

ISTITUTO NAZIONALE DI STATISTICA

DIPARTIMENTO DELLE STATISTICHE ECONOMICHE

**COEWEB: la banca dati on line
sulle statistiche del commercio con l'estero**

*Aspetti statistici ed informatici connessi alla
realizzazione della banca dati on line sulle
statistiche del commercio con l'estero*

Natale Renato Fazio (nafazio@istat.it)
Stefano Menghinello (menghine@istat.it)

Aggiornato al 29/01/2003

Premessa

La realizzazione della banca dati sulle statistiche del commercio con l'estero nasce dall'esigenza di fornire una risposta articolata e completa alla crescente richiesta informativa sui fenomeni connessi all'interscambio commerciale dell'Italia. Essa costituisce non solo un contenitore informativo dei principali risultati delle indagini sul commercio con l'estero ma mira ad organizzare tutta l'informazione disponibile, incluse le recenti elaborazioni su operatori e imprese che realizzano scambi commerciali con l'estero, in funzione dei seguenti obiettivi:

1. Accesso trasparente e coerente al contenuto informativo della banca dati;
2. Ottimizzazione dell'architettura informatica al fine di rilasciare informazioni on line in modo efficiente e tempestivo;
3. Creazione di un sistema informativo che fornisca non solo dati ma anche informazioni statistiche complete sui fenomeni in esame;
4. Modalità di accesso flessibile all'informazione statistica a seconda delle specifiche esigenze dell'utente;
5. Rispetto della normativa vigente in materia di riservatezza e delle relative procedure definite dal Dipartimento delle statistiche economiche per la diffusione delle statistiche del commercio con l'estero.

Nella costruzione della banca dati, tali obiettivi sono stati affrontati, in molti casi, congiuntamente (si pensi ad esempio ai punti 1 e 5), tuttavia sono di seguito analizzati separatamente al fine di evidenziare con chiarezza i principali caratteri e le potenzialità della banca dati.

1. Contenuto informativo e struttura statistica della banca dati

Le indagini sul commercio con l'estero costituiscono da sempre un patrimonio informativo di notevole valore per varie tipologie di utenza (associazioni industriali, ricercatori, enti internazionali). Il significativo numero di variabili presenti nelle rilevazioni sui flussi commerciali, nonché l'elevato numero di modalità associate ad alcune di esse (paesi, merci, territorio) conferiscono a tali statistiche un carattere di spiccata multidimensionalità che permette, potenzialmente, di realizzare infinite combinazioni tra variabili e modalità diverse. Recentemente, si sono aggiunte, oltre alla revisione e all'arricchimento delle elaborazioni relative ai numeri indici, nuove elaborazioni sugli operatori e sulle imprese che effettuano scambi commerciali con l'estero.

I nuovi software ed architetture delle banche dati permettono di gestire in modo semplice e funzionale notevoli moli di dati, tuttavia presentano un approccio sostanzialmente “anarchico”, che legittimando ogni tipologia di incrocio, può determinare risultati incompleti da un punto di vista informativo, incoerenti da un punto di vista statistico ed incompatibili con il principio di riservatezza delle statistiche del commercio con l'estero. Pertanto risulta necessario organizzare tale patrimonio informativo secondo una struttura che rispetti i seguenti principi:

