

L'impaginazione automatica di tavole statistiche

Claudia Albergamo (*)

Piero Crivelli (**)

(*) ISTAT - Direzione centrale per le indagini sulle istituzioni sociali

(**) ISTAT - Servizio Produzione editoriale

Introduzione¹

Il presente lavoro prende lo spunto dalle recenti esperienze di decentramento presso i servizi statistici delle funzioni di impaginazione delle tavole statistiche per le pubblicazioni correnti.

Gli autori hanno preso parte direttamente a quelle esperienze sia pure con ruoli diversi, l'uno facente parte del Dipartimento Diffusione che ha promosso l'innovazione, l'altra in forza presso una delle Direzioni Tecniche e quindi da utente che le ha recepite.

Fino a due anni fa il modo di produrre le pubblicazioni a stampa in Istat era basato su una elaborazione centralizzata. Il processo produttivo doveva essere modificato al fine di svincolarlo per tempo dall'elaboratore centrale, in via di dismissione. Inoltre ci si rendeva conto che presso i servizi tecnici le capacità di utilizzo degli strumenti di office automation per la formattazione di documenti testuali e di tabelle erano di buon livello. Si aggiunga che i passaggi di correzione delle bozze impaginate presso tipografie esterne costituivano una tra le principali componenti del tempo complessivo di stampa.

L'idea logicamente conseguente fu allora quella di semplificare drasticamente il processo fornendo supporto organizzativo ai servizi statistici affinché questi producessero direttamente, utilizzando PC e workstation, gli input cartacei per le tipografie, tagliando in tal modo tempi e costi. Gli originali forniti alle tipografie potevano essere in tal modo fotoriprodotti anziché fotocomposti, e quindi stampati.

Si decise di limitare il campo di intervento alle collane di pubblicazioni correnti *Annuari e Informazioni*.

Ovviamente non è stato tutto così semplice e, tra gli apprezzamenti per l'iniziativa messa in atto, si sono avute anche critiche: alcuni lamentavano difficoltà ad operare con strumenti nuovi, una assistenza discontinua, la sensazione di doversi improvvisare in un mestiere complesso, l'impaginazione, che in fondo spettava ad altri; dall'altra parte le persone del Dipartimento preposte alla funzione redazionale lamentavano un aggravio di lavoro nel controllo degli originali, che non sempre rispettavano gli standard editoriali in uso.

Altre critiche più tecniche: mancanza di automatismo e proceduralità da parte di strumenti come Excel adatti per esigenze di informatica individuale e non per elaborare massicce quantità di dati.

Tuttavia l'innovazione di processo è andata avanti fino ad includere la quasi totalità delle pubblicazioni correnti stampate negli ultimi due anni. Attualmente sono state stampate tutte le prime edizioni i cui originali sono stati realizzati direttamente presso i servizi e per alcune

¹ Del presente lavoro i capitoli 1, 2, 3 e 5 sono stati sviluppati da Piero Crivelli, il capitolo 4 da Claudia Albergamo mentre l'introduzione ed il capitolo 6 sono stati sviluppati in comune dai due autori.

pubblicazioni si è giunti alla seconda edizione. È quindi il momento di verificare i risultati della innovazione di processo perché nelle successive elaborazioni dei volumi si potrà riutilizzare in gran parte il lavoro già fatto.

Un fatto inaspettato nel processo in atto è consistito nella diversa rispondenza dei colleghi coinvolti. Mentre alcuni hanno subito il lavoro di impaginazione altri hanno fornito dei contributi spontanei, anche di rilievo, per migliorare il modo di lavorare ed il prodotto finale. E' questo il caso del programma di impaginazione automatica che viene qui illustrato.

Parallelamente alle iniziative innovative riguardanti le pubblicazioni su carta permanente, ed è sempre più pressante, l'esigenza di automatizzare il processo produttivo per la realizzazione dei prodotti editoriali su supporto magnetico, in particolare su CD ROM, ed elettronico on line. Anche su queste modalità si sono sviluppate esperienze interessanti di cui si parla più avanti.

Per tutte e tre le tipologie di prodotti si avverte la necessità di standardizzazione e di semplificazione. In particolare per le pubblicazioni su supporto magnetico occorre stabilizzare le funzionalità disponibili: visualizzazione, esportazione e stampa: inoltre la grafica delle schermate di presentazione e navigazione dovrebbe essere tale da dare ai prodotti caratteristiche di "riconoscibilità" al pari delle pubblicazioni su carta.

Dal punto di vista del processo produttivo, invece, è molto importante che più prodotti editoriali diversi possano essere realizzati indifferentemente a partire da un **unico input**. A tal fine i file di input dovranno contenere, oltre alle informazioni statistiche, anche altre "informazioni" gestibili dalle diverse applicazioni che realizzano i prodotti.

1. Il processo di produzione e i prodotti editoriali

Il processo di produzione dell'informazione statistica si compone, come è noto, della fase di raccolta dei dati, secondo le tecniche ritenute più adatte all'indagine, della fase di correzione, aggregazione e sintesi ed infine della fase di diffusione. Ognuna delle tre fasi è strettamente correlata con la successiva. Infatti la possibilità di realizzare determinate elaborazioni è subordinata al tipo di informazioni raccolte ovvero ad una opportuna progettazione del questionario. Le stesse elaborazioni possono risultare più o meno complesse in base alla tipologia di file utilizzati per memorizzare i dati e alla loro organizzazione all'interno del file stesso. La realizzazione di prodotti editoriali, ultimo step del processo produttivo, dipende anch'essa strettamente dalla precedente fase di aggregazione e sintesi dei dati elementari. Infatti, le diverse modalità di diffusione possono essere rese più agevoli da una adeguata predisposizione degli output della fase precedente. E' quest'ultimo aspetto che viene qui preso in esame.

Mentre fino ad un decennio fa l'informazione statistica era costituita soltanto da pubblicazioni su carta, la rapida diffusione dei supporti magnetici e il dirompente impatto di Internet hanno introdotto nuove modalità di diffusione, ciascuna con le sue problematiche. I dati necessitano quindi di un trattamento informatico diverso a seconda che si sia scelto di realizzare una pubblicazione cartacea, un prodotto su supporto magnetico (dischetti o CD ROM) oppure un'applicazione per la diffusione su reti locali o geografiche.

Le tre tipologie di prodotti editoriali presentano problematiche realizzative molto diverse una dall'altra: nella stampa su carta, ad esempio, si deve tenere conto delle linee per la delimitazione della testata, del corpo centrale e delle note, della numerazione di pagina, mentre per i prodotti su supporto magnetico può essere opportuno l'uso dei colori per le stesse finalità, l'utilizzo di apposite finestre dello schermo, di riferimenti ipertestuali e così via.

