

Documenti ISTAT

Il progetto “banca dati della diffusione congiunturale”

F. Rizzo, L. Vignola, D. Camol, M. Bianchi (*)

ABSTRACT

1. ANALISI DI FATTIBILITÀ

- 1.1 OBIETTIVI DELLO STUDIO
- 1.2 LA SITUAZIONE ATTUALE
- 1.3 L'IDEA PROGETTUALE
- 1.4 RIUSO DI COMPONENTI ESISTENTI E MINIMIZZAZIONE DEL RISCHIO
- 1.5 PROGETTO DI MASSIMA DELLA SOLUZIONE
- 1.6 MODALITÀ E TEMPI DI REALIZZAZIONE

2. ITABLE: UNA SOLUZIONE COMPLETA AL PROBLEMA

- 2.1 ANALISI DEI REQUISITI
- 2.2 ANALISI DELLE CRITICITÀ
- 2.3 REALIZZAZIONE DI UN PROTOTIPO
- 2.4 L'ARCHITETTURA COMPLESSIVA
- 2.5 IL DISEGNO DEL DATABASE
 - 2.5.1 L'information model per i dati e metadati del progetto SDMX*
 - 2.5.2 Gerarchia dell'informazione nell'information model*
 - 2.5.3 L'applicazione dell'information model al progetto Bollettino on line*
- 2.6 IL MODULO IDESIGNER
 - 2.6.1 Strutture dati e rappresentazioni logiche*
- 2.7 PROCEDURE DI AGGIORNAMENTO DELLA BANCA DATI
 - 2.7.1 Il modulo Check*
 - 2.7.2 Il modulo Loader*
 - 2.7.3 Il modulo iSCAN*
- 2.8 IL MODULO WEBTABLE
 - 2.8.1 Autenticazione*
 - 2.8.2 Ricerca*
 - 2.8.3 Presentazione della tavola*
 - 2.8.4 Uso di XML ed XSL per la realizzazione delle tavole*
 - 2.8.5 La realizzazione di un oggetto capace di creare il file XML*
- 2.9 iCONVERTER

APPENDICE A - LE TECNOLOGIE XML NELLA PUBBLICAZIONE DEI DATI STATISTICI

- 1. INTRODUZIONE
- 2. XML IL METALINGUAGGIO UTILIZZATO PER DEFINIRE ALTRI LINGUAGGI
- 3. COMPOSIZIONE DI UN DOCUMENTO XML
- 4. PRESENTAZIONE DI DOCUMENTI XML TRAMITE CSS
- 5. PRESENTAZIONE DI DOCUMENTI XML TRAMITE XSL
- 6. LA FORMATTAZIONE DEL DOCUMENTO XML ATTRAVERSO XSL-FO

Abstract

Il progetto “banca dati della diffusione congiunturale” ha come scopo primario la pubblicazione su Web dei dati presenti nella pubblicazione cartacea del Bollettino Mensile di Statistica (BMS). Il problema si è presentato di non facile soluzione in quanto il BMS raccoglie dati provenienti da Direzioni di produzione differenti. Questo avrebbe sicuramente significato considerare realtà con sistemi informativi e organizzazioni del lavoro non omogenee. Da subito, però, si è intuito che questa problematicità si sarebbe potuta tradurre in una opportunità, cioè nella possibilità di reingegnerizzare, sistematizzandolo, l'intero processo di pubblicazione di tali dati.

Il progetto è stato, quindi, suddiviso in due fasi:

1. caricamento e aggiornamento dei dati in un data base e loro pubblicazione su Web;
2. pubblicazione della versione cartacea attraverso procedure automatiche che estraessero i dati direttamente dal data base senza richiederli ai Servizi di produzione.

Evidentemente un progetto così ambizioso coinvolgerebbe tutte le fasi di realizzazione dell'informazione: a partire da quando i dati sono disponibili presso i Servizi di produzione fino al momento in cui vengono pubblicati su Web. L'intero processo dovrebbe essere governato da procedure capaci sia di diminuire le possibilità di errori sia di garantire la tempestività di pubblicazione.

Per diminuire i fattori di rischio, il progetto è stato scomposto in blocchi ben circoscritti ed indipendenti fra di loro, anche se coordinati da un'unica regia.

Il documento è organizzato in due capitoli. Il primo capitolo espone l'analisi di fattibilità che ha fornito alla dirigenza gli elementi necessari per identificare i vantaggi ma anche i fattori di rischio. Il secondo capitolo affronta il progetto da un punto di vista tecnico. L'appendice fornisce, in forma semplice e senza la pretesa di completezza, una panoramica sulle tecnologie utilizzate.

I contributi degli autori al documento sono nel seguente ordine:

capitolo 1 – Francesco Rizzo

capitolo 2, paragrafi 2.1, 2.2, 2.3, 2.4, 2.5 – Francesco Rizzo

capitolo 2, paragrafo 2.6, 2.9 – Dario Camol

capitolo 2, paragrafi 2.7 – Laura Vignola

capitolo 2, paragrafo 2.8 – Francesco Rizzo

appendice A – Mauro Bianchi

1. Analisi di fattibilità

1.1 Obiettivi dello studio

L'obiettivo fondamentale di questo studio è quello di fornire l'insieme delle informazioni necessarie a decidere per l'effettivo avvio della realizzazione del progetto e, quindi, dell'investimento necessario.

Queste informazioni riguardano la fattibilità tecnica e organizzativa, i benefici, i costi, i rischi, le scadenze temporali.

Il documento si propone, dunque, di fornire tutti gli elementi per verificare e valutare sufficientemente l'idea progettuale per rendere, quindi, concreta la fase realizzativa.

Obiettivo finale della realizzazione del sistema informativo è di contribuire al miglioramento dei processi di servizio, per cui è necessario che il progetto si collochi in un contesto di razionalizzazione complessiva del processo di servizio che si va ad affrontare. In altre parole verrà assunto un punto di vista complessivo sul processo esplicitando gli obiettivi di miglioramento ed indicando le necessarie iniziative collaterali dell'intervento informatico.

1.2 La situazione attuale

Il Bollettino Mensile di Statistica (BMS) rappresenta, attualmente, una delle pubblicazioni strategiche dell'Istituto. Viene fornito su supporto cartaceo sotto forma di volume con cadenza mensile. Contiene principalmente dati congiunturali sotto forma di serie storiche a varie frequenze. Per alcune serie esiste anche la componente territoriale. A tale proposito esiste una sezione definita "dati territoriali" in cui il livello di disaggregazione dipende dall'indagine e generalmente è a livello regionale o provinciale (tranne i trasporti il cui livello di disaggregazione possono essere i porti, gli aeroporti, paesi esteri).

Vengono prodotti undici numeri mensili e un numero riassuntivo (agosto/settembre)

Il flusso informativo attuale per la produzione del BMS è il seguente:

- i servizi interessati al bollettino preparano le tabelle in formato Excel non inserendo le testate e il piè di pagina ma sole le fiancate e la parte centrale. Collocano tali file su una directory condivisa di un determinato computer;
- esiste una procedura centralizzata che collegandosi alle varie directory condivise integra le tabelle con le testate e i piè di pagina;
- tutte le tabelle vengono inviate alla tipografia che prepara una prima bozza;
- tale bozza viene visionata dai referenti dei servizi che la correggono da eventuali errori o danno il consenso di correttezza.

Attualmente esiste un ritardo di uscita sul mercato del BMS di circa 6 mesi rispetto ai dati che vengono proposti.

Il BMS cartaceo attualmente è organizzato in due raggruppamenti:

- dati nazionali – le informazioni sono organizzate lungo la componente temporale;
- dati territoriali – le informazioni sono organizzate lungo la componente territoriale.

L'organizzazione dei dati in forma tabellare dipende dall'indagine e, in ogni caso, all'interno della stessa indagine è possibile trovare i dati organizzati in forma differente.

L'organizzazione delle informazioni è stata pensata per un supporto cartaceo, per cui dovrà essere pesantemente rivista per arrivare alla necessaria ottimizzazione per la versione on-line.

1.3 L'idea progettuale

L'esigenza di migliorare le procedure informatiche per la realizzazione del BMS, adeguandole alle possibilità offerte attualmente dalle tecnologie informatiche e alle esigenze dell'utenza, è alla base dell'idea progettuale.

Tale progetto oltre a definire nuove procedure, tenderà a migliorare il workflow del flusso informativo attualmente esistente tra Servizi produttori e Direzione di diffusione in un'ottica di *reengineering* del processo affinché lo stesso possa raggiungere miglioramenti in termini di efficacia ed efficienza.

Il progetto potrà essere suddiviso in due fasi temporali ben distinte:

- a. versione on-line del BMS;
- b. successiva automatizzazione della produzione cartacea.

In generale il progetto dovrà prevedere l'analisi, la progettazione e la realizzazione dei seguenti punti:

- protocollo di trasmissione dati tra Servizi produttore e Direzione di diffusione;
- definizione dei sistemi di codifica più adeguati;
- memorizzazione dei dati e metadati in un data base centralizzato;
- procedura di check per l'inserimento dei dati nel DBMS e procedure di RollBack per riportare lo stesso in una situazione di consistenza nel caso di inserimento di dati non corretti;
- procedure batch di controllo periodico della consistenza della banca dati;
- procedure per l'organizzazione delle informazioni in formato tabellare simile a quello cartaceo;

- presentazione dei dati in ambiente Web;
- eventuali procedure aggiuntive quali invio tramite e-mail a utenti particolari, ecc.;
- procedure batch di reporting per la realizzazione del BMS cartaceo.

1.4 Riutilizzo di componenti esistenti e minimizzazione del rischio

La realizzazione di qualunque sistema informativo, al di là della complessità realizzativa, dovrà tener conto dell'esistente. Ciò permetterà di non duplicare le informazioni, ma anche il flusso informativo, le esigenze strumentali (HW e SW), ed in ultima analisi l'impiego del personale.

Una analisi preliminare circa i dati contenuti attualmente nel BMS ha portato alla conclusione che:

- esiste già, almeno per quanto riguarda la componente economica delle informazioni, un flusso di dati codificato molto simile a quello che il progetto esprime tra i referenti dei dati dei vari Servizi produttori e una struttura centrale di gestione della banca dati Conistat¹;
- molte delle informazioni presenti nel BMS sono già presenti nel database da cui attinge Conistat.

Inoltre esistono già in corso, all'interno dell'Istituto, progetti, in fasi più o meno avanzate:

- tendenti a canalizzare le informazioni di diffusione in un unico contenitore da cui prelevare alla necessità;
- tendenti a standardizzare la codifica, e in generale la metainformazione che si accompagna al dato.

L'idea potrebbe essere quella di:

a) adattare il canale di alimentazione di Conistat in maniera tale che le modalità di alimentazione del BMS sia unificato con quello di Conistat. Questo permetterebbe una diminuzione del carico di lavoro presso i Servizi e una riduzione dei rischi di errori (gli stessi dati potrebbero essere passati con due valori diversi).

In particolare si potrebbe suddividere i Servizi interessati tra:

1. quelli che già forniscono i dati per la banca dati Conistat e quindi hanno già delle procedure codificate per fare ciò;
2. quelli che non sono presenti in tale banca dati e, quindi, sarà necessario realizzare le adeguate procedure.

¹ Conistat è la banca dati delle serie storiche congiunturali, accessibile all'indirizzo: <http://con.istat.it>

Nel primo caso l'operazione da compiere sarà quella di integrare il flusso dei dati con serie non presenti in Conistat ma presenti nel BMS; nel secondo caso oltre a realizzare le necessarie procedure informatiche, si potrà pensare ad un adeguamento del sistema di codifica (nel caso all'interno dell'Istituto esista un sistema di codifica centralizzato, si potrà pensare di adottare questo);

b) concordare con i progetti esistenti tempi e modi per:

1. la realizzazione di un contenitore comune e la sua gestione;
2. un sistema di codifica e di metainformazione.

Questo tipo di organizzazione del progetto porterà certamente ad un'ottimizzazione di risorse per la realizzazione dello stesso, in quanto parti del progetto potranno essere scorporate delegandole ad altre strutture o gruppi di lavoro.

1.5 Progetto di massima della soluzione

A fronte di una progettazione informatica dell'intero sistema sarà, inoltre, necessario pensare ad una progettazione dell'impaginazione grafica, che risulti adeguata ad una presentazione on-line del BMS.

L'intero workflow del sistema avrà come punto centrale di coordinamento delle varie operazioni un calendario degli eventi. Ciascun soggetto interessato compierà varie operazioni a secondo delle scadenze preventivate sul calendario. Tale calendario servirà anche per attivare i solleciti nel caso alcune operazioni subiranno ritardi rispetto alle date previste.

Il workflow può essere così schematizzato:

1. alla propria scadenza mensile ciascun referente dei dati (RD) presso il Servizio, attraverso un opportuno programma, carica la banca dati posizionata sulla intranet e produce un file che verrà inviato alla persona, addetta al caricamento (AC) della banca dati su internet, che procede nel passo successivo;
2. AC provvede al caricamento della banca dati su Internet, attraverso un opportuno programma (Loader);
3. ad un'ora prefissata sarà possibile pensare ad un processo *unattended* che verifica il rispetto delle scadenze così come sono state definite nel calendario. Nel caso tali scadenze non vengano rispettate, invia un messaggio di sollecito alla persona interessata.

Uno schema di massima delle operazioni da compiere nella progettazione e realizzazione dell'intero sistema è il seguente:

- analisi presso i servizi per capire le esigenze particolari ed avviare un protocollo per l'alimentazione strutturata della banca dati del BMS;
- progettazione e realizzazione della struttura del database;
- progettazione e realizzazione di adeguate procedure di amministrazione sia da utilizzare centralmente (gruppo di gestione centrale), sia dai singoli servizi produttori dei dati;
- realizzazione di procedure di controllo delle informazioni e della consistenza del DBMS;
- realizzazione delle procedure per la produzione delle informazioni via Web;
- realizzazione delle procedure per la produzione delle informazioni cartacee.

1.6 Modalità e tempi di realizzazione

Il progetto dovrà prevedere un gruppo di lavoro che si occupi di:

- progettazione informatica del sistema;
- progettazione editoriale del sistema;
- progettazione statistica del sistema;
- realizzazione del sistema:
 - realizzazione del DBMS;
 - realizzazione degli applicativi;
 - test del sistema;
 - messa in produzione del sistema.

Il gruppo potrà essere composto secondo le seguenti professionalità:

- 1 coordinatore con funzione anche di analista;
- 1 esperto di progettazione editoriale;
- 1 figura statistica di coordinamento nel settore economico;
- 1 figura statistica di coordinamento nel settore sociale;
- 2 programmatori;
- 1 referente dei dati presso ciascun Servizio di produzione interessato.

Per tutte le figure, tranne per i programmatori per i quali è previsto un impegno a tempo pieno per la durata dell'intero progetto, dovrà essere previsto un impegno almeno al 20% del loro tempo lavorativo nella durata indicata nella seguente tabella:

Funzione	Durata	Figure professionali
Analisi di fattibilità e progettazione	1 mese	Analista 100% P. editoriale 100% Statistici 100% Referenti 100%
Progettazione del DBMS e caricamento	1 mese	Analista 30% Programmatori 100%
Realizzazione degli applicativi	4 mesi	Analista 10% Programmatori 100%
Test e messa in produzione	1 mese	Analista 10% Programmatori 100%

E' possibile stimare, escludendo la parte del progetto che fa riferimento all'automatizzazione della produzione cartacea che dovrà essere valutata in seguito, un periodo di realizzazione di circa 6 mesi dall'avvio del progetto.