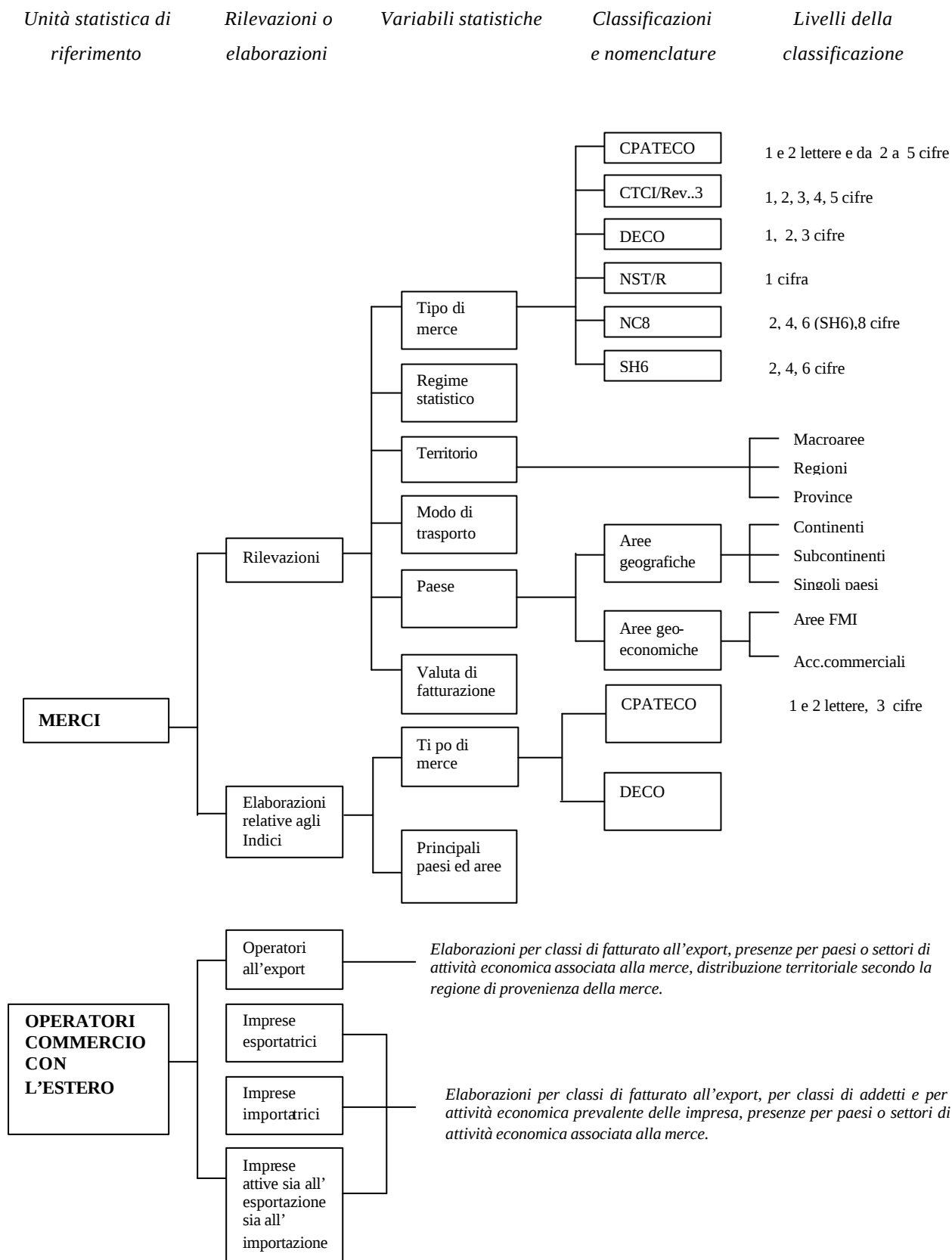
- *Coerenza* – E' necessario evitare la realizzazione di risultati incoerenti o ambigui da un punto di vista statistico.
- *Qualità dei dati* – E' opportuno mantenere le possibilità di incrociare variabili nonché di raggiungere elevati livelli di dettaglio entro limiti che rendano accettabile la qualità statistica dei risultati ottenuti.
- *Rispetto della riservatezza* - E' vincolante escludere dalla diffusione ogni modalità di incrocio o di dettaglio che risulti incompatibile con la normativa vigente in materia di tutela della riservatezza.