Un'esigenza comune a tutti gli Istituti di statistica è certamente quella di poter diffondere l'informazione in una o anche tutte le modalità suddette in relazione alle richieste dell'utenza e alle strategie di diffusione definite. Inoltre, non è infrequente che le decisioni circa la realizzazione di determinati prodotti editoriali vengano cambiate in corso d'opera aggiungendo, ad esempio, un CD ROM non previsto inizialmente ad un volume o decidendo di rendere disponibile in rete un insieme di tavole statistiche inizialmente previste soltanto su supporto cartaceo.

Sviluppare **modalità automatiche** di impaginazione e formattazione può essere vantaggioso sia nei casi citati di una pianificazione editoriale non rigida sia per conseguire una maggiore tempestività nel rilascio delle pubblicazioni.

E' peraltro vero che non tutte le pubblicazioni cartacee si prestano ad un discorso di impaginazione automatica perché l'insieme dei prodotti editoriali risponde ad obiettivi diversi. Ci sono pubblicazioni la cui vocazione principale è la tempestività, al fine di rendere l'informazione alla collettività il più presto possibile (si pensi ai dati congiunturali), pubblicazioni di analisi ed approfondimento, meno immediate dal punto di vista del rilascio, che riportano commenti ed indicatori di sintesi, per le quali uno stile editoriale curato, affidato a tipografie esterne o ad uffici specializzati interni, garantisce una adeguata qualità del prodotto finale e pubblicazioni di tipo promozionale per le quali è prioritario un aspetto accattivante, ricco di grafica e di colori, anche per le quali si rende necessario l'intervento di professionalità grafiche specifiche.

Del primo gruppo fanno certamente parte, considerando le pubblicazioni Istat, le collane *Annuari* ed *Informazioni*, che insieme assommano al 70% delle pubblicazioni stampate ogni anno. Tali pubblicazioni sono costituite da una parte testuale, solitamente breve, consistente in una introduzione, nella descrizione dell'indagine, nel commento dei risultati e nelle note metodologiche, e da una parte più ampia, contenente le tavole statistiche.

Per le ragioni appena esposte, dal momento che l'attività di diffusione dell'informazione statistica attraverso pubblicazioni cartacee o elettroniche è certamente parte del *core business* per un istituto di statistica come l'Istat, allora vale la pena di investire sia nella individuazione di prodotti software di mercato che nello sviluppo di software *ad hoc* e routine di supporto che si adattino alle specificità dei processi produttivi dell'ente e che assicurino un livello adeguato di **automatizzazione** e di **proceduralità** nella realizzazione di prodotti di diffusione.

2. Negli altri INS...

Al fine di effettuare una verifica ed una comparazione del percorso di cambiamento intrapreso per quanto riguarda l'impaginazione di alcune pubblicazioni presso i servizi statistici, è stato indirizzato nello scorso mese di settembre '99 un breve questionario di quattro domande ai colleghi che svolgono funzioni di interfaccia tra Eurostat e gli Istituti Nazionali di Statistica (INS) utilizzando il canale degli sportelli Eurostat presso gli INS, essendo tale canale sufficientemente collaudato. Ai destinatari è stato chiesto di ribaltare le domande ai loro colleghi presso le strutture produttive e/o di diffusione.

Le domande:

- Quali formati dei dati sono utilizzati per la realizzazione di pubblicazioni?
- Chi predispose il layout delle pubblicazioni (editori esterni, uffici specializzati interni, ...)?
- Le strutture che producono i dati realizzano direttamente le tavole o forniscono file sequenziali di dati disaggregati?
- È più importante la tempestività o un aspetto grafico curato ed accattivante?

Le risposte ottenute da parte di cinque INS (Svezia, Finlandia, Danimarca, Germania e Spagna) sono riportate integralmente in appendice. Da queste si evince che c'è sufficiente omogeneità tra noi e loro nel modo di trattare i dati e realizzare i prodotti di diffusione: i software Word ed Excel sono i più utilizzati per la formattazione delle tavole; inoltre, tutti hanno sottolineato l'esigenza di tempestività nella diffusione anche a costo di una essenzialità nel layout delle pubblicazioni.

Era invece meno scontato che le tavole statistiche venissero realizzate direttamente presso le strutture produttive in un formato *ready to print* o presso uffici specializzati interni. Ciò è confortante rispetto alla direzione di innovazione da noi intrapresa.

3. L'esperienza Istat

Come già detto in Istat è stata realizzata tra il 1997 e il 1999 una innovazione di processo consistita nel trasferire competenze e capacità di impaginazione dei volumi ai servizi produttori di statistiche per quanto riguarda le pubblicazioni numericamente più consistenti.

In particolare, il cambiamento ha interessato le collane *Annuari e Informazioni*.

In precedenza gli originali delle pubblicazioni venivano realizzati con una procedura denominata "512" (per la lunghezza massima dei record variabili prodotti), attiva sull'elaboratore centrale. Le decisioni di modifica dell'architettura informatica dell'Istituto, che prevedevano la dismissione degli elaboratori centrali in un'ottica di *downsizing*, hanno di fatto reso necessario l'abbandono di tale procedura e hanno determinato l'inizio della fase di innovazione di cui parliamo.

Nell'ambito della innovazione di processo sono stati tenuti dei corsi interni di formazione a cui hanno partecipato circa 100 persone delle due Direzioni Tecniche.

A supporto dell'innovazione è stato attivato un sito Intranet (*home intranet* → *attività e progetti* → *pubblicazioni*) da cui è possibile prelevare programmi di conversione, fogli di stile e indicazioni di standard disponibili per l'impaginazione dei volumi; tra questi anche un programma per consentire la migrazione dal formato "512" al formato Excel.