2. iTable: una soluzione completa al problema

2.1 Analisi dei requisiti

La necessità di realizzare un sistema integrato a supporto delle attività finalizzate alla pubblicazione dei dati congiunturali, sotto forma di tavole, è stato il punto di avvio per la fase di progettazione. Questo comporterà:

- un miglioramento dell'efficienza nella raccolta, validazione, memorizzazione e diffusione dei dati;
- un miglioramento della qualità attraverso la riduzione della possibilità di errore;
- un aumento dell'accessibilità e visibilità dell'informazione congiunturale.

Per quanto riguarda i requisiti da tenere in considerazione per la parte del sistema più vicina alla diffusione abbiamo considerato i seguenti vincoli:

- il layout delle tavole deve rispettare quello presente sulla pubblicazione cartacea dopo una loro razionalizzazione;
- una tavola può svilupparsi sia su una struttura organizzata su una sola pagina che su più pagine;
- la ricerca di una tavola deve avvenire principalmente attraverso un sistema che simuli la ricerca attraverso l'indice della pubblicazione cartacea, consentendo anche forme alternative di ricerca;
- la criticità dei tempi di risposta.

Per arrivare allo scopo il progetto ha considerato i vantaggi che le innovazioni informatiche allo stato attuale offrono: strumenti di sviluppo Web evoluti, XML, Web services, ecc..

2.2 Analisi delle criticità

In passato si è già cercato di realizzare un sistema per l'automatizzazione del BMS, ma i tentativi sono falliti per varie cause tra cui:

- organizzazione del lavoro disomogenea tra i vari Servizi fornitori dei dati;
- uso di software diversi per la fornitura dei dati da parte dei Servizi che utilizzano procedure automatizzate;
- disomogeneità di procedure di realizzazione dei dati in formato Excel, in quanto alcuni Servizi utilizzano procedure che producono direttamente questo tipo di file, altri li realizzano manualmente.

Da un'analisi effettuata presso i Servizi è risultato:

% tabelle	tipo di produzione
36 %	Automatizzata
64 %	Manuale

La soglia di accettabilità del rischio è stata definita come la possibilità di inserire nella versione on-line del BMS, almeno nella fase di partenza, tutte quelle tavole che possono essere composte con dati provenienti dal database Conistat. Questo comporterebbe la disponibilità immediata di oltre il 50% delle tavole.

Si dovrà pensare quindi ad un periodo di transizione in cui:

- alcuni servizi continueranno a fornire i dati solo al database Conistat. In questo caso sarà necessario realizzare le adeguate procedure affinché da tale database vengano estratti i dati in un formato Excel compatibile con quello che l'attuale procedura per la realizzazione del BMS accetta;
- altri servizi, non avendo disponibilità immediata a fornire i dati nel formato verticalizzato codice-periodo-valore così come accettato da Conistat, avranno bisogno di una procedura che vada a leggere dai loro file Excel e trasformi i dati nel formato verticalizzato.

2.3 Realizzazione di un prototipo

Per l'implementazione di un sistema così complesso, si è reso, necessario, la realizzazione di un prototipo che permettesse di fare le scelte più opportune durante la progettazione verificando:

- gli strumenti software da utilizzare;
- gli algoritmi più efficienti;
- i tempi di risposta del sistema alla richiesta di un utente;
- il disegno di un database capace di soddisfare oltre che le esigenze tipicamente di diffusione, anche le esigenze di memorizzazione con particolare riguardo alla frequenza di aggiornamento (dai 15 ai 20 aggiornamenti mensili) e alla possibilità di ripristino di situazioni consistenti dopo un aggiornamento con dati errati.

Durante la realizzazione del prototipo sono state considerate esperienze realizzate in Istat e/o l'utilizzo di software commerciali quali Business Object e Sas.

Non sono state riscontrate in Istat esperienze già realizzate che ci potessero aiutare nel progetto con particolare riguardo all'automazione dei processi di produzione di tavole statistiche complesse. Inoltre, l'uso dei software commerciali non ha prodotto i risultati attesi, che sarebbe stato possibile ottenere solo a fronte di una complessa programmazione. Si è resa, quindi, necessaria la realizzazione di una soluzione ad-hoc. Tale soluzione, che in primo momento potrebbe sembrare non economica, a regime rappresenterà la soluzione più efficiente in quanto "aperta" e, quindi, utilizzabile e modificabile nel tempo. Inoltre una programmazione modulare garantirebbe nel tempo interventi evolutivi senza dover riprogettare parti già consolidate.

Per quanto riguarda la tecnologia che abbiamo utilizzato nella realizzazione del prototipo abbiamo tenuto conto dei seguenti fattori per dar luogo ad una griglia di valutazione:

- facilità d'uso e tempi di sviluppo;
- stabilità;
- scalabilità;
- facilitazione nell'uso dell'XML;
- performance;
- facilità di gestione e manutenzione.

Si è arrivati alla conclusione di utilizzare la tecnologia .NET della Microsoft. In particolare questo modulo software utilizza ASP.NET 1.1.

L'uso di tale tecnologia ha avuto i seguenti vantaggi immediati:

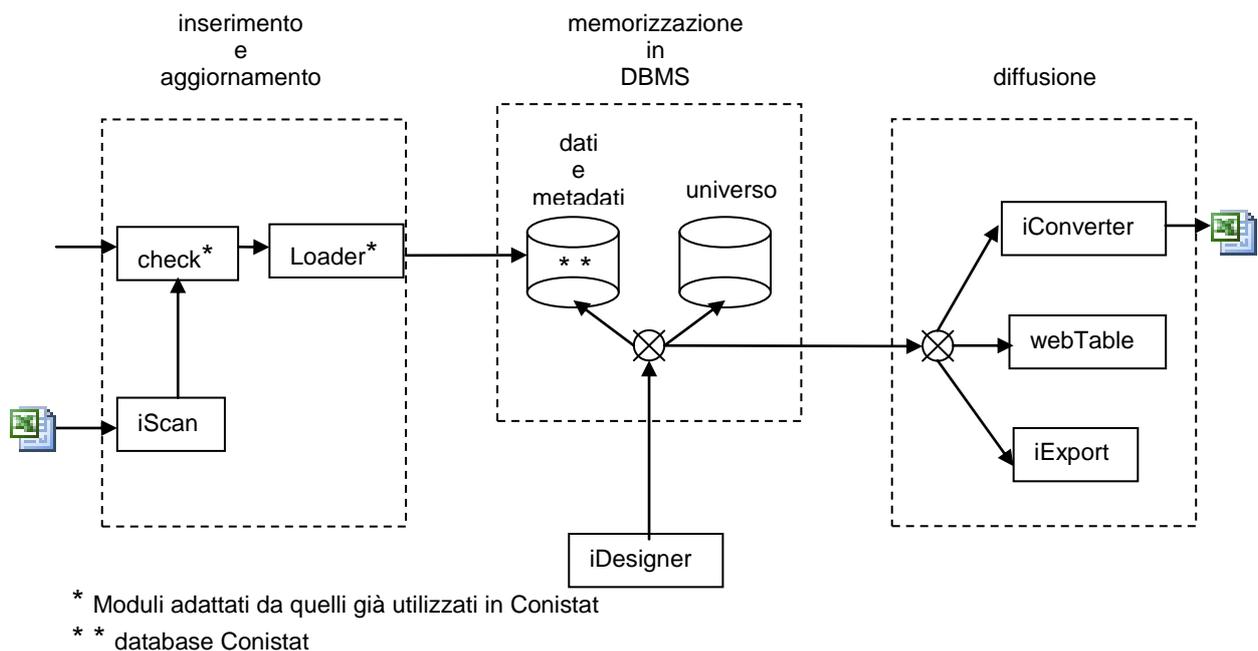
- le persone che avrebbero dovuto lavorare al progetto conoscevano già il linguaggio Visual Basic in ambiente Windows e Visual Basic Script in ambiente Web, quindi sono state facilitate nel migrare nel nuovo ambiente;
- ASP.NET permette di utilizzare un paradigma di programmazione simile a quello utilizzato in ambiente Windows: uso di oggetti forniti con il compilatore o da terze parti ai quali associare a design-time e/o a run-time proprietà e metodi. In particolare con oggetti complessi quali griglie è possibile utilizzare funzionalità di databind² ;
- l'oggetto dataset di ASP.NET permette di continuare a lavorare in ambiente relazionale, più vicino come logica alla programmazione classica, lasciando ad una funzionalità di tale oggetto la traduzione in formato XML;
- l'uso del debugger in ambiente Web ha permesso di ridurre notevolmente i tempi di sviluppo e, nello stesso tempo, rendere le procedure il più possibile prive di errori;

² Per databind si intende quella operazione mediante la quale vengono associati ad un oggetto alcuni record estratti da un database. Quindi l'oggetto stesso, a run-time, penserà a formattare tali dati in stringhe HTML

- l'uso di strumenti di tracing per evidenziare parti di codici che danno luogo ai classici "colli di bottiglia".

2.4 L'architettura complessiva

L'idea progettuale ha portato alla realizzazione di una soluzione integrata che permettesse di sistematizzare l'intero flusso informativo che dai Servizi di produzione arrivasse fino alla pubblicazione su Web e alla pubblicazione cartacea. Tale soluzione è stata denominata iTable:



Segue una breve descrizione dei vari moduli:

- *iDesigner* - permette di disegnare la struttura di ciascuna tavola, memorizzandola in un DBMS (universo);
- *iScan* - permette di acquisire, da tabelle Excel, sia la metainformazione descrittiva di ciascuna serie storica, sia i dati nel caso in cui il Servizio interessato non abbia immediata disponibilità a fornire gli stessi in un formato verticalizzato standardizzato;
- *Check & Loader* - permettono partendo da un file verticalizzato di costruire un file controllato (*check*) e di caricare nel data base il file stesso (*Loader*). Tali programmi sono un adattamento, alle nuove esigenze, della versione che attualmente viene utilizzata dai referenti dei dati che aggiornano la banca dati Conistat;
- *iConverter* - permette di esportare, in modalità batch, le tabelle nel formato Excel attualmente accettabile dalle correnti procedure per produrre il Bollettino Mensile di

Statistica cartaceo. Inoltre permette di esportare le tabelle in formato pdf pronto per essere inviato in tipografia per la stampa del Bollettino cartaceo;

- webTable: è l'interfaccia Web attraverso la quale un utente esterno accede al sistema per ricercare le tavole d'interesse.

2.5 Il disegno del database

Nel disegnare il database è stata presa in considerazione la possibilità di riutilizzare strutture già esistenti nel caso in cui risultassero compatibili con il progetto. A tale proposito è stato analizzato il database della banca dati Conistat. Tale database è nato per memorizzare e gestire in maniera ottimizzata dati organizzati in serie storiche. Dalle prove fatte attraverso il prototipo, è risultato che la struttura risultava compatibile con gli obiettivi del progetto a fronte di integrazioni al disegno iniziale. Inoltre l'uso di tale database ha evidenziato i seguenti vantaggi:

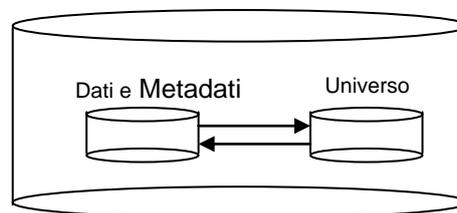
- utilizzo di un deposito dati che contiene già una buona percentuale delle informazioni necessarie al progetto;
- utilizzo di un deposito pensato per aggiornamenti ad alta frequenza che seguono un flusso informativo standardizzato;
- utilizzo di un deposito già ottimizzato nella struttura per memorizzare serie storiche.

Si è quindi progettata una infrastruttura di database organizzata in due aree logiche ben distinguibili:

- dati e metadati, mutuabile dal database Conistat;
- caratteristiche delle tavole, memorizzate in una nuova struttura denominata "Universo".

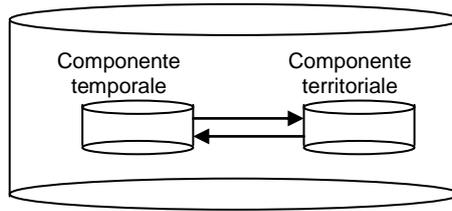
Gli schemi rappresentativi dei concetti sopra esposti sono i seguenti:

Database dell'informazione congiunturale

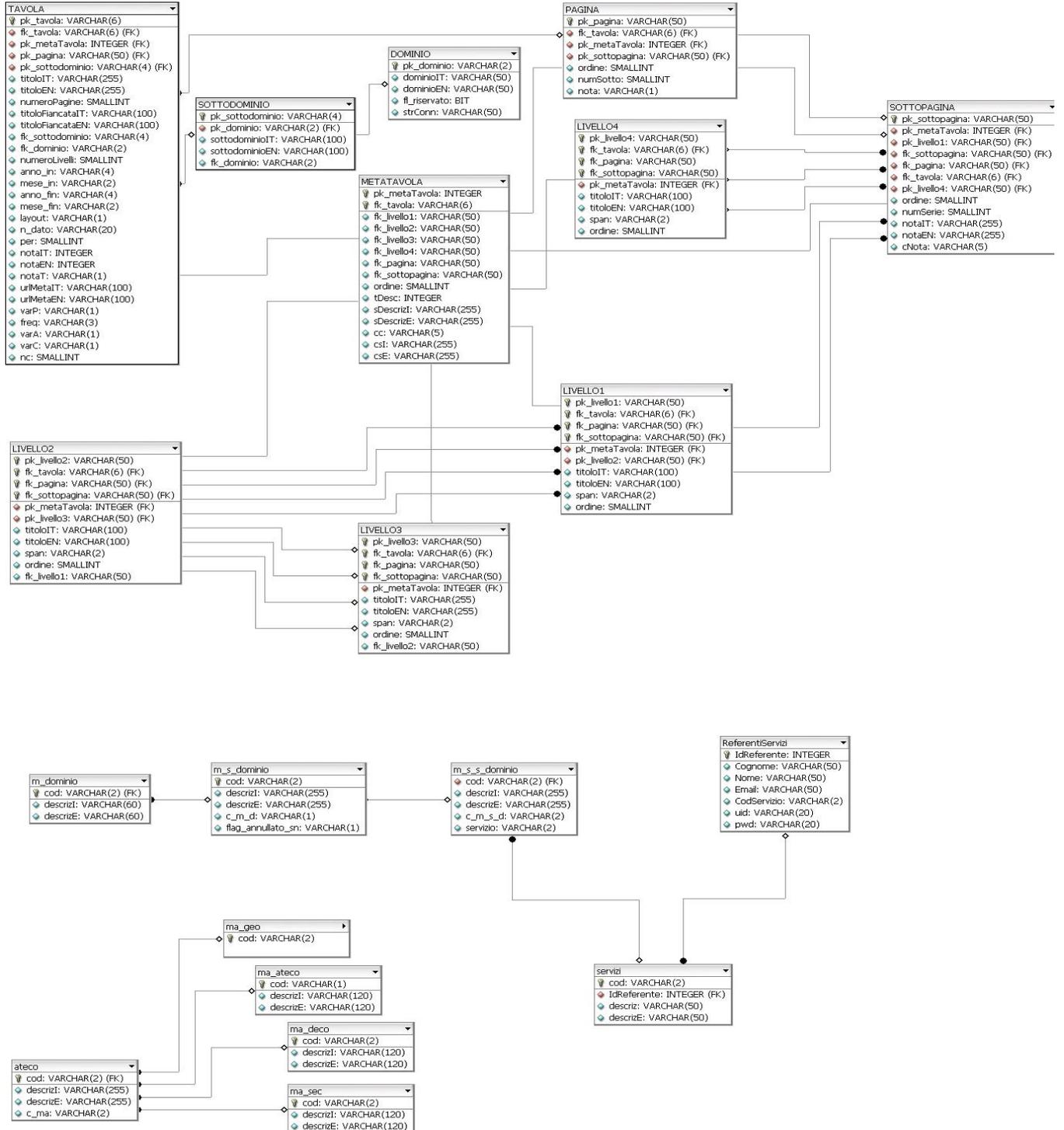


Per quanto riguarda la parte del database riguardante i dati e metadati, si è reso necessario fare una integrazione che tenesse conto dei dati a carattere territoriale:

Dati e metadati



Segue lo schema del data base nel suo complesso:



2.5.1 L'information model per i dati e metadati del progetto SDMX

Il progetto dell'information model dei dati e metadati ha cercato di garantire una compatibilità con quello proposto dal gruppo di lavoro che a livello internazionale ha dato luogo allo standard SDMX³.