I criteri statistici adottati nella definizione della struttura della banca dati sono i seguenti:

- individuazione delle unità statistiche di riferimento;
- individuazione delle principali variabili di classificazione ;
- individuazione delle principali nomenclature e classificazioni connesse alle variabili selezionate;
- definizione, per le sole statistiche relative ai flussi commerciali, di una “griglia” che disciplini le possibilità di incrocio tra variabili e la possibilità di raggiungere determinati livelli di dettaglio

I primi tre aspetti sono riassunti nella figura seguente, mentre per quanto riguarda l'ultimo punto, data la sua notevole rilevanza anche come strumento per la tutela della riservatezza, si rimanda al *Piano per la diffusione delle statistiche del commercio estero* contenuto nel sito e nei documenti in allegato.

Fig.1 – STRUTTURA DELLE INDAGINI STATISTICHE DEL COMMERCIO CON L'ESTERO



2. Progettazione informatica della banca dati

La banca dati on line sulle statistiche del commercio con l'estero è parte di un più ampio progetto strategico dell'ISTAT di innovazione tecnologica dell'intero patrimonio informativo ed informatico. Tale progetto, avviato dal 1995, ha subito una forte accelerazione con l'emergere delle problematiche relative all'anno 2000 e la definizione di accordi e convenzioni tra l'ISTAT ed altri enti (ICE, Confindustria) per la diffusione delle statistiche del commercio con l'estero.

L'idea di un sistema informativo statistico (S.I.S.) del commercio con l'estero (COE) è stata concepita a fine 1998. L'obiettivo è stato quello di giungere a una completa riprogettazione del sistema informativo centralizzato esistente, entro i termini imposti dalla data dell'anno 2000, creando un unico sistema che potesse comprendere tutte le fasi di produzione dell'informazione statistica.

Questo documento si concentrerà su un modulo del S.I.S. del COE che è appunto la banca dati delle statistiche del commercio estero su Internet (COEWEB). Esso costituisce un sottoinsieme del data warehouse del commercio estero che include, oltre alla diffusione standard, anche la possibilità di accesso ai dati elementari. Quest'ultima sezione è tuttavia limitata ai ricercatori ISTAT per fini istituzionali o di ricerca.

La progettazione informatica di COEWEB ha le sue premesse nei seguenti elementi:

1. analisi delle funzionalità richieste dall'utenza (riconoscimento utente, navigazione, querying e reporting , etc.);
2. analisi dei vincoli alla struttura dei dati imposte dal Piano per la diffusione;
3. scelta dell'approccio di fondo nella realizzazione della banca dati (logica relazionale, non relazionale, tecniche di data warehousing) in relazione alla struttura dei dati e alle modalità di accesso ad Internet;

I primi due punti, corrispondenti all'analisi dei requisiti dell'utente nei diversi paradigmi di ingegneria del software, sono fondamentali per la buona riuscita del progetto software. E' per tale motivo che a tale fase hanno collaborato numerosi esperti di diffusione dei dati del commercio estero nonché una rappresentanza di utenti esterni. Il problema in questa fase è essere certi di realizzare una applicazione "funzionale" nei termini dello standard ISO 9126 sulla qualità del prodotto software e quindi garantire le seguenti caratteristiche :

- adeguatezza: coerenza delle funzioni offerte dal prodotto rispetto alle esigenze dell'utenza;
- accuratezza: la correttezza e precisione dei risultati prodotti;

- sicurezza: capacità di prevenire accessi non autorizzati alle informazioni e funzioni gestite dal prodotto

Allo stesso tempo il prodotto da realizzare deve soddisfare le altre 5 caratteristiche principali dello standard ISO 9126 e cioè affidabilità, usabilità, efficienza, manutenibilità e portabilità.

L'analisi sui livelli ottimali per ciascuna caratteristica è stato punto di partenza per la successiva fase di progettazione che ha trasformato il modello dei requisiti in quattro livelli di dettaglio: la struttura dei dati, l'architettura del sistema, la rappresentazione dell'interfaccia ed i componenti.