Al termine del processo di innovazione sono stati raggiunti i seguenti risultati:

- è stata fornita una soluzione praticabile alla esigenza stringente di superamento della procedura cosiddetta 512;
- i Servizi mantengono il controllo totale dei dati fino al formato finale delle tavole pronte per la fotocoproduzione disponendo su PC delle stesse tavole anche per altre rielaborazioni;
- la scelta dei software per l'impaginazione, Word per i testi ed Excel per le tavole statistiche, essendo questi già largamente utilizzati nell'Istituto per produrre la documentazione corrente, ha comportato un investimento formativo minimo per il personale coinvolto ed un costo trascurabile;
- la elaborazione dei volumi fino alla copia pronta per la fotocoproduzione ha di fatto eliminato la lunga fase di correzione delle bozze; il ruolo del Dipartimento Diffusione si è in tal modo concentrato sulla funzione di supporto ed istruzione ai Servizi tecnici e sul controllo del rispetto degli standard di impaginazione opportunamente ridefiniti;
- con la realizzazione di un sito Intranet contenente le indicazioni di standard ed il materiale informatico di ausilio per l'impaginazione è stato definito un luogo virtuale condiviso in cui far confluire il *know how* necessario. Il sito è da considerare come una opportunità per

le persone interessate a lavorare in un modo diverso, meno burocratizzato e più cooperativo;

- i tempi di stampa dei volumi si sono ridotti notevolmente, grazie all'eliminazione di tutti gli oneri legati alla produzione di bozze e alla loro correzione.

I risultati ottenuti possono ritenersi soddisfacenti. Ad una analisi di massima effettuata sulle edizioni pubblicate tra il 1998 ed il primo quadrimestre 1999 (tabella 1) si evidenzia come il numero di volumi fotocomposti in tipografia si sia notevolmente ridotto, fin quasi a zero, mentre è molto aumentato il numero di volumi curati presso i Servizi, indice di un livello alto di partecipazione al progetto, anche se la partecipazione non è stata sempre spontanea ma a volte è derivata da precise indicazioni dei direttori di area.

Tabella 1 - Numero di pubblicazioni edite per quadrimestre dal 1/1998 al 1/1999

Quadrimestri	Volumi pubblicati		Tempo medio di pubblicazione (giorni)	
	Originali fotocomposti	Originali fotoriprodotti	Originali fotocomposti	Originali fotoriprodotti
1/98	8	6	163	42
2/98	11	17	143	52
3/98	7	22	113	69
1/99	2	20	165	37

Il tempo medio di elaborazione di una pubblicazione è stato decisamente più breve nel caso di fotoriproduzione, anche perché questa modalità include la correzione delle bozze.

Il costo di pre-stampa è risultato complessivamente più basso che in precedenza dal momento che il costo medio del supporto in pellicola di una pagina fotocomposta in tipografia è pari a circa 13.000 lire, secondo i contratti attualmente in vigore, contro le 5.700 lire di una pagina fotoriprodotta.

I limiti che si sono evidenziati nella innovazione di processo:

- l'utilizzo di Excel non vincola rigidamente l'utente ad osservare gli standard definiti. Nel caso in cui gli originali presentino evidenti difformità dagli standard prefissati ciò rende più oneroso il lavoro della U.O. Pubblicazioni che effettua l'attività di controllo redazionale e di supervisione del risultato finale;
- nella sua modalità di utilizzo più immediata Excel consente di trattare una tavola per volta e quindi obbliga ad applicare in maniera ripetitiva l'insieme dei comandi di formattazione;

- se è vero che il tempo di pubblicazione si stabilizza è peraltro vero che molta della complessità del lavoro editoriale viene trasferito sui colleghi dei servizi, i quali talvolta lamentano un eccessivo aggravio di lavoro;
- la mancanza di una effettiva proceduralità, non consente di costruire tutte le tavole previste in un processo "continuo", senza manualità e quindi senza la possibilità di introdurre errori dall'esterno, a meno di "blindare" Excel nell'ambito di un'applicazione guidata e controllata da programma.

Quello che è stato fatto quindi non è ancora sufficiente: occorre introdurre delle modalità di impaginazione che rendano il più possibile automatico il processo di realizzazione di prodotti editoriali e pensare i diversi prodotti in termini di **insieme integrato** nell'ambito di un unico processo.

4. Un programma di impaginazione automatica per le pubblicazioni cartacee

Il programma di seguito descritto è nato come contributo spontaneo per automatizzare il processo di impaginazione delle tavole statistiche relativamente alle pubblicazioni cartacee.

Il software, realizzato in Visual Basic per Excel, è stato sviluppato da C. Albergamo presso il Servizio Formazione, Istruzione e Lavoro. Questo utilizza i file ASCII ottenuti dalle elaborazioni dei dati e produce tavole in formato Excel, realizzate secondo gli standard Istat (standard grafici, di formattazione e di impaginazione), pronte per la stampa.

I principi su cui si basa sono:

- l'utilizzo di caratteri di controllo per determinare l'aspetto finale delle righe della tavola;
- la scomposizione della tavola in più elementi indipendenti tra loro: il titolo, la testata, il corpo della tavola (composto di fiancata e dati) e le note;
- la gestione parametrizzata ed esterna delle informazioni relative alle singole tavole.

Il programma ha come input due file, il primo contenente i dati, il secondo contenente una serie di oggetti e informazioni che vengono utilizzate per l'impaginazione.

Il file contenente i dati deve essere un file ASCII e dovrà avere le caratteristiche seguenti:

- virgola come separatore dei decimali;
- colonne separate da *pipe* "|";
- caratteri di controllo in fiancata.

Ogni carattere di controllo indica al programma le caratteristiche grafiche che deve avere la riga in cui si trova. Il programma riconosce i simboli, li interpreta e li elimina dalla fiancata. Nella figura 1 sono

stati elencati tutti i caratteri di controllo che il software è in grado di riconoscere con le relative funzionalità associate².

Figura 1 - Caratteri di controllo utilizzati dal programma con le relative funzionalità associate

“\$”	allinea il testo in fiancata con i numeri a cui si riferisce
“<N>”	imposta la riga in neretto
“<I>”	imposta la riga in italico
“<X>”	gestisce i rientri
“<D>”	elimina la riga
“<C>”	centra il testo
“<R>”	aggiunge una riga bianca
“<P>”	interruzione di pagina

Del file di input viene utilizzata la parte contenente i dati e la fiancata; viene invece esclusa la testata che deve essere predisposta a parte. La testata di una tavola infatti è un oggetto complesso in cui i livelli di raggruppamento, generalmente più d'uno, si collocano spesso a cavallo delle relative colonne di dati; inoltre il testo richiede la suddivisione su più righe. Tra i diversi raggruppamenti vengono tracciati dei fili che individuano i gruppi di colonne sottostanti per cui l'importazione automatica non è possibile o non è conveniente. Un esempio di un tale file di input è riportato in figura 2.