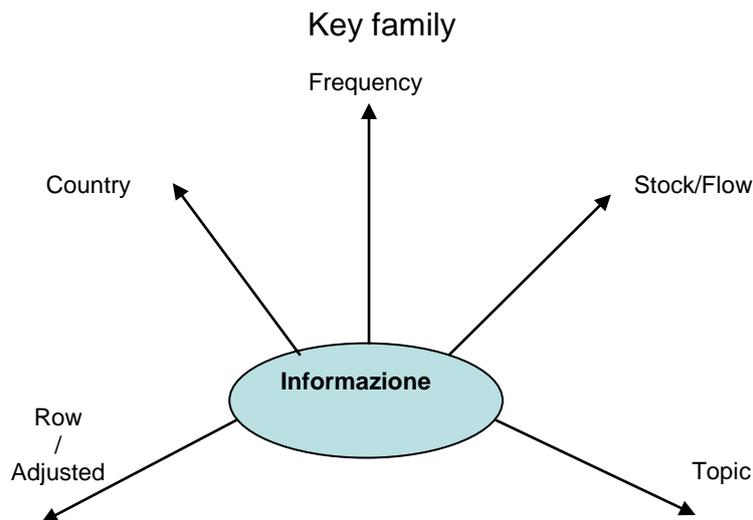
Di seguito vengono riassunte le principali caratteristiche di tale modello e vengono forniti alcuni concetti e definizioni su cui tale modello si basa:

- metadati: informazioni necessarie che permettono di interpretare correttamente i dati;
- osservazione: è il valore di una variabile misurata in un preciso momento (economic phenomenon);
- periodo: riferimento temporale in cui viene effettuata la misurazione di un fenomeno;
- frequenza: descrive l'intervallo di tempo che intercorre tra due osservazioni;
- descrittori o concetti statistici: i concetti fondamentali nel descrivere le osservazioni;
- dimensione: concetto statistico il cui contributo permette di individuare univocamente un dato o insieme di dati;
- key family: una o più dimensioni che possono essere immaginate come gli assi che descrivono le informazioni poste al loro centro;
- attributi: insiemi di concetti statistici usati per qualificare le osservazioni per una data key family;
- code list: insiemi di valori ammessi per i concetti statistici.

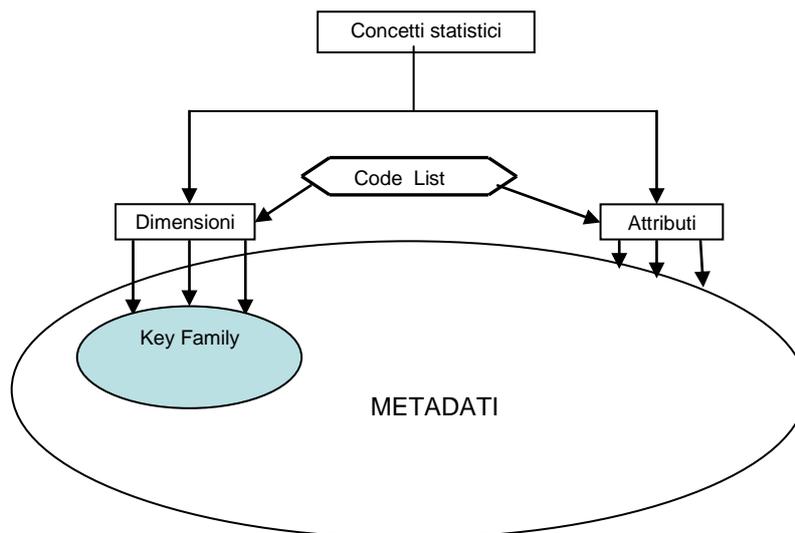
Data l'importanza della metainformazione nell'information model, la sua definizione viene meglio specificata:

- metadati strutturali: descrivono come i data set statistici sono organizzati. In generale svolgono una funzione centrale nel momento in cui i Sistemi Informativi Statistici vengono organizzati centralmente mediante un repository. In generale le key family forniscono il modello dell'organizzazione dei dati e, quindi, rappresentano i metadati strutturali;
- metadati metodologici: descrivono come i numeri vengono raccolti, analizzati e diffusi a seconda del dominio di appartenenza. Un esempio può essere considerato lo standard proposto dal FMI e poi fatto proprio da diverse organizzazioni centrali e nazionali.

³ SDMX – Statistical Data and Metadata eXchange è una iniziativa, sponsorizzata da vari organismi internazionali (Eurostat, BIS, ECB, IMF, OECD, UNSD) che ha il compito di sviluppare uno standard efficiente per lo scambio e la condivisione di dati e metadati.



Struttura logica dell'information model del database dei dati e metadati



2.5.2 Gerarchia dell'informazione nell'information model

I dati statistici sono spesso raggruppati per fornire un'informazione d'insieme. E' possibile, partendo dalla singola osservazione, tipicamente referenziata dal periodo in cui viene osservata, ma anche qualificata attraverso attributi sul suo stato (valore normale, break point, definitivo, provvisorio, ecc), considerare un insieme di osservazioni nel tempo. Tale raggruppamento, chiamato Time Series (TS) permette di vedere il trend del fenomeno che viene preso in considerazione. Una TS ha sempre associato il concetto di frequenza che

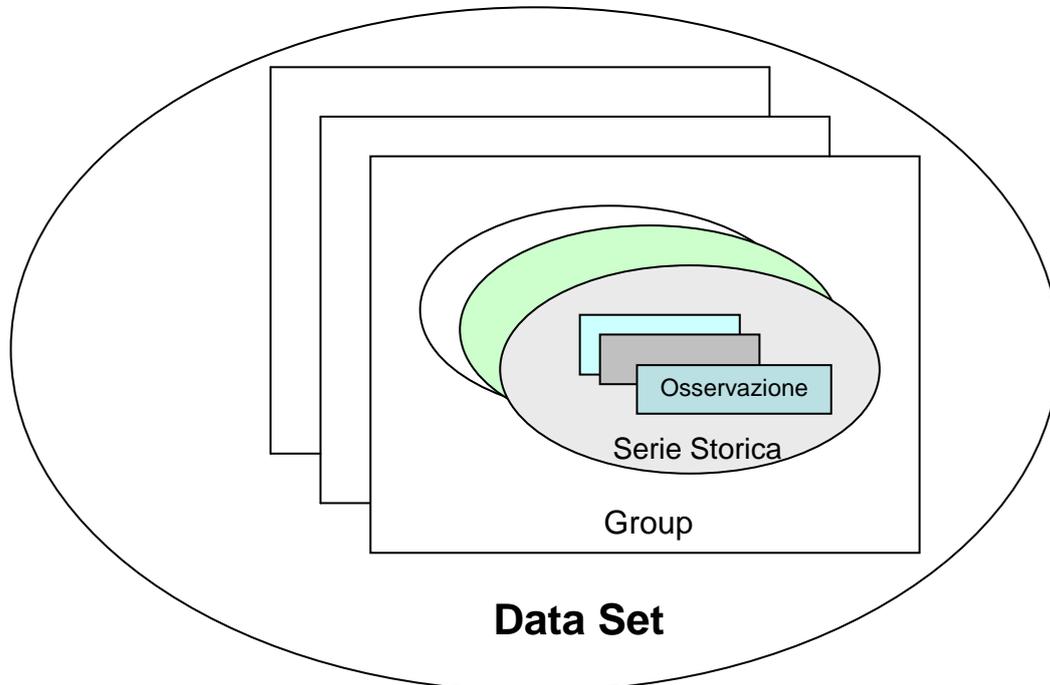
descrive ogni quanto tempo un fenomeno viene osservato. In generale la frequenza può essere costante o variabile.

Mentre il tempo è un concetto statistico di tipo dimensione sempre associato alla singola osservazione, la frequenza è un concetto statistico di tipo dimensione sempre associato ad una serie storica.

Alcune volte è necessario fornire un'informazione per la cui completezza vanno osservati diversi fenomeni. In tale caso un certo numero di serie storiche vengono raggruppate assieme formando un Group. Esiste un gruppo particolare definito Sibling Group che contiene un insieme di serie storiche identiche, tranne per una dimensione, come ad esempio la frequenza. In questo caso si dice che la dimensione non uguale è stata "wildcarded".

Infine esiste un raggruppamento chiamato Data Set che può includere uno o più Group. In genere un Data Set può rappresentare i dati che un Istituto nazionale di statistica scambia con un'istituzione centrale o, all'interno di una singola istituzione potrebbe rappresentare i dati di una singola struttura di produzione.

E' possibile quindi identificare la seguente gerarchia:



E' possibile inoltre considerare l'organizzazione dei dati in Cross-sectional, dove tale modalità permette di presentare un grande numero di osservazioni che vengono presentate in un preciso riferimento temporale. Volendo visualizzare tale concetto, si

potrebbe pensare ad una tabella in cui vengono presentati dei dati con la dimensione territoriale esplicitata a livello sub nazionale in un particolare anno.

Come definito precedentemente una key family permette di associare un insieme di descrittori (dimensioni) ad uno specifico set di dati statistici. Alla luce delle modalità di raggruppamento sopra esposte si può dire che una key family fornisce informazioni circa il livello al quale un particolare descrittore viene collegato.

Guardando alla gerarchia dei gruppi, in un determinato gruppo alcune dimensioni possono avere il proprio valore uguale in tutte le serie contenute nel gruppo, mentre altre dimensioni possono variare per ciascuna serie.

Per esempio consideriamo un Sibling Group rappresentato dall'Indice della Produzione industriale mensile in Italia per codice di Attività Economica ad una cifra. Le dimensioni rappresentanti i concetti di Soggetto ("dall'Indice della Produzione industriale"), di Territorio ("Italia") e Frequenza ("mensile") sono uguali per ciascun membro del gruppo. Il concetto descrittore "codice di Attività Economica ad una cifra" cambia per ciascuna serie dentro il gruppo: ci sarà una serie per "C – Estrazione di minerali", una per "D – Attività manifatturiera", una per "E – produzione e distribuzione di energia elettrica, gas e acqua", ecc.

In generale la regola è: le dimensioni vengono collegate al livello di gruppo dove esse diventano variabili. Così, se all'interno di un insieme di dati raggruppati in un livello le serie hanno una dimensione uguale, tale dimensione va collegata a quel livello.

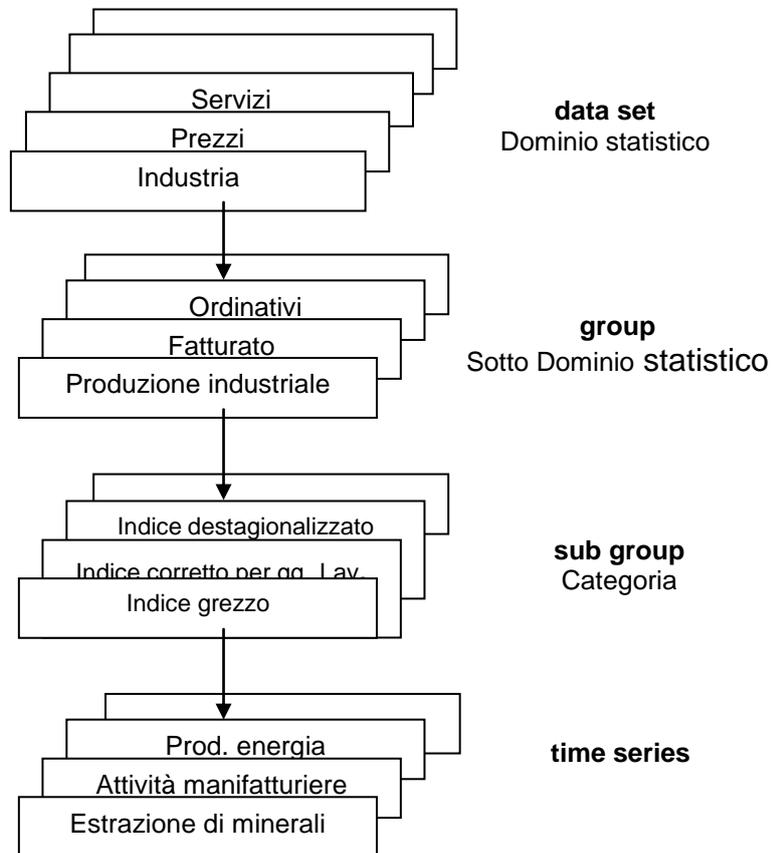
Nel nostro esempio essendo i descrittori Soggetto, Frequenza e Territorio uguali per tutti le serie, essi andranno a far parte della key family che individuerà il gruppo; mentre il codice di attività economica, variando per ciascuna serie, farà parte della key family che specificherà la singola serie.

Generalizzando si potrebbe dire che un determinato livello di raggruppamento è il più alto livello strutturale dove i suoi sottogruppi condividono lo stesso valore.

Nel caso in cui volessimo considerare l'intero Data Set dei dati dell'Industria Italiana comprendente i Sibling Group del Fatturato, degli Ordinativi e della Produzione, la dimensione Territorio ("Italia") essendo condivisa da ciascun Sibling Group, verrà collegata a livello di Data Set.