La progettazione della struttura dati, soprattutto per “magazzini dati “ di grandi dimensioni come quello del commercio estero (lo spazio fisico totale del S.I. del COE è di circa 120 GB, lo spazio fisico della banca dati COEWEB è di 50 GB) si è rilevata un passaggio fondamentale per raggiungere gli obiettivi di tempestività nel reperimento delle informazioni imposti dal sistema. La ridondanza che viene introdotta a livello di database è tipica in ambiente data warehousing in cui, usualmente, non solo si denormalizza l'architettura di base ma si creano fisicamente tabelle dei fatti con dati aggregati già precalcolati ed indici ottimizzati sulle tabelle. Questa soluzione sarebbe inopportuna per database di piccole dimensioni, ma nel nostro caso, dove ogni mese l'aggiornamento a livello di dati elementari implica un inserimento nel data warehouse di circa 2.5 milioni di record, è l'unico modo per permettere una facile gestione dei dati secondo schemi definiti puntualmente nel piano della diffusione.

Per quanto riguarda l'architettura del sistema dal punto di vista tecnologico si è scelto di utilizzare standard di mercato quali il RDBMS Oracle per la gestione dei dati, l'architettura a componenti COM/DCOM di Microsoft con IIS quale web server ed ASP per la logica di business. Tale soluzione si è rilevata ottimale e permette, tra l'altro, un più facile reperimento di risorse esperte per la manutenzione rispetto ad altre soluzioni molto più costose, più complesse in fase di sviluppo e manutenzione e non soddisfacenti nel rapporto qualità/prezzo.

La figura 3 mostra l'architettura del modulo di diffusione del sistema informativo statistico del commercio estero. Tale architettura si compone di due parti:

- area Intranet (DWCOE);
- area Internet (COEWEB);

Per quanto riguarda l'area Intranet, dall'ambiente di produzione, che comprende la fase di acquisizione dei dati, la correzione automatica e quella interattiva da parte dei revisori, il dato definitivo proveniente dalle due indagini Intrastat ed Extrastat viene reso omogeneo e quindi alimenta il meccanismo di aggiornamento del data warehouse

L'area Internet è in pratica un mirroring di un sottoinsieme dell'area Intranet aggiornato mensilmente off-line per garantire il massimo livello di sicurezza.

Questa architettura del sistema permette di strutturare facilmente i livelli di accesso ai dati: attualmente, mentre gli operatori del servizio del commercio estero ed i ricercatori autorizzati hanno la possibilità di utilizzare applicazioni ad un basso livello di dettaglio o di operare, a scopo di studio o analisi, a livello di microdati, gli uffici regionali e il data-shop possono accedere solo al livello stabilito dal piano per la diffusione. Val la pena accennare che un utente esterno, accedendo ai dati dal web non potrà mai ottenere dati riservati per barriere non solo applicative ma anche strutturali, poiché il database a cui accede è stato depurato fisicamente da qualunque dato riservato.

