una sorta di mini-sito per la singola impresa. In questo modo, quando il rispondente si collega, può, ad esempio, accedere alle indagini che lo riguardano, e in generale alle informazioni ufficiali dell'Istituto;
- implementazione di procedure per l'accesso facilitato e integrato alla piattaforma di data-capturing, con funzioni del tipo single sign-on per l'identificazione del rispondente;
- disponibilità, per le imprese rispondenti, di accesso a servizi informativi specifici (dati e analisi ad-hoc), e ai servizi usuali di diffusione dei dati, con collegamento "guidato" verso i sistemi di diffusione dell'ISTAT;
- realizzazione di strumenti di supporto alla compilazione del modello elettronico, con possibilità di utilizzo, da parte del rispondente, di dati strutturati già in suo possesso. Nello specifico si ipotizza la possibilità di una importazione di dati in formato XBRL (eXtensible Business Reporting Language) all'interno del questionario elettronico, per quelle variabili compatibili con le tassonomie stabilite dalla comunità di esperti che sviluppano questo linguaggio;
- possibilità di utilizzo di Fogli Elettronici "open" e in ambienti diversi da Windows per la compilazione dei modelli elettronici; in questo senso si possono prevedere più versioni del modello scaricabile, e diverse linee di acquisizione a seconda del formato.

Relativamente ai processi di back-end di trattamento dei dati elettronici, sono in fase di analisi e progettazione diverse soluzioni. Il miglioramento delle procedure di acquisizione e controllo dei modelli elettronici si rende necessario in quanto, allo stato attuale i modelli elettronici in formato Excel, vengono elaborati in ambiente Windows (linguaggio VBA) su pc dedicati e potenziati, ma sottoposti a sistemi di sicurezza (antivirus, backup, aggiornamento sistemi operativi) non diversi da quelli utilizzati per le usuali postazioni di lavoro.

Sono in studio diverse ipotesi per consentire l'elaborazione e il controllo dei modelli Excel in ambiente Linux. In questo modo i modelli elettronici pervenuti rimangono residenti sui server di sede (IRAN, NEPAL), e possono usufruire dei sistemi di sicurezza di cui sono dotati i server dell'Istituto.

Nello specifico sono in fase di studio e di progetto i seguenti componenti applicativi:

- procedura di lettura di fogli elettronici Excel realizzata in linguaggio PHP. Con questa procedura l'acquisizione dei modelli elettronici potrebbe avvenire direttamente sul server INDATA, oppure su uno dei server di sede (IRAN o NEPAL);
- procedura di revisione interattiva dei modelli elettronici in ambiente Linux, tramite lo strumento Foglio Elettronico della suite OpenOffice;
- procedura di acquisizione dei modelli elettronici, con output dei dati in formato XML, allo scopo di agevolare l'elaborazione dei dati dei questionari, e consentire il salvataggio degli stessi su database Oracle (caratteristica XML DB della Release 2 di Oracle 9i, già presente in Istituto).

RIFERIMENTI

1. Balestrino R., Politi M., "A complete system of data capture to improve timeliness of short statistics", Eurostat Seminar on short-term statistics- Improving timeliness and co-operation, 2001
2. Balestrino R., Larese P., "The use of a data capture tool in statistical surveys on enterprises", ECE/Eurostat/OECD Meeting on the management of statistical information systems (MSIS), 2004
3. Mick P. Couper, 2000. "Web Surveys: a Review of Issues and Approaches," Public Opinion Quarterly, volume 64, number 4 (Winter), pp. 464-481.
4. Garritano F., "Generare file Excel personalizzati da PHP", documento interno Istat
5. Fanfoni L., "Descrizione delle soluzioni implementate nella rilevazione elettronica per l'indagine sul Sistema dei Conti delle Imprese, edizione 2003", documento interno Istat
6. Fanfoni L., "Le soluzioni implementate nella rilevazione elettronica per l'indagine annuale sulla produzione industriale (PRODCOM), anno 2004", documento interno Istat