2.5.3 L'applicazione dell'information model al progetto Bollettino on line

Lo schema seguente individua la struttura gerarchica conseguente all'applicazione dell'information model:



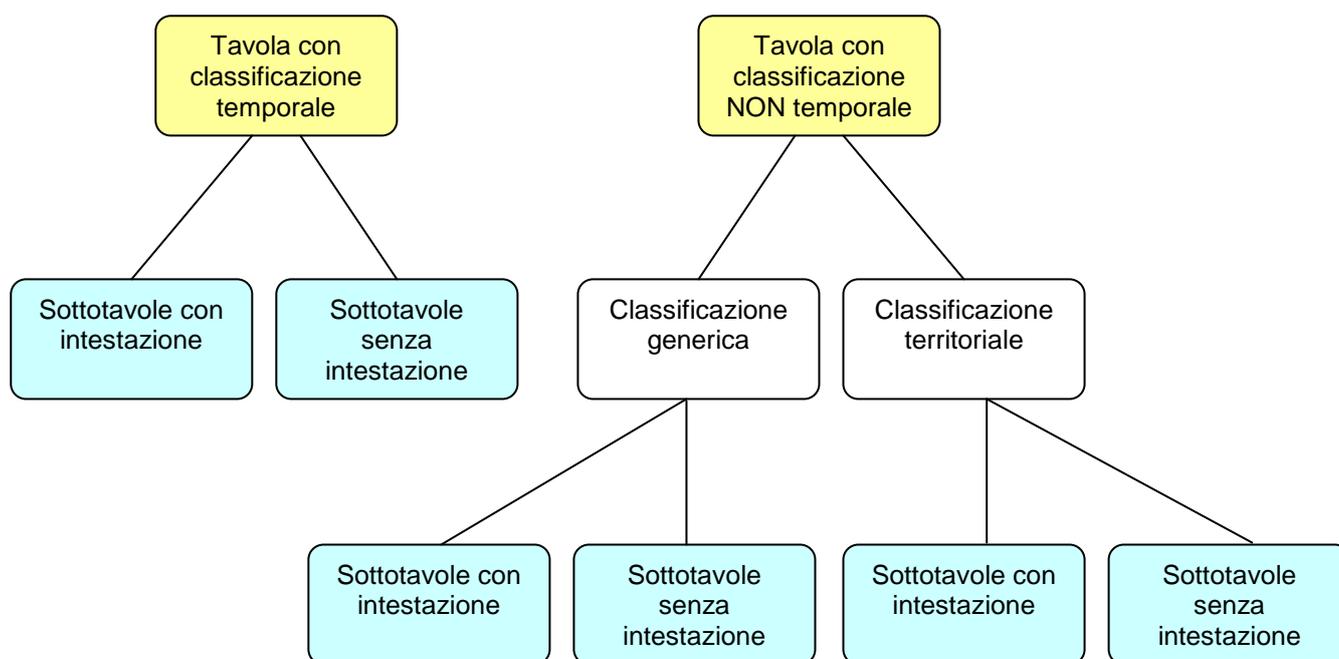
2.6 Il modulo iDesigner

Il modulo iDesigner permette di descrivere la struttura delle tavole memorizzandole in un database definito *Universo*. Sviluppato in Microsoft Visual Basic 6.0, iDesigner è un'applicazione client-server perfettamente integrata nel suo ambiente operativo e nell'insieme dei moduli del sistema iTable.

Con interfaccia utente semplice ed intuitiva evita al personale statistico la necessità di lunghi periodi di training.

L'aver creato un sistema di così immediato utilizzo è stato possibile grazie ad un'attenta analisi sull'eterogeneità delle strutture riscontrabili fra le tavole dei servizi.

Ciò ha permesso di estrapolare il seguente insieme minimo di modelli base e, quindi, di semplificare un software che altrimenti avrebbe avuto un notevole incremento di complessità dal punto di vista dell'interazione.



Questo modulo permette al progettista della tavola di descriverne una partendo dalla selezione delle variabili che la costituiranno. Tale selezione avviene attraverso una ricerca ad albero nel database che rappresenta il contenitore delle serie storiche congiunturali. Quindi sarà necessario definire la testata attraverso l'inserimento delle opportune descrizioni e, nel caso di tavole Cross-sectional, definire la formattazione delle classificazioni di fianco.

L'interfaccia è stata sviluppata con una particolare attenzione all'approccio WYSIWYG (What You See Is What You Get), che permette al progettista della tavola di seguire un

percorso privo di ambiguità. Tale peculiarità è garantita, inoltre, dall'aver progettato tutto il processo produttivo della descrizione in maniera sequenziale: step by step. Si è prevenuto il possibile divenire modale, tipico dei wizard, dotando l'interfaccia di un tasto per l'annullamento di qualsiasi operazione in corso e dando la possibilità, nella maggioranza degli step, di ripercorrere i passi precedenti.

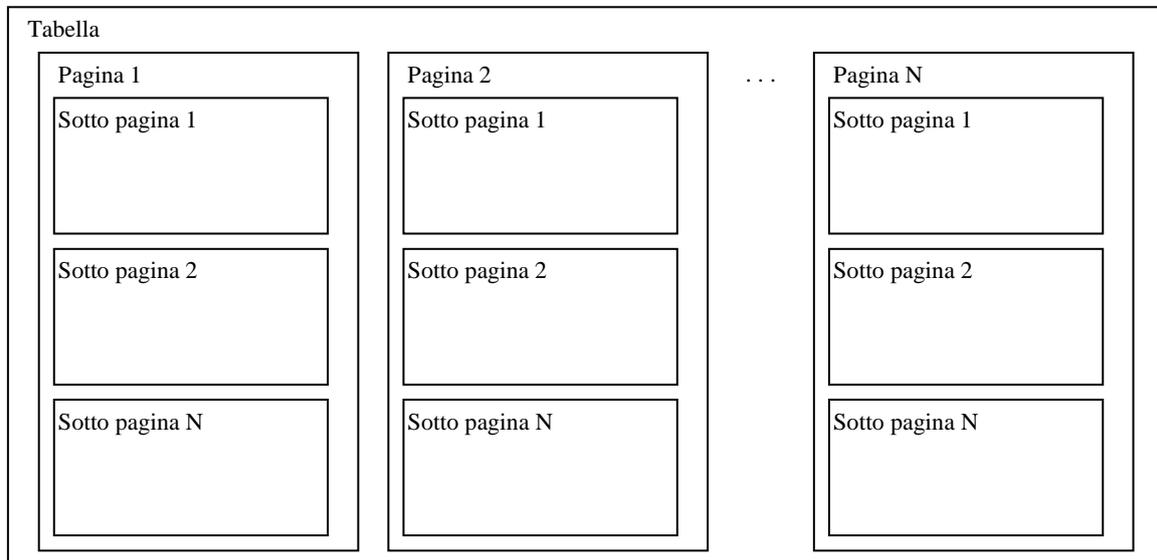
Sono state previste, inoltre, le seguenti funzionalità:

- la **gestione tipologia utente**, con opportuno login, per limitare il dominio delle informazioni proponibile al progettista della tavola semplificando, così, le sue scelte;
- il raggruppamento in **sottoform di informazioni omogenee**, che consente al progettista di affrontare l'inserimento in maniera più specifica e modulare;
- la presenza, nella maggioranza dei casi, nei sottoform di un tasto "**Salva ed esci**" che, favorendo le modifiche limitate e fortemente mirate, aumenta la confidenza con il sistema facendo sentire il progettista libero di sbagliare ed in maniera veloce poter correggere i propri errori;
- l'implementazione di una form apposita per la **ricerca di stringhe, in italiano ed inglese**, che permette una rapida compilazione dei campi che presentano testo simile. Tale form permette, inoltre, una facile traduzione, nel caso di stringa già esistente, fra le due lingue del sistema;
- la possibilità di inserire **oggetti in maniera ordinata** garantisce la sospensione della sessione di lavoro e permette di ricominciare, eventualmente, aggiungendo elementi precedenti a quelli già inseriti;
- la particolare attenzione data all'implementazione della **form dedicata all'inserimento delle modalità di classificazione**, dotando l'utente di tutti gli strumenti possibili per rendere l'operazione esaustiva e nel contempo rapida, cercando di limitare gli errori nel momento più impegnativo nell'utilizzo del iDesigner;
- la visualizzazione di **un messaggio di aiuto** ad ogni passo raggiunto, oltre all'elenco di tutti gli step ancora da dover compiere per poter completare una descrizione totale o parziale della tavola, rende più comprensibile all'utente ogni sua azione.

Sulla base di confronti diretti con gli utenti progettisti di tavole, l'aggiornamento del sistema di interfaccia sta permettendo un continuo apporto di migliorie all'utilizzo del iDesigner.

2.6.1 Strutture dati e rappresentazioni logiche

L'astrazione dei modelli di tavola ha portato alla generazione di una serie di oggetti che vengono utilizzati nell'engine del software. Per iDesigner una tavola è rappresentata logicamente come nel seguente schema:



1. una tavola può essere costituita da una o più pagine;
2. ciascuna pagina può essere costituita da una o più sottopagina;
3. ciascuna sottopagina contiene un certo numero di serie storiche o di dati Cross-sectional.

Inoltre ciascuna tavola presenta un determinato layout. La versione 1.0 del sistema gestisce 2 layout, entrambi orientati alle serie storiche. Nella versione 2.0 verrà gestito un layout orientato ai dati Cross-sectional.

Seguono gli schemi dei layout 1 e 2:

Sub page 2

Sub page 1

LAYOUT 1

Time series

LEVEL 3

LEVEL 2

LEVEL 1

	Estrazioni di Minerali										Attività manifatturiere		
	Indice generale	Totale	Estrazione di minerali energetici		Estrazione di minerali non energetici						Industrie alimentari e del tabacco		Industrie a bevande
			Estrazione di petrolio greggio e gas naturale	Estrazione di minerali non energetici	Totale	Estrazione di pietra	Estrazione di ghiaia, sabbia e argilla	Estrazione di minerali per le industrie chimiche e la fabbricazione di concimi	Produzione di sale	Estrazione di altri minerali e prodotti di cava	Totale	Totale	
2002	97,8	107,9	117,1	85,1	92,5	72,0	78,6	100,3	110,8	97,2	105,0	105,4	
2003	96,8	110,0	119,7	86,0	96,1	74,1	38,9	87,7	104,2	95,5	107,0	107,4	
2003 1-11	97,4	110,5	119,7	87,7	95,7	77,5	40,9	87,5	105,0	96,3	108,1	108,5	
2004 1-11	98,0	107,1	115,6	86,0	93,9	75,8	26,3	84,5	103,9	96,7	107,2	108,5	
2003 12	102,0	110,5	123,2	79,1	100,3	58,1	25,2	91,3	102,0	100,9	117,2	117,6	
2003 1	90,4	105,0	120,2	67,0	100,3	35,7	16,6	90,2	95,9	87,4	95,2	95,8	
2004 1	90,9	105,4	117,2	76,1	86,7	67,0	15,9	94,4	78,8	87,1	96,3	96,7	
2004 2	99,9	105,9	116,4	80,0	76,7	79,6	20,3	88,6	89,8	97,8	105,2	105,8	
2004 3	110,0	87,6	82,7	99,6	101,2	92,3	41,9	85,4	123,2	109,5	110,6	111,8	
2004 4	99,2	104,8	110,7	90,1	103,5	76,2	36,6	70,8	109,6	98,9	99,8	101,1	
2004 5	103,3	110,8	116,4	96,9	104,2	87,1	33,2	80,9	116,2	103,5	101,5	102,8	
2004 6	102,5	111,4	115,7	100,9	99,1	99,5	24,3	80,1	113,4	102,6	100,3	101,4	
2004 7	108,2	111,1	118,4	93,0	100,8	84,5	30,5	83,0	106,5	108,0	108,6	110,0	
2004 8	52,8	102,1	120,9	55,3	76,0	40,1	12,0	64,3	60,1	47,9	88,1	89,9	
2004 9	105,2	108,3	117,4	85,6	88,9	77,7	28,0	88,3	106,2	105,1	126,5	128,2	
2004 10	103,7	115,3	126,4	87,8	97,7	70,9	23,1	92,9	122,8	103,0	125,0	126,8	
2004 11	102,1	115,0	129,0	80,3	96,2	58,8	23,0	100,4	116,1	100,5	117,2	118,7	

FOLLOW

	segue Attività manifatturiere										Industrie tessili e dell'abbigliamento		
	segue Industrie alimentari e delle bevande e del tabacco										Industrie tessili		
	segue Industrie alimentari e delle bevande										Industrie tessili		
	Pesce e prodotti a base di pesce	Frutta e ortaggi	Oli e grassi vegetali e animali	Industria lattiero-casearia e dei gelati	Lavorazione delle granaglie e di prodotti amidacei	Prodotti per alimentazione animali	Altri prodotti alimentari	Industria delle bevande	Industria del tabacco	Totale	Totale	Preparazione e filatura di fibre tessili	
2002	111,0	109,4	105,0	101,0	100,4	115,7	106,7	106,3	85,8	93,9	93,0	92,4	
2003	115,0	107,8	101,3	104,2	102,5	112,4	109,4	108,1	92,1	90,6	89,4	87,8	
2003 1-11	116,9	109,9	102,5	105,7	103,1	111,8	110,5	109,5	94,2	90,9	90,7	89,2	
2004 1-11	111,0	114,2	103,5	102,9	102,7	113,8	112,1	106,1	53,2	89,4	87,3	84,7	
2003 12	124,4	142,0	92,8	88,1	104,3	109,1	135,1	108,1	103,5	92,1	93,9	95,2	
2004 1	94,8	84,3	87,5	88,7	96,0	119,3	96,8	92,5	69,6	87,0	74,9	72,3	
2004 2	86,0	77,9	88,6	101,6	100,1	120,0	98,5	88,6	79,5	83,7	78,1	85,5	
2004 3	121,3	91,8	97,4	105,4	96,7	109,4	113,5	98,2	79,9	95,0	88,0	94,6	
2004 4	142,4	90,9	112,9	118,0	109,3	125,1	111,2	110,5	58,1	105,0	99,9	106,3	
2004 5	123,3	77,3	102,9	118,1	102,9	120,7	92,9	106,8	44,6	92,5	97,9	96,5	
2004 6	131,2	81,0	117,4	116,1	105,1	111,9	94,8	111,0	48,9	103,1	101,4	111,6	
2004 7	125,7	112,5	98,8	106,0	102,2	112,2	89,8	116,8	54,5	100,3	94,6	95,3	
2004 8	128,3	127,7	136,1	105,0	107,9	110,6	101,6	125,9	47,7	107,2	97,1	87,7	
2004 9	47,9	125,3	59,2	88,7	77,0	109,6	89,8	82,0	15,0	38,5	22,6	16,1	
2004 10	108,9	160,8	107,2	93,8	103,4	112,3	153,8	111,2	56,0	91,0	99,1	80,6	
2004 11	104,2	171,9	105,7	91,7	113,5	109,5	151,4	106,1	50,3	81,9	93,0	80,9	
2004 11	101,8	138,9	112,3	87,8	111,3	110,9	135,5	109,8	51,1	85,7	88,6	76,3	

FOLLOW TO NEXT PAGE

La scelta di questa struttura dati ha permesso agli sviluppatori di relazionarsi in maniera semplificata ed intuitiva alla struttura fisica delle tavole. L'astrazione data dalle classi, inoltre, ha reso meno complicato il continuo colloquio con i dati contenuti nell' *Universo* .

Riguardo alla comunicazione client-server si è cercato di limitare il più possibile ritardi di risposta evitando frequenti richieste dati al database.

L'aver scelto un'interfaccia di tipo step-by-step avrebbe reso necessario un continuo scambio di dati con il database remoto. Infatti, per dare maggior libertà di movimento fra gli step il sistema sarebbe stato costretto a caricare e salvare i dati singolarmente ad ogni passo.

Per evitare tale "collo di bottiglia" si è preferito limitare questa interazione in due momenti: un iniziale ingente caricamento di dati, ed una finale memorizzazione di tutta la struttura dati.

Tale scelta ha portato ad una maggiore fluidità nell'intero processo descrittivo in termini di risposta del programma a scapito di una maggiore attesa nella fase iniziale di recupero dati ed al completamento di un insieme minimo di operazioni per poter avere la possibilità di salvare i dati sul database.

Al primo problema si è posto rimedio provvedendo alla creazione di una tavola temporanea che viene mantenuta in memoria per tutta la sessione di lavoro (di tale opzione viene comunque lasciata libertà di scelta all'utente al momento di salvare le prime modifiche) e limitando, quindi, i caricamenti solo nel caso in cui si decida di cambiare la tavola su cui si vuole lavorare.