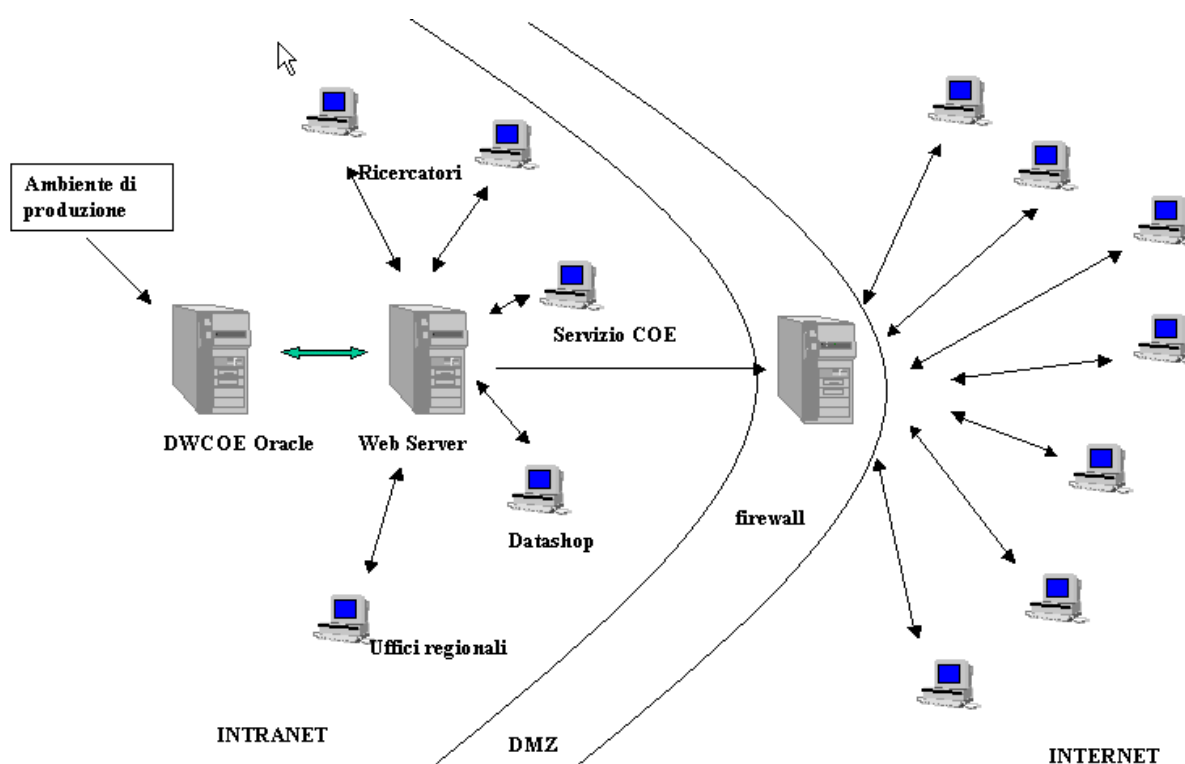
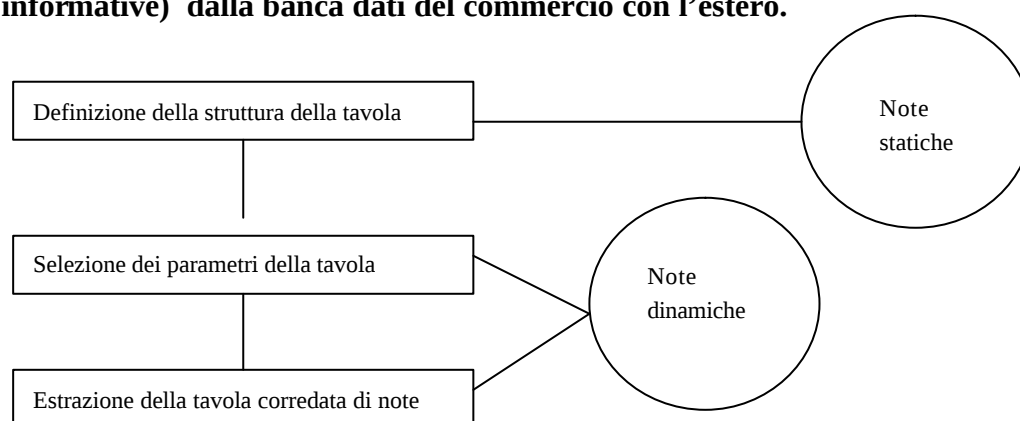


Fig.3 - Architettura del data warehouse commercio estero

3. Dal dato numerico all'informazione statistica

Un sistema informativo statistico, ed ancor di più una banca dati on-line su Internet, non può prescindere dalla messa a disposizione all'utente dei metadati senza i quali l'informazione statistica rischia di divenire una scatola nera fuorviante e rischiosa. Le "insidie" del dato statistico, connesse a problemi definitori e classificatori, infatti, sono numerose ed è ormai evidente come, per fornire una corretta informazione statistica, sia sempre necessario corredare i dati con metainformazioni. Nel caso delle statistiche del commercio con l'estero tale affermazione acquista una rilevanza particolare poiché la mole di informazioni, generali o puntuali, necessaria per poter analizzare correttamente il dato è notevole e di non facile accesso. Si pensi ad esempio ai significativi cambiamenti che ogni anno occorrono nella geonomenclatura (scissioni di paesi, cambiamenti di denominazione, modifica dei confini delle aree geoeconomiche), nella nomenclatura delle merci (scissioni, accorpamenti, cambi di codifica o alfabeto). Tale obiettivo è stato raggiunto corredando la banca dati di informazioni sulla struttura delle rilevazioni ed elaborazioni nonché di dettagliate note metodologiche.