Il programma è in grado di gestire qualsiasi file che rispetti le specifiche descritte, indipendentemente dal modo in cui è stato prodotto. Tuttavia è stata rivolta una particolare attenzione alle elaborazioni realizzate utilizzando il SAS, essendo questo lo strumento maggiormente utilizzato presso l'Istat per le analisi statistiche. Si è cercato, quindi, di risolvere le problematiche tipiche della tabulazione SAS. Chiunque si sia trovato a dover realizzare tavole in SAS presso l'Istat si sarà sicuramente scontrato con il problema dei “di cui”: è infatti uso comune presso l'Istituto evidenziare solo alcune delle modalità che concorrono a formare un totale e, dal momento che il SAS elabora comunque tutte le modalità di una variabile, questo problema è stato risolto permettendo in fase di impaginazione di eliminare righe e colonne; per eliminare le righe basta far precedere il testo in fiancata dal simbolo “<D>” (come descritto in fig. 1), mentre come eliminare le colonne sarà descritto in seguito.

² Se è necessario utilizzare più di un carattere di controllo sulla stessa riga, questi dovranno comparire nell'ordine in cui sono stati elencati in figura 1

Figura 2 - File ottenuto da una elaborazione SAS

	studia			non studia			NON LAVORANO			
		Non	Non		Non					
		lavo-	lavo-		stu-					
		rano	rano		diano					
		non	non							
	Cerc-	cerc-	Tota-		Tota-					
	ano	ano	le		le					
	Lavo-	lavo-	lavo-			stud-	stud-	stud-	stud-	
	rano	ro	ro			ia	ia	ia	ia	
<C>ITALIA										
TIPI DI SCUOLA										
<N>Totale scuole superiori	17,8	22,9	59,1	100,0	100,0	100,0	69,3	30,6	18,5	81,4
Istituti Professionali (b)	31,4	32,4	36,0	100,0	100,0	100,0	31,8	68,1	7,3	92,6
Istituti Tecnici	21,9	25,8	52,2	100,0	100,0	100,0	55,3	44,6	12,1	87,8
Licei	13,7	18,5	67,7	100,0	100,0	100,0	93,3	6,6	65,0	34,9
Istituti Magistrali	20,6	34,0	45,2	100,0	100,0	100,0	57,5	42,4	27,8	72,1
Altri tipi di scuole	20,1	27,9	51,9	100,0	100,0	100,0	59,3	40,6	22,4	77,5
<R>RIPARTIZIONI										
<N>Italia	17,8	22,9	59,1	100,0	100,0	100,0	69,3	30,6	18,5	81,4
Nord-Ovest	21,7	13,0	65,2	100,0	100,0	100,0	80,0	19,9	17,7	82,2
Nord-Est	24,4	11,3	64,2	100,0	100,0	100,0	81,0	18,9	18,2	81,7
Centro	19,4	21,4	59,0	100,0	100,0	100,0	72,8	27,1	19,8	80,1
Sud	13,0	32,7	54,1	100,0	100,0	100,0	61,1	38,8	18,8	81,1
Isole	11,5	34,6	53,8	100,0	100,0	100,0	61,6	38,3	17,4	82,5

La possibilità di eliminare righe e colonne non risolve esclusivamente il problema dei “di cui”. In figura 3 è riportato un esempio in cui la possibilità di eliminare alcune colonne è stata di fondamentale importanza: confrontandola con la figura 2 si può notare che di tutte le colonne ottenute dall’elaborazione SAS, solo cinque erano di interesse. Nel caso in esame la presenza delle colonne in più è dovuta al fatto che il rapporto nell’ultima colonna ha come denominatore un insieme, coloro che lavorano, che include anche una fetta di popolazione non inclusa nel resto della tavola, coloro che non studiano. Lavorando con la procedura *Tabulate* del SAS questo non è consentito.

Figura 3 - Risultato dell'impaginazione della tavola a figura 2

Tavola 2 - Diplomati del 1995 che nel 1998 studiano, per condizione occupazionale e per tipo di scuola e ripartizione geografica_Maschi e femmine (composizione percentuale)

RIPARTIZIONI GEOGRAFICHE (a) E TIPI DI SCUOLA	NON LAVORANO			Totale	Studenti che lavorano (per cento occupati)
	Lavorano	Cercano lavoro	Non cercano lavoro		
ITALIA					
TIPI DI SCUOLA					
Totale scuole superiori	18	23	59	69	19
Istituti Professionali (b)	31	32	36	32	7
Istituti Tecnici	22	26	52	55	12
Licei	14	19	68	93	65
Istituti Magistrali	21	34	45	58	28
Altri tipi di scuole	20	28	52	59	22
RIPARTIZIONI					
Italia	18	23	59	69	19
Nord-Ovest	22	13	65	80	18
Nord-Est	24	11	64	81	18
Centro	19	21	59	73	20
Sud	13	33	54	61	19
Isole	12	35	54	62	17

(a) Le ripartizioni si riferiscono all'ubicazione della scuola.
(b) Sono incluse le scuole magistrali.

Un'altra caratteristica del programma che evidenzia la specificità del prodotto rispetto alla tabulazione SAS è la funzionalità associata al carattere di controllo “\$”. Questo simbolo, infatti, indica al programma che la descrizione in fiancata si riferisce ai dati che si trovano nella riga successiva: è questo il caso dei subtotali ottenuti con la procedura *Tabulate* del SAS. In figura 4 sono rappresentate tre righe di una tavola ottenuta con la stessa procedura in cui la dicitura “Istituti professionali” rappresenta un sottotale e si riferisce ai dati che seguono. La presenza del carattere “\$” fa sì che il programma allinei i dati al testo.

Figura 4 - File ottenuto da una elaborazione SAS

<N>Totale scuole superiori	31993	31993	25707	118089	143796
\$<N>Istituti Professionali (b)					
	5868	5868	7516	28097	35613
Industriali	2158	2158	2158	11081	13239
Servizi comm., turist., pubbl.	1932	1932	2586	11543	14129