Per quanto riguarda le operazioni minime da compiere si è diviso il problema nel caso di inserimento o di modifica: nel primo l'utente deve almeno inserire una sottotavola, nel secondo è agevolato dall'utilizzo di opportuni pulsanti per i salvataggi immediati.

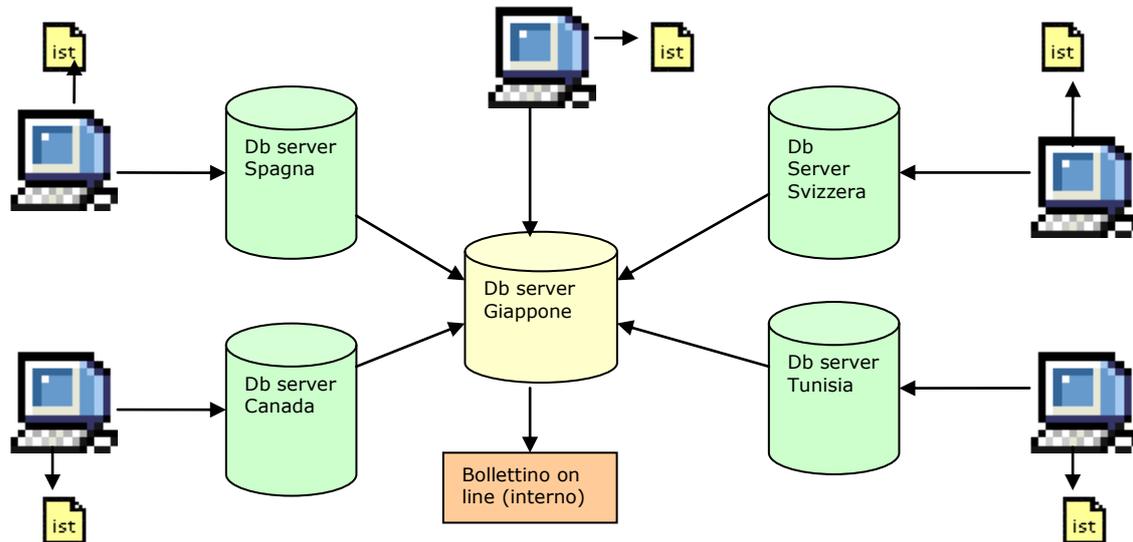
Per proteggere i dati del sistema, da ogni eventuale problema di memorizzazione/cancellazione, si è prevista un'opportuna gestione delle transazioni.

In caso di errore in inserimento o modifica tutte le operazioni effettuate sul database dall'ultimo salvataggio vengono eliminate ed il database ripristinato al suo ultimo stato corretto.

2.7 Procedure di aggiornamento della banca dati

La struttura della banca dati della diffusione congiunturale risulta essere un'integrazione della banca dati di serie storiche denominata Conistat già in funzione da diversi anni. Risulta, quindi, composta da due blocchi:

- una struttura distribuita che prevede in ogni sede Istat la presenza di un database server contenente i dati e metadati di Conistat relativi ai Servizi produttori che si trovano in quelle sedi e la presenza di un server centralizzato che contiene, per la parte riguardante i dati, i link ai dati dei database server locali;
- un database server centralizzato, periodicamente aggiornato con i dati di tutti i Servizi.



Sulla prima struttura accede un'applicazione Web interna analoga a quella esterna (che invece accede alla seconda struttura) e che viene utilizzata dai Servizi per controllare i dati prima di inviare i file di dati all'esterno.

Le procedure che provvedono al popolamento delle due strutture sono rispettivamente il Check ed il Loader.

2.7.1 Il modulo Check

La procedura Check, già precedentemente utilizzata per l'aggiornamento dei dati sul database Conistat è stata modificata per utilizzarla anche nel caso dell'aggiornamento dei dati del BMS.

Il Check è stato sviluppato in Visual Basic 6 e lavora in ambiente client-server in cui i client risiedono sulle macchine dei vari Servizi produttori di dati e i dati risiedono in un sistema di database distribuiti.

Come è stato precedentemente affermato il deposito dati utilizzato è quello di Conistat (vedi par. 2.5), che presenta una struttura distribuita.

Lo scopo di questa procedura è sia quello di effettuare un controllo sui dati prima dell'inserimento nel database, sia l'aggiornamento automatico dei data base server locali situati nelle varie sedi.

Una volta avvenuto il collegamento vengono effettuati i controlli sui codici e sui dati. In questa fase sono state fatte delle modifiche per adattare il software usato per Conistat al Bollettino on line.

E' stato infatti introdotto un database locale in cui sono presenti tutte le serie che il Servizio deve inviare.

Il motivo di questo cambiamento è dovuto al fatto che il Bollettino mensile on line a differenza di Conistat presenta le serie in formato tabellare per cui è necessario che le variabili presenti nella tavola vengano inviate tutte dal servizio per non portare ad errori nella visualizzazione della tavola stessa.

Il database locale viene precaricato la prima volta prelevando le informazioni dal database server centrale e aggiornato in maniera incrementale le volte successive. Il motivo per cui è stato deciso di utilizzare un database locale è dovuto solo a problemi di performance.

Si effettua quindi il controllo dei codici, che può avvenire secondo due modalità: una totale e una parziale. La procedura di controllo parziale controlla solo che i codici inviati nel file .dat siano presenti all'interno del database server. La procedura di controllo totale oltre ad effettuare un controllo analogo a quello della procedura parziale, confronta le variabili presenti nel file .dat relative all'ultimo periodo con quelle che sono all'interno del database locale.

Nel caso in cui venga riscontrata la mancanza di una data variabile, viene chiesto all'utente se inserire un valore per la variabile mancante o inserire un valore "-" che indica che la variabile non viene rilevata ma compare ugualmente sulla tavola.

Una risposta negativa ad entrambe le alternative interrompe il programma.

Se un dato viene segnalato come non rilevato, viene memorizzato all'interno del database come non rilevato (flag_nr della tabella contenente le serie), questo è il motivo per cui gli aggiornamenti al database (esclusa la prima volta) devono essere incrementali, in questo modo non vengono sovrascritti i codici segnati come non rilevati.

Avviene anche un controllo sui dati che riguarda il controllo della percentuale di tolleranza: data una percentuale di tolleranza nel menù di impostazioni, il programma controlla in base alla frequenza della serie la differenza del valore corrente con quello dell'anno precedente (nel caso di frequenza annuale), con quello dello stesso mese dell'anno precedente (nel caso frequenza mensile) e con quello dello stesso periodo dell'anno precedente nel caso di altri tipo di frequenza. Se tale valore supera quello di tolleranza

viene segnalato un “alert” che non impedisce il proseguimento del check ma che avverte l’utente di tale anomalia.

Se il controllo dei codici passa senza errori viene aggiornato il database server locale e viene creato un file con estensione .ist, che viene utilizzato per l’aggiornamento del database a cui accede l’applicazione esterna.

Il file .ist che viene generato contiene dati non ancora presenti nel database o dati che devono essere modificati.

La procedura crea inoltre (sempre su richiesta dell’utente) un file di ripristino in cui vengono memorizzati (così come accade nelle funzioni di rollback dei database) i record preesistenti con accanto l’operazione usata (inserimento o modifica) per l’aggiornamento del record.

In questo modo, in caso di un caricamento errato c’è la possibilità (prima di far rigirare il Check con i dati corretti) di ripristinare la situazione precedente al caricamento errato.

2.7.2 Il modulo Loader

Il modulo Loader è utilizzato in calce al modulo Check per aggiornare i database interni (intranet), e serve anche per l’aggiornamento del database esterno (internet).

Il modulo Loader è stato sviluppato in ambiente Client-Server in cui il data base Server è Access ed è stato realizzato con Visual Basic 6.

I file .ist inviati dai Servizi vengono inseriti all’interno del database secondo un calendario precedentemente fissato.

I Servizi possono inviare anche aggiornamenti di dati precedentemente inseriti, è fondamentale per questo rispettare l’ordine cronologico con cui vengono inseriti i file.

Se deve essere inserito il file relativo ad un determinato servizio il software controlla che non ci siano altri inserimenti precedenti, relativi al Servizio, non ancora effettuati.

Ogni volta che viene effettuato un nuovo inserimento viene aggiornato un file di news in cui compaiono tutti gli inserimenti più recenti.

La durata di permanenza di una news viene fissata nel menù delle impostazioni.

Ogni volta che viene effettuato un nuovo inserimento il software elimina tutte le news più vecchie del valore impostato precedentemente e controlla che non esistano news relative allo stesso Servizio nello stesso giorno, infine inserisce la nuova news.

2.7.3 Il modulo iSCAN

Lo scopo di questo nuovo modulo è quello di consentire agli utenti che posseggono i dati in formato Excel l'inserimento di nuove classificazioni e di associare la metainformazione alla struttura della tavola.

La procedura controlla che tutte le variabili di classificazione della tavola siano presenti; nel caso in cui una variabile di classificazione sia assente, la inserisce assegnandogli un codice incrementale; se invece trova una nuova voce di territorio avverte l'utente che la voce presente sulla tavola Excel non esiste all'interno del database. Le voci territoriali non vengono inserite automaticamente in quanto in diverse tavole possono comparire delle diciture diverse di stesse voci (ad esempio la voce "Trento e Bolzano" si può trovare anche come "Trento-Bolzano") per cui, per mantenere uniformità nelle descrizioni territoriali, si preferisce inserirle manualmente.

Per ogni variabile che viene introdotta viene creato un record nella tabella metadati e vengono aggiornate le tabelle relative alla tavola indicando la posizione all'interno della tavola dell'incrocio.

Nella maschera di inserimento è necessario inserire oltre a tutte le informazioni sul foglio Excel anche altre informazioni riguardanti i metadati, ovvero il periodo di inizio e di fine delle serie, l'unità di misura, la tipologia (grezzo o destagionalizzato).

2.8 Il modulo webTable

Questo modulo permette ad un utente che utilizza un browser Web⁴ di accedere all'interfaccia di navigazione per ricercare una tavola d'interesse e visualizzarla.

In particolare vengono implementate tre macro funzioni:

- autenticazione;
- ricerca;
- presentazione della tavola.

2.8.1 Autenticazione

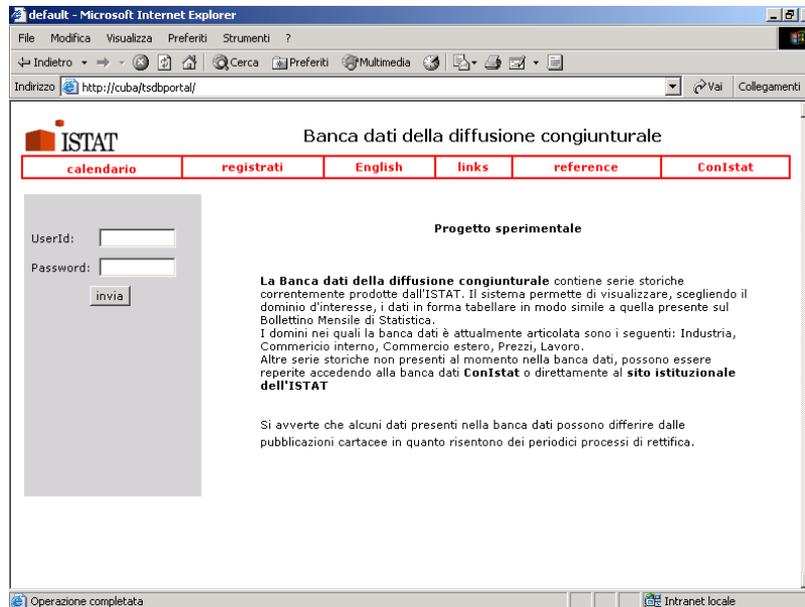
La necessità di garantire all'utente un unico punto di autenticazione nel momento in cui accede ai dati congiunturali, attraverso Internet, ha portato alla realizzazione di un semplice sistema di Single Sign On (SSO). In pratica un utente che ha necessità di accedere a delle informazioni in differenti banche dati, magari posizionate su due server differenti, non deve più avere due identificativi diversi, ma può utilizzare uno stesso identificativo.

⁴ Il sistema è stato testato con i seguenti browser: MS-Explorer 5.0 o sup; Mozilla 1.7 o sup

In attesa che l'Istat implementi a livello generale un sistema di SSO, quello utilizzato nel progetto è stato realizzato attraverso un Web service.

Il Web service è stato posizionato su un computer che agisce da server per tale servizio. Esso ha la possibilità di interfacciare il database in cui sono memorizzati gli utenti abilitati ad accedere alle banche dati.

In pratica la funzione di autenticazione agisce come client SOAP (Simple Object Access Protocol) che invia al Web service la userid e la password che un utente specifica nei campi di una maschera preposta allo scopo. Il Web service verifica nel database se tale utente è abilitato all'accesso alla banca dati, quindi, invia indietro al client una risposta positiva o negativa.



2.8.2 Ricerca

La progettazione dell'interfaccia di ricerca ha privilegiato l'impostazione "per indice" attualmente presente sulla pubblicazione cartacea.