Fig.4 – Schema di base relativo alla costruzione di una tavola statistica completa (dati numerici e note informative) dalla banca dati del commercio con l'estero.



Le note metodologiche sono sia di tipo statico sia di tipo dinamico (Fig.4). Le note di tipo statico possono essere consultate nella sezione *Note metodologiche* della banca dati che contiene informazioni di carattere generale sulle definizioni statistiche, nonché sulle modalità di rilevazione delle statistiche del commercio con l'estero. Ulteriori informazioni relative alle elaborazioni dei numeri indici e alle nuove statistiche per operatore ed impresa sono contenute nella stessa sezione.

Per quanto riguarda le classificazioni e le nomenclature, indicazioni dettagliate sono riportate nella sezione *Classificazioni* della banca dati che riporta informazioni sulle principali classificazioni e nomenclature di tipo geografico, merceologico e territoriale. In tale sezione è inoltre presente un motore di ricerca che consente di trovare i codici merceologici delle varie classificazioni in corrispondenza di una parola chiave indicata (per esempio: calzature).

Le note di tipo dinamico sono più di 5000 e, grazie ad una complessa gestione del modulo di “inquiring e reporting”, vengono attivate automaticamente in calce alla tavola (sempre fig.4) riportando informazioni puntuali relative ai parametri e alle modalità selezionate in funzione dell’intervallo temporale definito. Si pensi ad esempio a note che avvertono su cambiamenti di codici o descrizione relative a singole merci, province o paesi selezionati.

4. Modalità di accesso flessibile all’informazione statistica

Ulteriore elemento caratterizzante è la presenza simultanea di due diverse modalità di consultazione (“consultazione tematica” e “ricerca puntuale”) che attingono ad un comune patrimonio informativo ma che permettono all’utente di accedere alla banca dati in modo flessibile e finalizzato alle proprie esigenze conoscitive. La *ricerca puntuale* permette infatti all’utente di costruire, nell’ambito dell’area tematica considerata (dati nazionali, territoriali ecc), la propria tavola definendone struttura e contenuti. Nel caso della *consultazione tematica* si privilegia invece la rapidità di accesso ed il contenuto informativo a discapito della struttura della tavola, già definita, per cui è tuttavia possibile selezionare le modalità di interesse (singoli paesi, merci, regioni).

Per quanto riguarda le funzionalità previste dalla banca dati esse possono essere schematicamente raggruppate nelle seguenti categorie:

-Informative

-transazionali

All’interno delle funzionalità informative vi è la possibilità di usufruire di:

-note metodologiche estese sulle rilevazioni Intrastat ed Extrastat;

-riferimenti normativi;

-note metodologiche sulla elaborazione dei numeri indice del commercio estero ;

-note metodologiche sugli operatori ed imprese che effettuano interscambi commerciali;

-un glossario completo di tutti i termini propri della statistica doganale;

-tutte le classificazioni utilizzate all’interno del sistema;

-motore di ricerca per parole chiave all'interno di tutte le classificazioni merceologiche e derivate per consentire all'utente di trovare la corrispondenza codice-[merce cercata] che permetterà, anche ai meno esperti di classificazioni, di trovare le informazioni di loro interesse.

Nel dominio transazionale vi sono le funzionalità di registrazione e di modifica del profilo utente, ma soprattutto vi si trovano le funzionalità che costituiscono il nucleo centrale dell'applicazione, quelle cioè che permettono la visualizzazione, la stampa, il salvataggio su file di tavole statistiche costruite dall'utente.