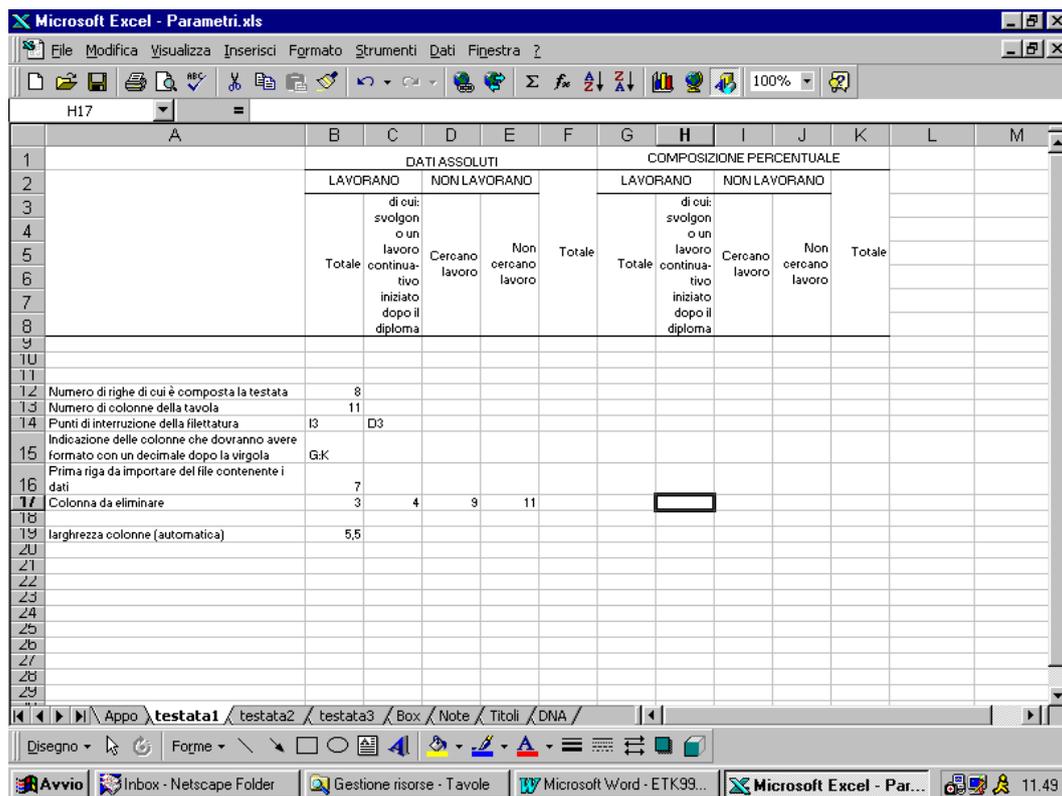
Il file contenente le testate, i titoli e le note è un una cartella di lavoro Excel composta di alcuni fogli. Il primo foglio, denominato “Appo” è un foglio di appoggio in cui il programma compone le diverse parti della tavola, prima di salvarla come file esterno. I fogli successivi contengono le testate, sono denominati “testata1”, “testata2”, e così via. All’interno di ognuno di

questi fogli, oltre alla testata si trovano alcuni parametri che forniscono ulteriori informazioni relative alla testata stessa.

Come si può vedere dalla figura 5, per ogni testata, in alcune celle a partire dalla riga 12, vengono indicati:

- il numero di righe di cui è composta la testata: la testata, infatti secondo gli standard dell'Istituto è realizzata su più righe tutte della stessa altezza. Il programma imposta automaticamente l'altezza delle righe sia per il corpo della tavola che per la testata;
- il numero di colonne della tavola, inclusa la fiancata: questo parametro viene utilizzato per gestire in automatico la larghezza delle colonne, in relazione alla larghezza totale della tavola, con possibilità di variare la larghezza della prima colonna;
- punti di interruzione della filettatura: il programma aggiunge una piccola colonna per interrompere la filettatura. Se sono presenti più punti di interruzione della filettatura questi devono essere indicati a partire da destra verso sinistra;
- indicazione delle colonne che dovranno avere formato con un decimale dopo la virgola: vengono indicate le colonne contenenti valori percentuali;
- prima riga da importare del file contenente i dati: questo parametro indica al programma di quante righe è composta la testata nel file contenente i dati. Infatti questa parte del file viene esclusa fin dalla fase di importazione dei dati in Excel;
- colonne da eliminare: i vantaggi legati alla possibilità di eliminare alcune colonne sono già stati commentati;
- larghezza delle colonne: questo parametro viene utilizzato per impostare la larghezza di tutte le colonne esclusa la prima. E' possibile inserire direttamente il valore nella cella oppure farlo calcolare al programma. Infatti, in fase di impaginazione, se la cella è vuota, il programma calcola automaticamente il valore e lo memorizza nella cella stessa.

Figura 5 - Foglio contenente una testata con relativi parametri, utilizzabile dal programma



A seguire, dopo i fogli contenenti le testate, si trovano altri quattro fogli: nel primo, denominato “BOX”, è collocato l’elenco delle descrizioni della fiancata, ovvero i diversi testi che dovranno comparire nell’angolo in alto a sinistra delle tavole (nell’esempio in figura 3 “ RIPARTIZIONI GEOGRAFICHE (a) E TIPI DI SCUOLA”). Il secondo, denominato “Note” contiene alcune caselle di testo in cui si trovano le note, le caselle di testo si chiamano “NOTA1”, “NOTA2”, ecc. Il terzo, organizzato come il precedente, contiene i titoli, le caselle di testo contenenti i titoli si chiamano “Titolo1”, “Titolo2” ecc. In fine nel foglio “DNA” si trovano elencate tutte le tavole sulle quali si sta lavorando (ad esempio un capitolo di un volume) e ad ognuna di esse viene associato un box , una nota e un titolo. Ad esempio, guardando la figura 6 si capisce che per realizzare la tavola 2.1 viene utilizzato come box il testo che si trova nella prima riga del foglio “Box”, nella tavola non compaiono note, viene utilizzata la prima testata (“testata1”) e la casella di testo “Titolo1” per il titolo. E’ sempre in questo foglio che viene indicato al programma il nome del file contenente i dati e la dimensione della prima colonna.

La realizzazione di questo file di input è piuttosto semplice, in quanto, insieme al programma , viene fornito un modello contenente alcuni fogli predisposti per contenere le testate e i fogli "Appo", "Box", "Note", "Titoli" e "DNA", già nominati. Dove necessario, all'interno dei fogli si trovano alcune caselle di testo già opportunamente nominate. È sufficiente, quindi predisporre le testate negli appositi

fogli³, e i relativi parametri seguendo quanto indicato nel foglio stesso; scrivere i titoli e le note nelle apposite caselle di testo, i box nella colonna A del foglio "Box" e compilare il foglio "DNA". All'interno di questo foglio l'inserimento dei parametri è guidato con note poste in opportune celle, visualizzabili al passaggio del mouse, come si può vedere dalla figura 6.

Dopo aver predisposto i file di input è necessario indicare al programma dove si trovano e il loro nome: dalla schermata principale (figura 7), visualizzata all'avvio del programma, facendo clic sul bottone "imposta / modifica parametri" si apre una maschera (figura 8) in cui viene chiesto all'utente di indicare la directory di input (la directory in cui si trovano sia i file contenenti i dati che il file con le testate, le note e i titoli), la directory di output (directory in cui si vogliono salvare le tavole), alcune caratteristiche relative al file contenente i dati e il nome del file con le testate, le note, e i titoli. Queste informazioni vengono memorizzate, quindi devono essere impostate soltanto la prima volta che si utilizza il programma.