In particolare ciascuna tavola è identificabile tramite un percorso gerarchico composto da un dominio e un sottodominio:

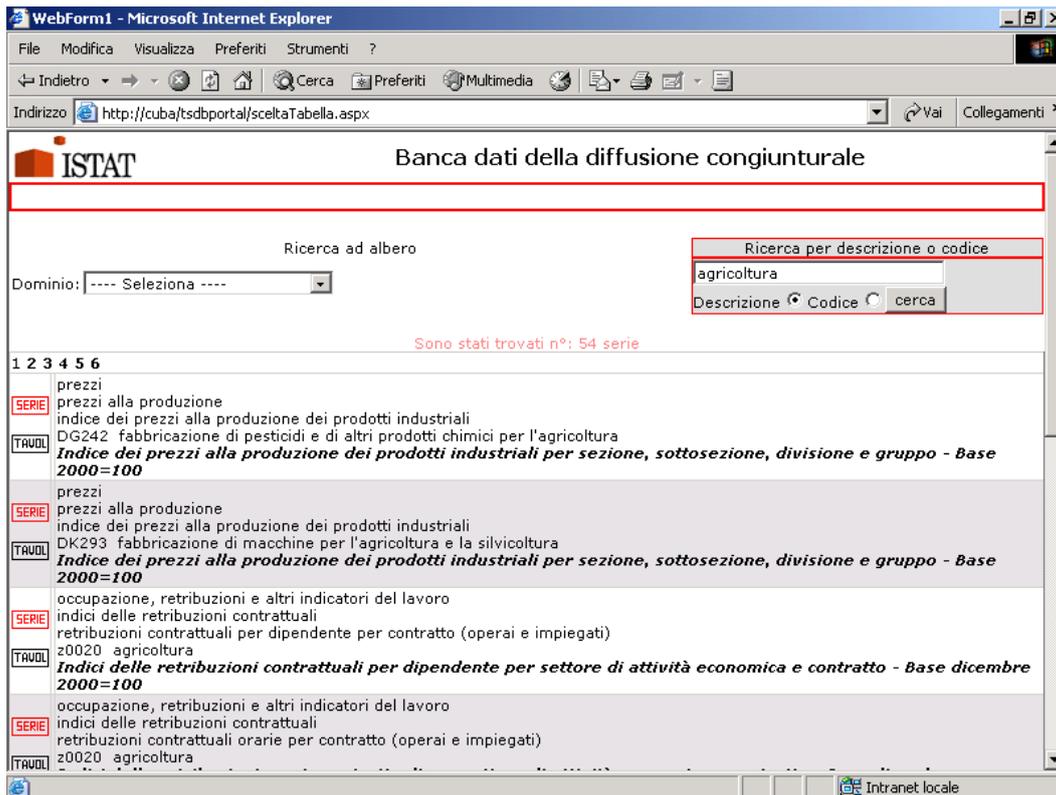
DOMINIO
Meteorologia
Popolazione
Sanità
Istruzione
Giustizia

DOMINIO
Agricoltura e foreste
Industria
Trasporti
Commercio interno
Commercio con l'estero
Credito
Mercato monetario e finanziario
Prezzi
Lavoro
Finanza pubblica
Contabilità Nazionale

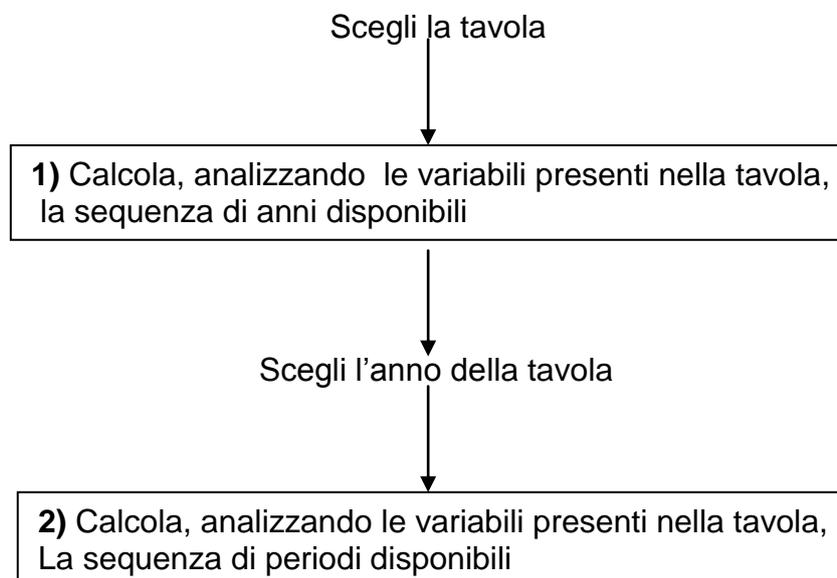
Scegliendo l'opportuno dominio, il sistema presenta i sottodomini di competenza per quel dato dominio. Per esempio, se l'utente sceglie il dominio "Lavoro", i sottodomini saranno:

SOTTODOMINIO
Rilevazione Forze di Lavoro
Indicatori del lavoro nelle grandi imprese della industria e dei servizi
Retribuzioni
Costo del lavoro, retribuzioni e oneri sociali

Come funzionalità aggiuntiva è stata implementata una ricerca per parola chiave o per codice. Per esempio se l'utente ricerca per la parola "agricoltura", il sistema, nel caso la ricerca vada a buon fine, propone tutte le tavole in cui tale parola compare nelle descrizioni delle singole variabili. A questo punto il sistema dà la possibilità o di accedere alla singola variabile per una data tavola, oppure alla tavola nel suo insieme.



Prima di avviare la presentazione della tavola selezionata, l'utente può definire l'anno e il periodo iniziale dai quali far partire la tavola. Il sistema presenta valori di anno e periodo compatibili con la tavola scelta. L'algoritmo che si applica è il seguente:



In particolare:

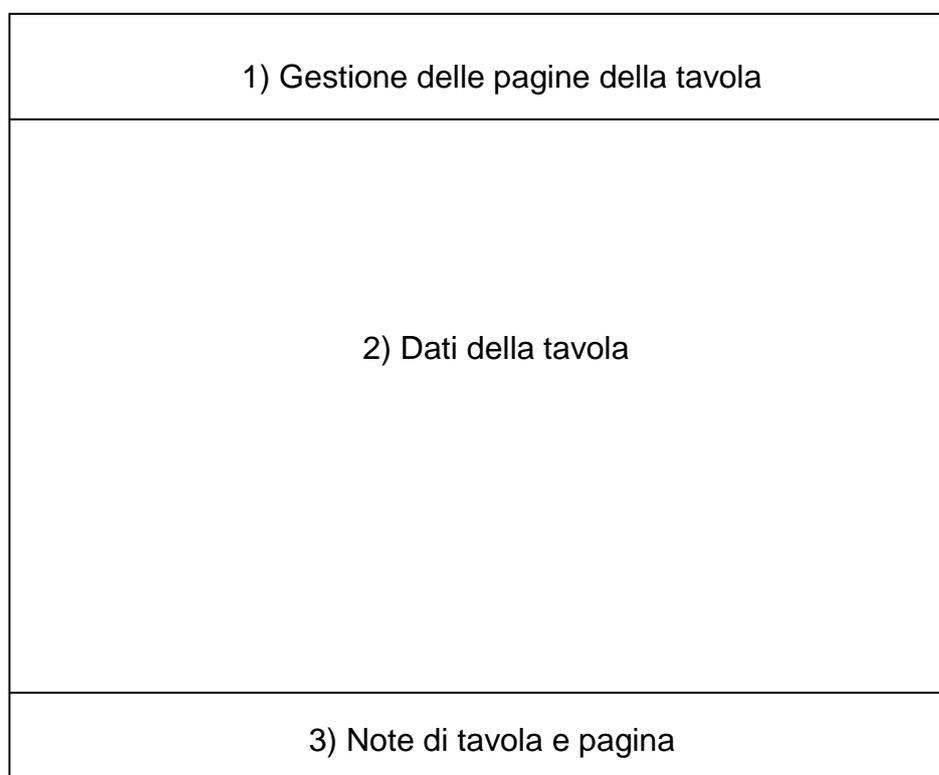
nel punto 1) il sistema confronta l'anno finale e iniziale di ciascuna variabile (serie storica) presente nella tavola e crea un intervallo delimitato dal massimo anno iniziale e minimo anno finale;

nel punto 2) il sistema confronta il periodo finale di ciascuna variabile presente nella tavola per l'anno scelto e impone il valore minimo trovato come valore massimo dell'intervallo.

2.8.3 Presentazione della tavola

Questa funzionalità tiene conto del tipo di layout scelto in fase di progettazione della tabella attraverso il programma iDesigner.

In ogni caso viene presentata a video una schermata divisa in tre aree:



1) area gestione delle pagine.



Questa area è suddivisa su più zone:

- zona della paginazione - potendo ciascuna tavola essere organizzata su più pagine, questa zona risulta preposta alle funzionalità di paginazione. Il sistema calcola il numero di pagine per una data tavola e presenta un sistema di paginazione sia

grafica (incremento o decremento del numero di pagina per salti di uno) che numerica, con la possibilità di saltare direttamente alla pagina d'interesse;

- zona informativa - in tale zona viene segnalato l'anno e il periodo scelto per la tavola, una icona che simbolizza se esistono note di pagina, una icona che permette di fare una esportazione su excel, una icona che permette di accedere ai metadati metodologici delle variabili presenti nella tavola;

2) area note di tavola e di pagina.

In questa area il sistema presenta una icona dalla quale è possibile accedere, in modalità pop-up, ad una pagina contenente la nota allegata alla tavola. Inoltre vengono ricercate e formattate le note della pagina che si sta visualizzando. Tali note possono far riferimento a una qualunque variabile e/o una sottopagina;

3) area dei dati.

In questa area vengono presentati i dati organizzati per colonne. Ciascuna colonna contiene i dati, in serie storiche, di una data variabile. Nella presentazione dei dati, il sistema segue le seguenti regole:

- nel caso siano disponibili, vengono presentati 13 mesi con sequenzialità inversa a partire dell'anno e periodo scelto. Così, se viene scelto l'anno 2004 e il periodo novembre, il sistema presenterà una serie storica che va da novembre 2003 fino a novembre 2004. Viene quindi presentato il valore cumulato per gennaio – periodo scelto per l'anno corrente e l'anno precedente, quindi la media annuale dei due anni precedenti a quello scelto;
- i valori delle cumulate vengono presentati solo se definiti nell' iDesigner e se l'utente non ha scelto come periodo da cui partire gennaio o dicembre;
- i valori delle medie annuali vengono calcolati solo se definiti nell' iDesigner;
- il sistema nel calcolare gli annuali e le cumulate distingue se il valore della variabile in questione è un indice (in tal caso vengono calcolati come media) o è un valore assoluto (in tal caso vengono calcolati come somme).

Nell'area dati, legate a ciascuna variabile, vengono presentate delle icone (triangoli neri) che permettono di accedere ad una pagina di pop-up con le seguenti informazioni:

- metadati strutturali della singola serie;
- tutta la serie storica disponibile per quella variabile;
- nel caso sia stato definito nell'iDesigner, la variazione congiunturale e tendenziale della variabile.

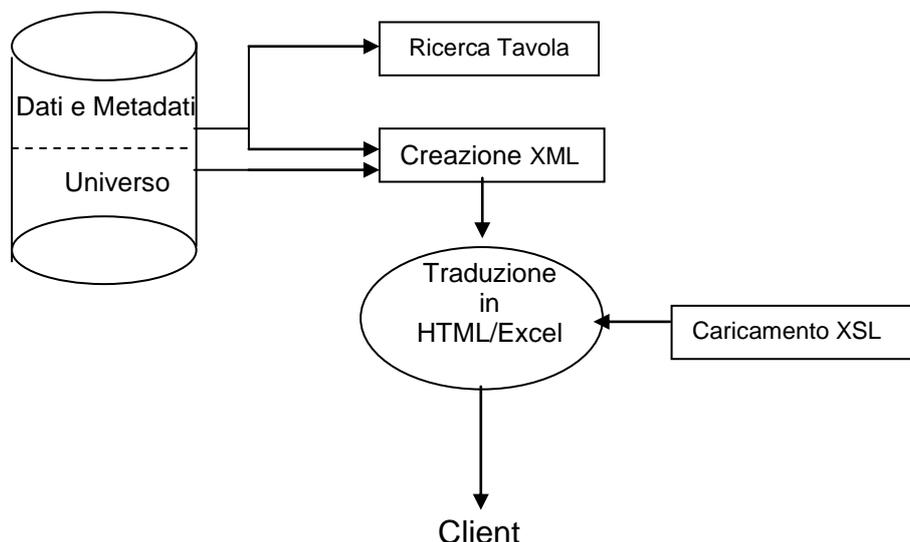
2.8.4 Uso di XML ed XSL per la realizzazione delle tavole

La realizzazione di tavole dinamiche, prelevando sia la metainformazione che i dati da un database, ha comportato la sperimentazione, attraverso il prototipo, di varie possibilità. L'ipotesi da cui si è partiti è stata quella di ottenere qualcosa di molto simile a ciò che viene proposto nella pubblicazione cartacea, e di realizzare qualcosa di modulare in grado di essere ampliato in futuro senza dover cambiare completamente l'impianto del sistema. Dunque, il sistema dovrà permettere di formattare le informazioni statistiche in tavole complesse con le seguenti caratteristiche:

- singola pagina con più sottopagine ciascuna con una propria testata. In pratica questo si può pensare come una pagina su cui collocare più tabelle statistiche;
- singola pagina con una singola testata, ma sottopagine identificate da una quarta dimensione;
- testate sviluppate su 3 livelli in grado di dare un senso di raggruppamento alle variabili
- sviluppo della dimensione tempo o territorio o categoria sulla prima colonna;
- calcolo di medie (annuale e cumulata) nel caso di loro definizione attraverso iDesigner.

Si è arrivati, quindi, a un modello di sistema basato su XML per l'organizzazione dei dati per la singola pagina e XSL per la formattazione.

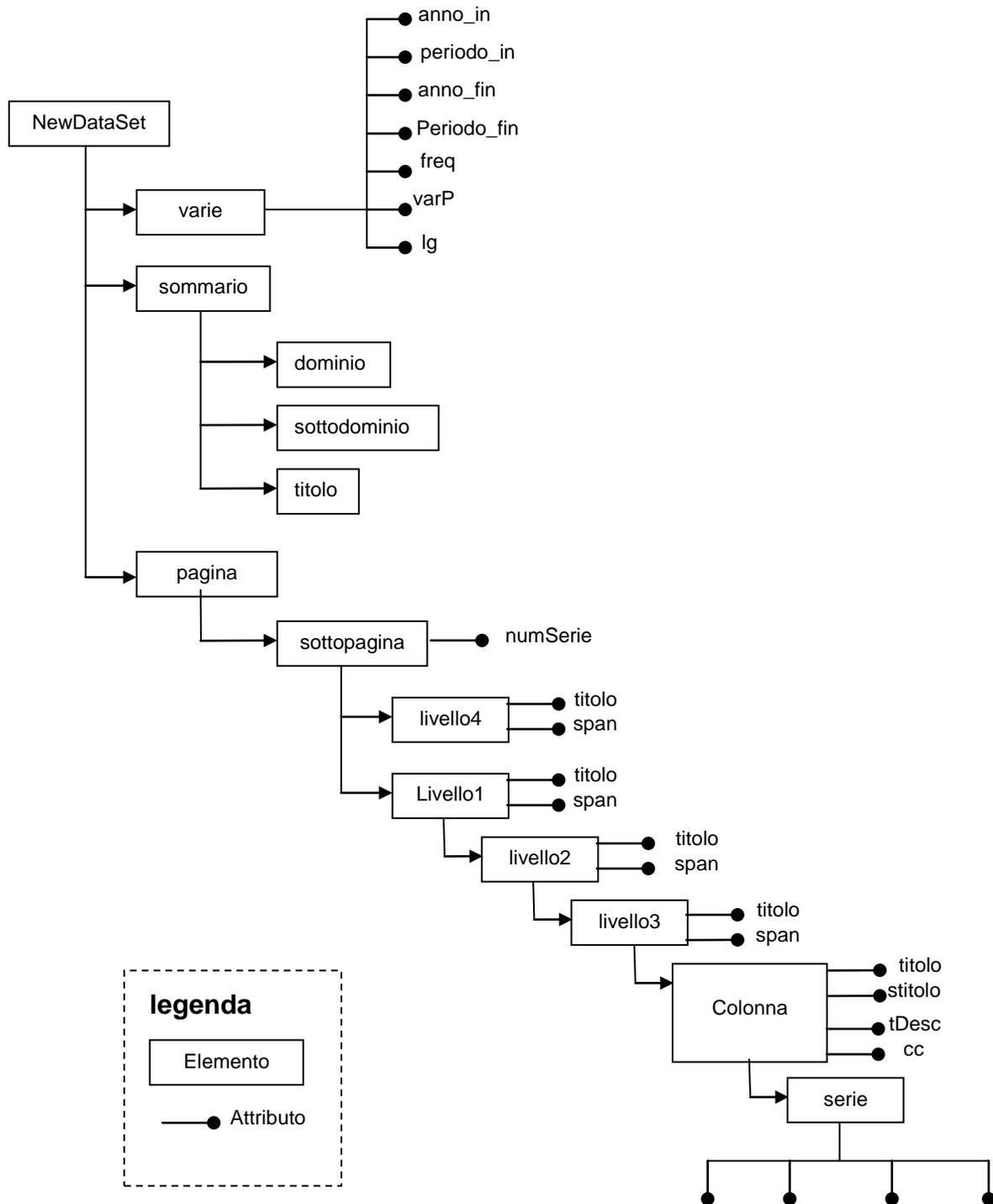
Il seguente schema illustra il modello:



La traduzione XML + XSL = HTML viene fatta sul server, quindi il client riceve soltanto uno stream in formato HTML. Tale operazione sarebbe stato possibile farla realizzare direttamente al client, ma non tutti i client sono abilitati a farlo.