Il meccanismo per la costruzione di tavole ad-hoc (in numero pressoché illimitato) è abbastanza semplice ed è composto da tre fasi:

- Scelta del layout della tavola e delle variabili interessate;
- Applicazioni di filtri sulle variabili interessate;
- Esecuzione dell'interrogazione e output della tabella richiesta.

Le tavole statistiche che possono essere prodotte rientrano nel piano della diffusione delle statistiche del commercio estero che disciplina le possibilità di compiere incroci e di raggiungere specifici livelli di dettaglio merceologico, geografico o territoriale. Tali tavole possono essere raggruppate nelle seguenti tipologie:

- Tabelle dei dati nazionali
- Tabelle dei dati territoriali
- Tabelle per modo di trasporto
- Tabelle per regime statistico
- Serie storiche
- Tabelle sui numeri indice
- Tabelle sugli operatori del commercio estero

5. Rispetto dei principi di riservatezza nella diffusione delle statistiche del commercio con l'estero

Un ulteriore aspetto saliente della banca dati è la sua completa compatibilità con le procedure di tutela della riservatezza nella diffusione delle statistiche del commercio con l'estero recentemente riviste e rese più efficaci dal Dipartimento delle statistiche economiche. La crescente domanda informativa sulle statistiche del commercio con l'estero deve infatti essere soddisfatta garantendo un'opportuna tutela ai soggetti economici interessati.

Tale obiettivo è piuttosto complesso poiché deve tener conto della particolare natura delle statistiche del commercio con l'estero nonché dei problemi di conflittualità, in termini di ulteriore

esposizione di dati personali, che possono essere indirettamente generati dalle nuove elaborazioni su operatori ed imprese.

Per un più approfondito esame di tali problematiche si rimanda al documento redatto dalla Dott.ssa Linda Cicalini. Di seguito si riportano le principali indicazioni operative definite nell'ambito della Dipartimento sulle statistiche economiche:

La diffusione delle statistiche relative ai flussi commerciali è autorizzata in funzione degli incroci e delle modalità previste dal Piano per la diffusione. Rispetto a tale piano risultano escluse le sole merci per cui è stato richiesto dalle imprese il vincolo di riservatezza passiva. Tutte le informazioni contenute nel piano per la diffusione sono presenti in banca dati.

Le elaborazioni relative alle merci non previste dal Piano di diffusione sono soggette al vincolo di riservatezza attiva attualmente vigente per le statistiche sulle imprese.

Le elaborazioni relative agli operatori e alle imprese che effettuano scambi commerciali con l'estero sono soggette al vincolo di riservatezza attiva. Nel caso degli operatori tale vincolo è innalzato ad almeno 9 operatori presenti per ogni incrocio al fine di evitare problemi di conflittualità con le tradizionali statistiche sulle merci.

Conclusioni

Il rilascio della banca dati COEWEB è un risultato che risponde a due degli obiettivi principali del piano d'azione sull'e-government: migliorare l'efficienza interna dell'amministrazione ed offrire a tutti l'accesso telematico alle informazioni ed ai servizi erogati.

L'elevata efficienza e flessibilità della nuova banca dati ha permesso di ridurre notevolmente i tempi di rilascio delle informazioni statistiche e di realizzare on line un'ampia gamma di interrogazioni che prima richiedevano la compilazione di appositi programmi. La definizione di nuove funzioni ed il completamento dell'offerta informativa sulla base di un piano per la diffusione organico e coerente ha notevolmente ampliato la mole di informazioni e profili di analisi possibili. Partendo dal presupposto che le statistiche del commercio con l'estero forniscono con cadenza mensile un ricco patrimonio informativo sui flussi commerciali dell'Italia utili a operatori sia economici sia istituzionali, il rilascio di un accesso internet per tale banca dati ha notevoli potenzialità di catturare l'interesse di una vasta gamma di utenti nazionali ed internazionali.