³ Se necessario è possibile aggiungere nuovi fogli per ulteriori testate. In questo caso è necessario rinominare il foglio "testatan" e copiare il contenuto di un foglio relativo ad un'altra testata in modo da trovarsi le indicazioni relative ai parametri da inserire.

Figura 6 - Foglio “DNA” del file contenente le testate, le note e i titoli

	A	B	C	D	E	F	G	H	I
1	NOME TAVOLA	BOX	NOTE	TESTATA	nome tavola	TITOLO			
	Tavola 2.1 -								
2		1		1	tav2_1a	1	20		
	Tavola 2.1 segue -								
3		1		2	tav2_1b	2	20		
	Tavola 2.1 segue -								
4		1		3	tav2_1c	3	20		
	Tavola 2.2 -								
5		5	2	1	tav2_2a	4	20		
	Tavola 2.2 segue -								
6		5	2	2	tav2_2b	2	20		

Figura 7 - Schermata principale

Impaginazione automatica di tavole statistiche

Impagina:

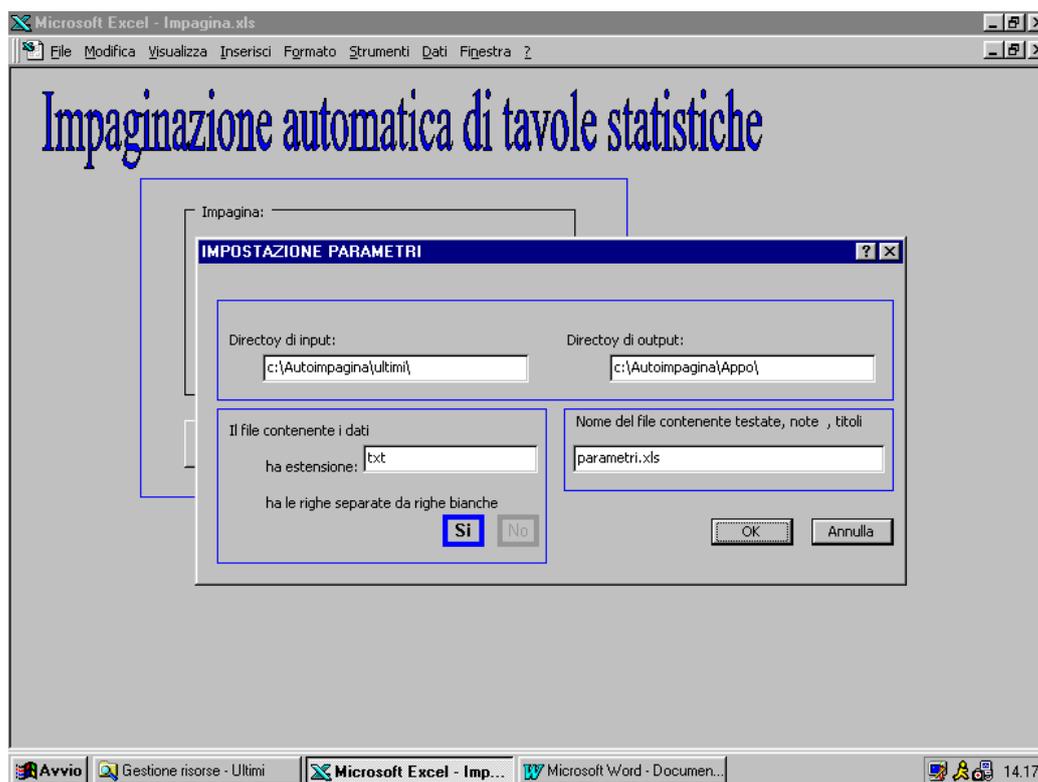
tutto

Selezione Tavola 2.2 segue -

Annuari Informazioni Altro

inserisci / modifica parametri

Figura 8 - Maschera di inserimento

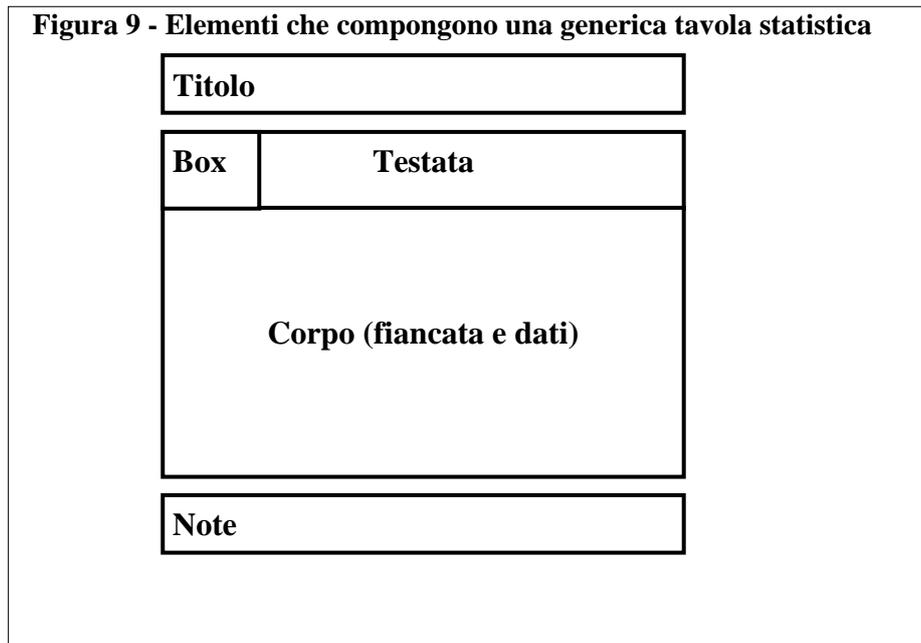


A questo punto le tavole sono pronte per essere impaginate. È possibile trattare una tavola per volta, selezionandola da un menù a discesa, oppure impaginare 'tutto', ovvero tutte le tavole che si trovano elencate nel foglio "DNA" del file sopra descritto.

Le impostazioni per la stampa possono essere quelle standard per la collana *Annuari* o per quella *Informazioni* (facendo clic rispettivamente sui bottoni "Annuari" e "Informazioni"), oppure si possono definire delle impostazioni personalizzate, facendo clic sul bottone "Altro".