Inoltre per ragioni di performance non viene creato fisicamente un file XML su disco, ma viene mantenuto uno stream XML in memoria.

L'organizzazione dei dati nel formato XML è la seguente:



Di seguito vengono analizzati gli elementi e gli attributi del formato XML:

- elemento radice <NewDataSet>;
- elemento <varie> presenta i seguenti attributi:
 - attributo anno_in – indica l'anno in cui inizia la serie storica
 - attributo periodo_in - indica il periodo in cui inizia la serie storica
 - attributo anno_fin – indica l'anno scelto dall'utente
 - attributo periodo_fin – indica il periodo scelto dell'utente
 - attributo freq – indica la frequenza di tutte le serie storiche presenti nella tavola
 - attributo varP – indica se visualizzare il calcolo delle variazioni congiunturali e tendenziali nella finestra di accesso alla singola serie contenuta nella tavola
 - attributo lg – indica la lingua scelta dall'utente (attualmente italiano o inglese)
- elemento <sommario> contiene i seguenti sottoelementi:
 - <dominio>
 - <sottodominio>
 - <titolo>
- elemento <pagina> contiene i seguenti sottoelementi:
 - sottopagina
 - attributo numSerie – specifica il numero di variabili presenti nella sottopagina
 - <livello4> - questo elemento è significativo solo in tavole con layout di tipo 2
 - attributo titolo
 - attributo span
 - <livello3>
 - attributo titolo
 - attributo span
 - <livello2>
 - attributo titolo
 - attributo span
 - <livello1>
 - attributo titolo
 - attributo span
 - <colonna> - specifica la variabile
 - attributo titolo
 - attributo stitolo – descrizione breve

- attributo tDesc – nel caso assuma il valore 1 verrà visualizzata la descrizione breve
- attributo cc – identificativo della nota associata alla variabile
- <serie>
 - attributo freq – la frequenza della serie
 - attributo anno – anno di quella osservazione
 - attributo mese – periodo di quella osservazione
 - attributo valore – valore dell'osservazione

Segue un estratto di un esempio di un file XML

```
<?xml version="1.0" standalone="yes" ?>
<NewDataSet>
<varie anno_in="2001" periodo_in="1" anno_fin="2004" periodo_fin="12" freq="12"
  varP="S" lg="it" />
<sommario>
  <dominio>Lavoro</dominio>
  <sottodominio>Retribuzioni</sottodominio>
  <titolo>Indici delle retribuzioni contrattuali per dipendente per settore di
    attività economica e contratto - Base dicembre 2000=100</titolo>
</sommario>
<pagina>
  <sottopagina numSerie="13">
    <livello4 titolo="COMPLESSO" span="13" />
    <livello1 pkLivello1="02" titolo="" span="1">
      <livello2 pkLivello2="08" titolo="" span="1" fkLivello1="02">
        <livello3 pkLivello3="17" titolo="" span="1" fkLivello2="08">
          <colonna cod="6093" titolo="INDICE GENERALE" stitolo="" tDesc="0"
            cc="">
            <serie freq="14" anno="2002" valore="104,0" />
            <serie freq="14" anno="2003" valore="106,3" />
            <serie freq="12" anno="2003" mese="12" valore="107,4" />
            <serie freq="12" anno="2004" mese="1" valore="107,6" />
            <serie freq="12" anno="2004" mese="2" valore="108,3" />
            <serie freq="12" anno="2004" mese="3" valore="108,4" />
            <serie freq="12" anno="2004" mese="4" valore="108,6" />
            <serie freq="12" anno="2004" mese="5" valore="109,1" />
            <serie freq="12" anno="2004" mese="6" valore="109,2" />
            <serie freq="12" anno="2004" mese="7" valore="109,9" />
            <serie freq="12" anno="2004" mese="8" valore="109,9" />
            <serie freq="12" anno="2004" mese="9" valore="110,0" />
            <serie freq="12" anno="2004" mese="10" valore="110,1" />
            <serie freq="12" anno="2004" mese="11" valore="110,2" />
            <serie freq="12" anno="2004" mese="12" valore="110,9" />
          </colonna>
        </livello3>
      </livello2>
    </livello1>
  </sottopagina>
</pagina>
</NewDataSet>
```

2.8.5 La realizzazione di un oggetto capace di creare il file XML

Innanzitutto c'è da dire che per motivi di performance non viene creato un file XML, ma uno stream in memoria che rappresenta una sequenza di caratteri che, se visualizzati, danno luogo ad una struttura XML. Per far ciò abbiamo utilizzato un oggetto, messo a disposizione dal framework .NET, denominato dataset. Un dataset nella sua forma più semplice può rappresentare in memoria una tabella di un database, popolata con tutti i dati della tabella fisica oppure un suo sottoinsieme attraverso una operazione di "where" in una query SQL. Nella sua forma più completa può rappresentare, in memoria, un intero database, con le sue tabelle, le sue relazioni le sue regole di integrità. C'è da dire, inoltre, che il dataset è un oggetto disconnesso. Questo significa che, nel momento in cui vengono estratti i dati da un database fisico, è possibile chiudere la connessione verso tale database a tutto vantaggio della scalabilità, e quindi realizzare svariate operazioni sul dataset disconnesso, come aggiunta di record nelle sue tabelle, filtraggi, aggiunta di nuove tabelle come risultato di operazioni su altre tabelle, eccetera.

Una delle funzionalità più interessanti di questo oggetto è la sua convertibilità in un file XML e viceversa. Questo ha permesso di lavorare in un ambiente relazionale e, infine, utilizzare la funzionalità prima esposta per convertire il tutto in uno stream XML senza doverci preoccupare di implementare le regole classiche di qualunque file XML cioè di buona formattazione (well-formed) e di validità.

L'oggetto creato da luogo ad un dataset rappresentativo dello stream XML necessario. Tale oggetto presenta un solo metodo che accetta alcuni parametri e restituisce un dataset.

Il dataset restituito dal nostro oggetto presenta tabelle realizzate in due modi:

- 1) attraverso operazioni di filtraggio sulle tabelle fisiche del database;
- 2) attraverso la creazione a run-time.

Per ciascun campo delle varie tabelle è possibile impostare la proprietà ColumnMapping ai valori Element/Attribute/Hidden che permette in fase di traduzione verso l'XML di convertire il valore di quei campi in elementi, attributi o non facenti parte dell'XML.

2.9 iConverter

Per soddisfare la necessità di disporre in maniera repentina dei dati componenti le tavole del Bollettino in formato Excel da poter inviare alla tipografia per la produzione della versione cartacea, si è dotato il sistema iTable di un software specifico alla produzione di tali file.

A differenza degli altri moduli desktop del sistema iTable il presente software è stato sviluppato in Microsoft Visual Basic .NET 2003 per poter utilizzare algoritmi già sviluppati nel modulo webTable (componente del sistema iTable dedicato alla visualizzazione online delle tavole).

La possibilità di riutilizzare il codice già presente, la diffusione del Framework .NET sulle piattaforme Microsoft odierne, la possibilità di installarlo rapidamente sulle versioni di Windows meno recenti hanno favorito la scelta della nuova versione di Visual Basic.

Ciò ha condizionato la scelta del paradigma di programmazione verso OOPS (Object Oriented Programming System) che, anche se non strettamente necessaria nell'implementazione dell'interfaccia e nello sviluppo dei processi di interazione/elaborazione, si adatta efficacemente nell'utilizzo degli oggetti COM (Component Object Model) e della riutilizzazione del codice del webTable.

Come rapidamente accennato la produzione del file Excel è stata permessa dall'utilizzo dell'oggetto COM Microsoft Excel 11.0 Object Library.

Da un punto di vista puramente tecnico l'architettura COM è un potente strumento che consente di modellare il software a runtime in quanto combina tra di loro dei componenti durante l'esecuzione di un'applicazione. Un po' come prelevare qualcosa che ci serve da uno scaffale mentre si sta eseguendo un preciso lavoro. Infatti COM consente di costruire dei moduli software "riutilizzabili", dotati di una serie di funzioni che un qualsiasi oggetto di un'altra applicazione può adoperare.

Questa potenzialità, adottata nelle librerie Microsoft Office, ha permesso di interagire con semplicità con Excel; si è potuta instaurare una comunicazione diretta che ha permesso l'apertura di nuovi workbooks e la creazione di fogli su cui poter inserire i dati direttamente dal codice Visual Basic dell'iDesigner. L'aver scelto questa soluzione ha permesso una rapidità di sviluppo molto superiore ad un'eventuale creazione di un convertitore in formato Excel. iConverter prevede un'interfaccia immediata ed un'elevata semplicità di utilizzo.

Il compito dell'utente è limitato alla selezione del dominio e sottodominio della tavola, per restringere l'elenco dei titoli di tavola da presentare, la selezione del titolo e la data delle serie richiesta. Sarà compito, totalmente automatico, dell'iDesigner l'apertura e la generazione del nuovo foglio Excel rispettante il formato richiesto per la versione cartacea delle tavole.

Appendice A - Le tecnologie XML nella pubblicazione dei dati statistici

1. Introduzione

Nell'ambito della diffusione dei dati statistici nell'arco di pochi anni molto è cambiato. E' possibile individuare schematicamente le seguenti fasi:

- fino a tutti gli anni '80 diffusione su carta;
- dagli anni '90 diffusione anche su supporto magnetico (tabelle Excel e timidi sistemi di interrogazione personalizzati);
- a fine anni '90 diffusione su supporto magnetico (banche dati) e prime esperienze in ambito Internet (tabelle *HTML* o *Excel*) con contenuto tipicamente statico;
- dall'anno 2000 piena diffusione su Internet sia di contenuti statici e che dinamici (banche dati).

Ora che la tecnologia su Internet è pienamente matura sia in termini di strumenti di sviluppo che di livelli di accesso (linee veloci e a basso costo), molta dell'informazione statistica verrà spostata online ed è proprio in quest'ambito che la concorrenza sarà più agguerrita. Le tecniche di diffusione delle informazioni su Internet sono passate nel giro di poco tempo dalle classiche pagine *HTML* in cui attraverso un sistema di hiperlink era possibile saltare tra vari documenti anche e spesso a costo di seguire percorsi tortuosi, a sistemi dinamici in cui l'utente può personalizzare la modalità di presentazione dell'informazione di cui ha bisogno. In particolare, per quanto riguarda l'informazione statistica, non risulta più sufficiente la semplice consultazione di tabelle statiche, ma l'utente sempre più ha la necessità di fare estrazioni e formattazione dei dati personalizzate ed eventualmente elaborare online.

In questo scenario le seguenti linee guida saranno i vincoli imprescindibili dai quali nessun sistema informativo orientato alla diffusione potrà prescindere:

- qualità - intesa come correttezza dei dati diffusi. In quest'ambito non esiste più il sistema delle bozze, quindi, la messa in linea di dati che contengono qualche errore dovuto anche ad una distrazione, significa la veloce propagazione dell'errore stesso, in tempi rapidissimi, tra l'utenza, e l'impossibilità di far giungere una *errata corrige* agli utenti che hanno acceduto ai dati errati;
- aumento dell'informazione;
- tempestività;
- facilità di reperimento.

Il rispetto di tali vincoli, come presupposto necessario per l'apprezzamento da parte dell'utenza dei servizi offerti, sarà sicuramente facilitato dall'adeguamento dei propri sistemi informatici con tecnologie innovative e che offrono le loro migliori performance proprio in ambito Internet.

A tale scopo già da qualche tempo in vari settori della scienze e della tecnica sono state sviluppate varie esperienze in questo senso. In particolare l'attenzione si sta concentrando sulla nuova tecnologia nata su Internet che va sotto il nome di XML(*eXtensible Markup Language*). Dopo una prima fase prototipale, sono nati dei veri e propri standard di settore quali, per esempio:

AML (*Astronomical Markup Language*) - nell'ambito della condivisione di dati astronomici;

CML (*Chemical Markup Language*) - nell'ambito della condivisione di dati chimici;

CIDS (*Component Information Dictionary Specification*) - uno standard per i componenti elettronici;

cXML (*Commerce XML*) - nell'ambito del commercio elettronico;

ICE (*Information & Content Exchange*) - standard per le agenzie di stampa.

Anche nell'ambito statistico è presente il progetto SMDX (*Statistical Data e Metadata Exchange*) che si prefigge di creare uno standard che faciliti l'interscambio di dati e metadati statistici. In particolare, proponendo realizzazioni XML, almeno nella versione attuale, sia dello standard GESMES/CB nello scambio di dati sotto forma di serie storiche, sia dello standard DSBB per ciò che riguarda il formato di diffusione dei metadati.

Questa appendice si concentrerà principalmente sui vantaggi e le possibilità che l'adozione di XML offre. Lo scopo principale non sarà quello di fornire un manuale d'uso dell'XML, per il quale si rimanda a testi specialistici, ma di illustrare la sua applicabilità nell'ambito della diffusione dell'informazione statistica. Dopo una breve introduzione tecnica sul linguaggio, verranno forniti alcuni esempi realizzati nell'ambito della *Direzione centrale per la diffusione della cultura e dell'informazione statistica* dell'ISTAT.

2. XML il metalinguaggio utilizzato per definire altri linguaggi.

L'HTML (HyperText Markup Language) è considerato il linguaggio base del Word Wide Web. Questo linguaggio consente, infatti, di creare in maniera standardizzata pagine di informazioni formattate che possono essere facilmente accedute tramite Internet da un utente comunque dislocato. L'HTML prende in considerazione soprattutto il modo in cui le informazioni vengono presentate e non il tipo o la struttura di tali informazioni. Per colmare tale deficienza sono stato sviluppate da parte del Word Wide Web Consortium le

specifiche del linguaggio XML . L'XML è un linguaggio di markup aperto⁵ e basato su testo che fornisce informazioni di tipo strutturale e semantico relative ai dati veri e propri. Questi dati sui dati, o metadati, offrono un contesto aggiuntivo all'applicazione che utilizza i dati e consente un nuovo livello di gestione e manipolazione dell'informazione basata sul Web.

Il contesto di applicazioni dell'XML non è delineato in maniera netta, attualmente due sono le sue maggiori applicazioni: creazione di documenti (soprattutto Web) e interscambio di dati. Volendo semplificare si potrebbe dire che XML serve per rappresentare digitalmente documenti la cui informazione è strutturata gerarchicamente. Oltre ad XML esistono dei linguaggi di supporto XSL, XSLT, XPath, che consentono l'elaborazione dei dati rappresentati da XML.

I vantaggi derivanti dall'utilizzo di tale tecnologia nell'ambito della diffusione e dello scambio dei dati si possono così riassumere:

- XML è una famiglia di tecnologie standardizzate a livello mondiale da un ente indipendente: il Word Wide Web Consortium;
- XML è supportato dai maggiori fornitori di software e molti tools di sviluppo poggiano le loro basi su tale linguaggio;
- l'uso dei Web services, che si basano proprio su XML, può rappresentare una piattaforma ideale per diffondere automaticamente informazioni attraverso sistemi di scambio diretto tra sistemi informativi differenti;
- XML, permette di realizzare procedure generalizzate per costruire tabelle statistiche complesse separando le informazioni dalla loro formattazione.