La tavola (o le tavole, se si è scelto di impaginare 'tutto') viene quindi composta a partire dai suoi elementi (figura 9), ed impaginata secondo gli standard: il corpo della tavola viene importato in Excel dal file contenente i dati e viene trattato attraverso l'interpretazione dei caratteri di controllo che si trovano nella fiancata. Successivamente vengono copiati la testata, le note, i titoli dai fogli che li contengono, secondo quanto indicato nel foglio "DNA". Infine vengono realizzate in automatico tutte le impostazioni di carattere generale, si tratta di quelle impostazioni dettate dallo standard dell'Istituto che devono essere impostate su ogni tavola (ad esempio l'altezza della riga, il formato delle celle, la riga nera sottile in fondo alla tavola ecc...), vengono impostati i margini e tutte le informazioni relative alla stampa e viene salvato il file in formato Excel.

Figura 9 - Elementi che compongono una generica tavola statistica



Un tale modo di lavorare rende automatica la procedura di tabulazione, riducendo notevolmente i tempi di lavorazione delle tavole statistiche, garantendo il rispetto degli standard e riducendo drasticamente la possibilità di introdurre errori. Inoltre, una volta impostato il file contenente gli elementi che concorrono per la realizzazione della tavola, ogni successiva modifica alla tavola stessa risulta semplice da apportare e, poiché ogni elemento può essere utilizzato da più tavole, una eventuale modifica di un elemento verrà riportata su tutte le tavole che lo utilizzano. Non è quindi necessario ripetere la modifica tante volte quante sono le tavole su cui si intende intervenire.

Il livello di generalizzazione del programma può, naturalmente essere aumentato prevedendo, ad esempio ulteriori caratteri di controllo in fiancata per gestire formati attualmente non previsti, oppure prevedendo la possibilità di utilizzare come file contenente i dati anche file con caratteristiche diverse da quelle descritte.

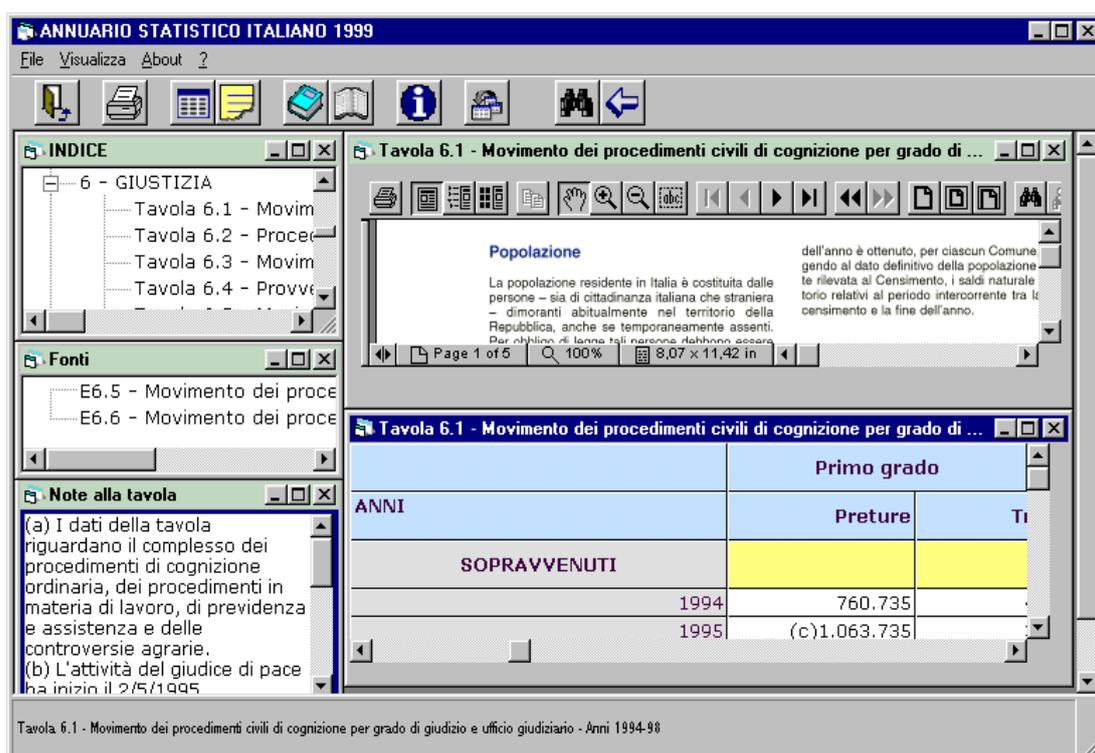
5. Altre realizzazioni

Parallelamente e indipendentemente dal programma per l'impaginazione automatica di tavole statistiche, descritto nel capitolo precedente, è stata sviluppata presso il Centro stampa dell'Istituto da F. Basilotta un'applicazione Visual Basic per la realizzazione di prodotti editoriali su CD ROM. Con tale sistema sono già stati pubblicati, tra l'altro, il Rapporto annuale 1999 e l'Annuario statistico italiano 1999.

Il programma prende in input tavole Excel già predisposte, dal momento che questa modalità di lavoro è molto comune in Istat soprattutto quando il contributo alle pubblicazioni generali è fornito da molti servizi statistici diversi.

L'applicazione è guidata dall'indice generale dei capitoli e delle tavole e consente la visualizzazione in apposite finestre dello schermo espandibili se necessario, dei file di testo, in formato PDF, dei dati, il cui formato di input è Excel o Access, delle note, dei file di metadati agganciati attraverso riferimenti ipertestuali (figura 10). Sono previste, oltre alla visualizzazione, le funzioni di esportazione (in formato .xls o .txt o .pdf) e la stampa su carta. Inoltre consente la ricerca per stringa effettuata sulle parti testuali e sui titoli delle tavole.

Figura 10 - Annuario statistico italiano 1999: aspetto delle finestre di visualizzazione



Per realizzare la formattazione delle tavole vengono interpretati e gestiti, come per il programma di impaginazione descritto nel capitolo precedente, dei caratteri di controllo a livello di riga e, similmente, gli elementi che compongono la tavola (titolo, testata, corpo, note) sono gestiti separatamente in quanto presentano problematiche diverse.

Uno dei primi metodi per la formattazione automatica che hanno introdotto l'uso di macro in ambiente Excel per realizzare tavole pronte per la stampa è stato messo a punto da A. Giordano, descritto in un documento a circolazione interna dell'ottobre 1997, e utilizzato per comporre l'Annuario statistico "Nascite: caratteristiche demografiche e sociali - Anno 1995".