3. Composizione di un documento XML

Un documento XML è formato da componenti denominati *elementi*. Ciascun elemento rappresenta un componente logico del documento e può contenere altri elementi (*sottoelementi*) o del testo.

Gli elementi possono avere associate altre informazioni che ne descrivono le proprietà. Queste informazioni sono chiamate *attributi*. L'organizzazione degli elementi segue un ordine gerarchico che prevede un elemento principale chiamato *root element* o semplicemente *root*, che contiene tutti gli elementi di un documento.

In genere la struttura di un documento XML può essere rappresentata graficamente tramite un albero noto come *document tree*. Segue un esempio di documento XML che rappresenta una tabella statistica:

⁵ XML è un metalinguaggio che permette di creare dei linguaggi personalizzati di markup

```

<?xml version="1.0" encoding="utf-8" ?>
<!--Rappresentazione XML di una tabella Statistica -->
<tabella Titolo="Indice del Fatturato">
<nota>
Per gli aspetti metodologici si veda la nota informativa del 28 marzo 2003
</nota>
  <pagina numero="1">
    <colonna titolo=="Totale generale">
      <osservazione anno="1996" periodo="01" valore="76,1" />
      <osservazione anno="1996" periodo="02" valore="83,9" />
      <osservazione anno="1996" periodo="03" valore="92,4" />
    </colonna>
    <colonna titolo="Estrazione di minerali">
      <osservazione anno="1996" periodo="01" valore="45,7" />
      <osservazione anno="1996" periodo="02" valore="42,8" />
      <osservazione anno="1996" periodo="03" valore="73,2" />
    </colonna>
    <colonna titolo="Attività manifatturiera">
      <osservazione anno="1996" periodo="01" valore="12,1" />
      <osservazione anno="1996" periodo="02" valore="25,9" />
      <osservazione anno="1996" periodo="03" valore="31,1" />
    </colonna>
  </pagina>
  <pagina numero="2">
    <colonna titolo="Alimentari">
      <osservazione anno="1996" periodo="01" valore="89,1" />
      <osservazione anno="1996" periodo="02" valore="81,7" />
      <osservazione anno="1996" periodo="03" valore="85,9" />
    </colonna>
    <colonna titolo="Tabacchi">
      <osservazione anno="1996" periodo="01" valore="89,1" />
      <osservazione anno="1996" periodo="02" valore="73,9" />
      <osservazione anno="1996" periodo="03" valore="75,1" />
    </colonna>
  </pagina>
</tabella>

```

La prima riga identifica il documento come un documento XML e ne specifica la versione; per inserire un elemento è necessario prima definirlo e poi aggiungerne il contenuto; per specificare un attributo di un elemento si inserisce il nome dell'attributo con il relativo valore all'interno del tag di apertura dell'elemento; per rappresentare un commento nel linguaggio XML si utilizza la forma <!-- Commento -->, così come rappresentato nella seconda riga.

La struttura logica di un documento XML non segue una regola generale, ma dipende da come il progettista decide di organizzare gli elementi.

Un documento XML è un file di testo che può essere creato manualmente attraverso qualunque editore di testo, oppure automaticamente attraverso dei programmi software che estraggono i dati da un deposito (generalmente da un data base).

Ogni documento XML comprende sia una *struttura logica* che una *struttura fisica*. La struttura logica rappresenta un modello che indica quali elementi includere in un documento e con quale ordine. La struttura fisica contiene i dati effettivi utilizzati in un documento.

Come già evidenziato XML possiede, come caratteristiche fondamentali, la capacità di fornire una struttura ai documenti e di rendere i dati autodescrittivi. Quindi non definisce *tag*, ma fornisce le regole per poter definire i tag di cui si ha bisogno. Logicamente questa libertà di definizione potrebbe generare anarchia se non venisse in qualche modo controllata attraverso il rispetto delle regole strutturali e grammaticali. A questo scopo il linguaggio prevede due principi fondamentali:

- i documenti XML devono essere ben formattati (*well-formed*);
- i documenti XML devono essere validi.

I documenti ben formattati devono rispettare determinate regole ortografiche per la definizione della struttura del documento e per l'uso corretto dei tag:

- ogni documento deve contenere un elemento di massimo livello (root) che contenga tutti gli altri elementi del documento. Le sole parti che rimangono fuori del livello di root sono i commenti e le direttive di elaborazione;
- ogni elemento deve avere sia un tag di apertura che uno di chiusura. Nel caso di elementi vuoti, è possibile utilizzare un tag unico nella forma abbreviata (*/>*);
- gli elementi devono essere correttamente nidificati, cioè i tag di chiusura devono seguire l'ordine inverso dei rispettivi tag di apertura;
- XML è sensibile alle maiuscole (*case sensitive*);
- i valori degli attributi devono sempre essere racchiusi tra singoli o doppi apici;
- XML definisce cinque elementi sintattici, denominati *entità*, che devono essere utilizzati al posto dei corrispondenti caratteri:
 - *&*; definisce il carattere &
 - *<*; definisce il carattere <
 - *>*; definisce il carattere >
 - *"*; definisce il carattere “
 - *'*; definisce il carattere ‘

Le regole *well-formed* non pongono nessun vincolo sui tag che possono essere utilizzati in un determinato documento. In altre parole abbiamo bisogno di definire una grammatica per il linguaggio basato su tag che il progettista *inventa* per un dato documento. Tale grammatica dovrà indicare quali sono i vocaboli che possiamo utilizzare (tag) e con che

struttura possiamo comporre le nostre frasi. Ci sono vari meccanismi per definire questa grammatica, in particolare i DTD (*Document Type Definition*) e gli schemi XML.

Il DTD, attraverso un linguaggio XML-based, descrive i tag utilizzabili, la loro reciproca relazione e altre informazioni sugli attributi. Un DTD può essere contenuto internamente allo stesso documento XML, o creato esternamente come documento a se stante, ed allegato al documento in oggetto. La sintassi da utilizzare per la realizzazione dei DTD non segue le stesse regole dei documenti XML ed inoltre la grammatica che si riesce a definire non sempre è soddisfacente.

Gli schemi XML consentono un vero controllo degli errori, in modo che i record strutturati XML possano essere analizzati e verificati non solo per controllare se sono ben formati, ma anche se sono internamente validi. Sfortunatamente le specifiche degli schemi XML sono abbastanza difficoltose, quindi ci limiteremo a dire che la definizione di un elemento avviene tramite la seguente sintassi:

```
<xs:complexType name="nome tipo">
...
</xs:complexType>
```

e che possiamo definire i componenti di uno schema dichiarati al livello massimo, cioè allo stesso livello dell'elemento root, come schemi a livello globale e quindi utilizzabili nel resto dello schema; quando dichiariamo un insieme di variabili globali in un programma abbiamo definito un *namespace*. Il meccanismo dei namespace consente di definire la struttura di un documento XML basandosi su diversi schemi.

4. Presentazione di documenti XML tramite CSS

CSS è un linguaggio che consente di controllare la formattazione di un documento strutturato tramite tag, sia esso in HTML che in XML. Esso consente di definire caratteristiche come la dimensione ed il tipo dei caratteri, il colore dello sfondo e così via, basandosi su regole la cui sintassi rispetta il seguente schema:

```
selettore { proprietà: valore; proprietà: valore ... }
```

Il *selettore* specifica a quale elemento la regola deve essere applicata, mentre la parte racchiusa tra parentesi graffe elenca le caratteristiche da impostare ed il relativo valore. Si riporta di seguito, un esempio di foglio di stile utilizzato:

```
/* Foglio di stile CSS predefinito per un nuovo progetto
BODY
{ BACKGROUND-COLOR: white;
  FONT-FAMILY: Ariel;
  FONT-SIZE: 7pt;
  FONT-WEIGHT: normal;
```

```

    LETTER-SPACING: normal;
    TEXT-TRANSFORM: none;
    WORD-SPACING: normal}
.txt7n
{ font-size: 7pt;
  font-family:Arial;
  font-style:normal;
  line-height:90%;
  margin:0; }
.txt7n2
{ font-size: 7pt;
  font-family:Arial;
  font-weight:bold;
  font-style:normal;
  line-height:90%;
  margin:0; }

```

Generalmente un foglio di stile CSS da applicare ad un documento XML viene salvato in un file di testo con estensione .css (anche se l'estensione in realtà è irrilevante). Nel documento XML possiamo quindi inserire un riferimento ad esso nella parte iniziale del documento stesso, come nel seguente esempio:

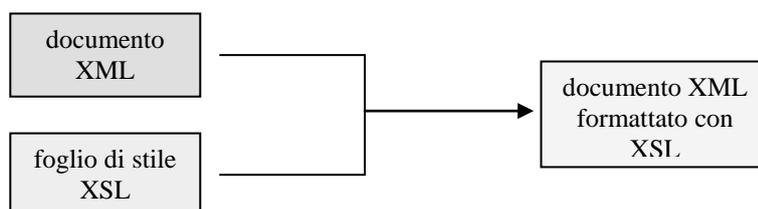
```

<?xml version="1.0" ?>
<?xml - stylesheet type="text/css" href="stilestititistico.css" ?>

```

5. Presentazione di documenti XML tramite XSL

L'XML affronta da un punto di vista semantico un documento, fornendo la struttura e il contesto per i dati che contiene, senza far riferimento all'aspetto dei dati nel momento in cui verranno visualizzati. A questo scopo è stato creato, quindi, il linguaggio XSL (*Extensible Stylesheet Language*) che si basa sul meccanismo dei fogli di stile. In generale i fogli di stile permettono di definire delle classi di formattazione applicabili ad un dato documento.



Diversamente da CSS, che utilizza il documento stesso come struttura di base a cui aggiungere la formattazione, XSL è formato da tre componenti di base: XSL vero e proprio, XSLT, cioè XSL *Transformations*, che permette di convertire un documento XML in un altro formato e Xpath, un linguaggio progettato per indicizzare le parti di un documento XML.

In questo modo, invece di imparare un linguaggio completamente nuovo per imporre uno stile ad una pagina, XSL appare simile ai tag di XML o HTML.

L'elaborazione ed il flusso di controllo avvengo attraverso alcuni elementi ben distinti:

- **caricamento di parametri**

```
<xsl:value - of select"nome elemento">
</xsl:value - of>
```

- **creazione ed assegnazioni di variabili**

```
<xsl:variable name="nome variabile" select="valore variabile">
</xsl:variable>
```

- **ciclo**

```
<xsl:for - each select="nome variabile" >
    contenuto
</xsl:for - each>
```

- **test di controllo semplice**

```
<xsl:if test="valore Booleano">
    <xsl:template>
        contenuto
    </xsl:template>
</xsl:if>
```

- **test di controllo multiplo**

```
<xsl:choose>
    <xsl:when test ="valore Booleano">
        <xsl:template>
            contenuto
        </xsl:template>
    </xsl:when>
    <xsl:when test ="valore Booleano">
        <xsl:template>
            contenuto
        </xsl:template>
    </xsl:when>
    <xsl:otherwise>
        <xsl:template>
            contenuto
        </xsl:template>
    </xsl:otherwise>
</xsl:choose>
```

Attraverso questi semplici costrutti si può meglio capire come generare un foglio di stile

XSL da applicare ad un documento XML, come riportato di seguito:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
<xsl:output method="html"/>
<xsl:template match="/">
<html>
<head>
<title><xsl:value-of select="//NewDataSet/sommario/titolo"/></title>
<Link href=http://bms.istat.it/styles2.css type="text/css" real="stylesheet" />
</head>
<body>
<xsl:variable name="periodoFin" select="//NewDataSet/varie/@periodo_fin"/>
<xsl:variable name="annoFin" select="//NewDataSet/varie/@anno_fin"/>
<xsl:variable name="lg" select="//NewDataSet/varie/@lg"/>
<xsl:variable name="varP" select="//NewDataSet/varie/@varP"/>
<!-- tabella di intestazione -->
<table border="0" cellpadding="0" cellspacing="0" bordercolor="red" width="700">
```

```

        <tr>
            <td class="txt7n3" align="center">
                <xsl:value-of select="//NewDataSet/sommario/dominio" />
            </td>
        </tr>
        <tr align="left" >
            <td class="txt">
                <xsl:value-of select="//NewDataSet/sommario/titolo"/>
            </td>
        </tr>
    </table>
    <xsl:for-each select="//NewDataSet/pagina">
        <!-- tabelle generali -->
        <table cellspacing="0" cellpadding="0" border="0" bordercolor="blue" width="600">
            <xsl:for-each select="sottopagina">
                <xsl:variable name="numserie" select="@numSerie"/>
                <tr>
                    <td valign="top">
                        <!-- prima sottotavola -->
                        <table align="center" width="100%" border="0" bordercolor="green"
                            cellspacing="0" cellpadding="0">
                            <!-- disegno tutte le righe -->
                            <tr>
                                <td height="1" colspan="{ $numserie+1}" style="FONT-SIZE:
                                0px;">&#160;</td>
                                </tr>
                                <tr>
                                    <td rowspan="4" width="1%">&#160;</td>
                                    <xsl:for-each select="livello1">
                                        <xsl:variable name="span" select="@span"/>
                                        <xsl:choose>
                                            <xsl:when test="$span = 1">
                                                <td class="txt7n" valign="bottom" align="center"
                                                colspan="{ $span}">
                                                    <table width="98%"><tr><td><xsl:value-of
                                                    select="@titolo"/></td></tr>
                                                    </table>
                                                </td>
                                            </xsl:when>
                                            <xsl:otherwise>
                                                <td class="txt7n" valign="bottom" align="center"
                                                colspan="{ $span}">
                                                    <table border="0" bordercolor="red" width="98%">
                                                        <tr align="center">
                                                            <td class="txt7n">
                                                                <xsl:value-of select="@titolo"/>
                                                            </td>
                                                        </tr>
                                                    </table>
                                                </td>
                                            </xsl:otherwise>
                                        </xsl:choose>
                                    </xsl:for-each>
                                </tr>
                            <!-- seconda riga della testa -->
                            <tr><xsl:for-each select="livello1/livello2">
                                <xsl:variable name="span" select="@span"/>
                                <xsl:variable name="titolo" select="@titolo"/>
                                <xsl:choose>
                                    <xsl:when test="string-length($titolo)=0">
                                        <td class="txt7n" align="center" colspan="{ $span}">
                                            <table>
                                                <tr align="center">
                                                    <td>
                                                        <xsl:value-of select="@titolo"/>
                                                    </td>
                                                </tr>
                                            </table>
                                        </td>
                                    </xsl:when>
                                </tr>
                            </tr>
                        </td>
                    </tr>
                </table>
            </xsl:for-each>
        </table>
    </table>

```

```
        <xsl:otherwise>
        <td class="txt7n" valign="bottom" align="center"
colspan="{\$span}" >
        <table><tr><td class="txt7n"><xsl:value-of
select="@titolo"/></td></tr>
        </table>
        </td>
<!-- CONTINUA →
```

6. La formattazione del documento XML attraverso XSL-FO

Le proprietà di formattazione XSL sono applicate agli oggetti di formattazione, ovvero visualizzazioni astratte del documento che definiscono determinate aree a cui le proprietà possono essere applicate. Queste proprietà si identificano attraverso il tag *<fo:nome oggetto>* e si applicano per:

- impaginazione e layout;
- blocchi;
- documenti in linea;
- tabelle;
- elenchi;
- link e formattazioni multiple;
- documenti fuori linea.