Vengono gestiti dati ottenuti da elaborazioni SAS, COBOL, TPL ed altri formati impostando, per ogni tavola del volume, una cartella di lavoro Excel costituita da due fogli di lavoro e da un modulo. Nei fogli di lavoro vengono posti rispettivamente i dati di output delle precedenti elaborazioni e la maschera, cioè l'ossatura della tavola, mentre il modulo contiene la macro (l'insieme di istruzioni eseguibili) elaborata per importare i dati di output ed allocarli all'interno della maschera predisposta.

L'utilizzo del metodo in questione è particolarmente indicato nel caso in cui la tavola rimane inalterata da un anno all'altro. In questo caso, infatti, si ottiene un cospicuo risparmio di tempo poiché, restando pressoché costanti la maschera e la macro, l'unica modifica concerne il foglio di lavoro contenente i dati, i quali devono essere ovviamente aggiornati di volta in volta. Il programma consente anche di introdurre nella "maschera" funzioni di controllo e di calcolo (come ad esempio totali o percentuali). Tuttavia il vantaggio della presenza di una maschera *ad hoc* per ogni tavola costituisce anche un limite rispetto ad una soluzione di tipo generalizzato.

Un'altra tappa importante tra le esperienze interne all'Istituto di automatismo nella impaginazione e formattazione di tavole statistiche è il prodotto CONVEX che permette di ottenere fogli elettronici, formato Microsoft Excel, dai file approntati per la Fotocomposizione con la procedura cosiddetta "512". L'applicazione è stata completamente sviluppata in ambiente Windows utilizzando il "Visual Basic 4 Enterprise", sfruttando al meglio l'interfaccia grafica utente (GUI). È stato realizzato da S. Baldanza e presentato in Istat in un seminario nel Luglio 1996.

La documentazione relativa alle due applicazioni è consultabile in Intranet (*home intranet* → *attività e progetti* → *pubblicazioni*).

Per la diffusione in ambiente Web il formato delle tavole dovrà essere visualizzabile con i più diffusi *browser*. Tuttavia, per questa modalità di diffusione e le possibili automazioni di processo, anche se la riflessione è stata meno approfondita, valgono le considerazioni fatte per i prodotti su CD ROM anche per i quali la modalità di visualizzazione con un *browser* potrebbe essere quella che pone meno problemi per il futuro. In tal caso l'applicazione non dovrebbe farsi carico dei problemi di visualizzazione ma limitarsi a fornire al browser le adeguate informazioni su come caratterizzare graficamente gli oggetti presenti nello schermo.

6. Sviluppi futuri e generalizzazioni

Il quadro descritto finora mette in evidenza un contesto di innovazione, in cui le diverse realtà coinvolte hanno partecipato cercando e determinando soluzioni, più o meno generalizzate, per le difficoltà che hanno incontrato nello svolgimento del proprio lavoro. Il programma per

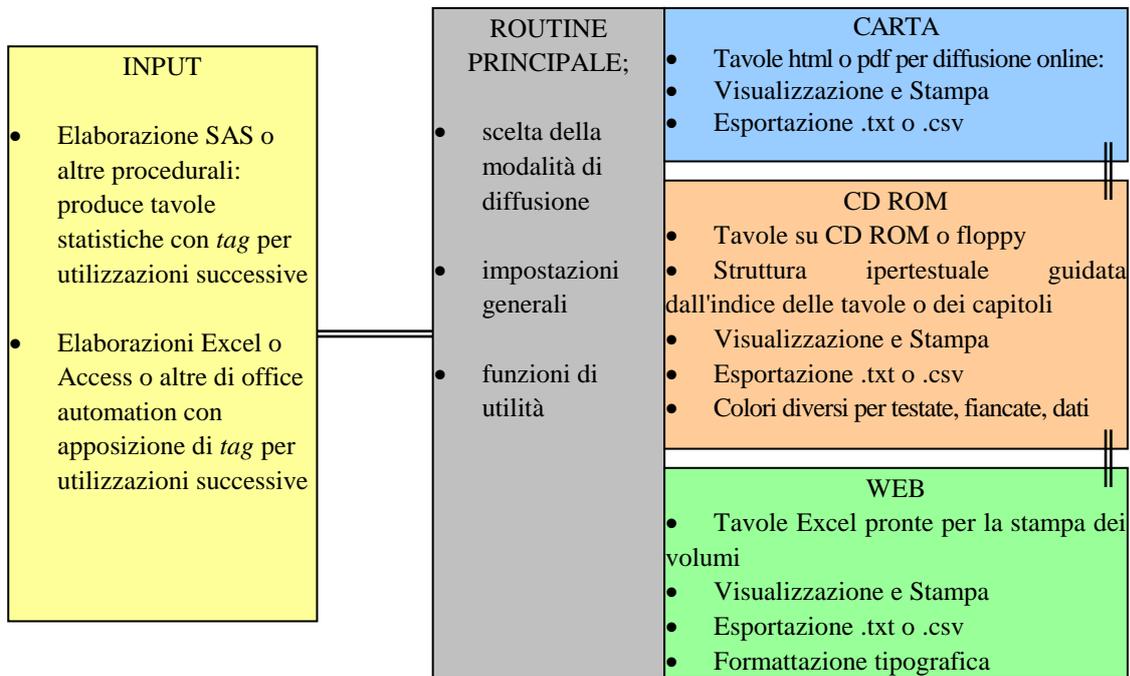
l'impaginazione automatica di tavole statistiche, descritto nel capitolo 4, così come le altre applicazioni informatiche descritte nel capitolo successivo, è frutto della necessità di risolvere problemi specifici. Ognuno affronta un problema diverso portando un contributo interessante ed efficace per la realizzazione di un obiettivo più ampio: gestire in termini di **sistema integrato** le diverse modalità di diffusione. Ovvero si vuole poter indifferentemente realizzare una pubblicazione cartacea, la visualizzazione su CD ROM o su Web a partire da un unico file di uscita dalla fase di elaborazione e aggregazione dei dati. Ciò consentirebbe di diffondere su supporti diversi le stesse informazioni statistiche senza dover ripetere l'elaborazione e consentirebbe inoltre di modificare le decisioni relative alla tipologia di prodotti di diffusione con un onere organizzativo ed economico minimo.

L'elemento chiave di un tale approccio è costituito dall'individuazione di un **formato standard per i file** contenenti i dati che possa essere trattato in modo differenziato dalle diverse routine a seconda della tipologia di prodotto editoriale prescelto. Ciò che consente un uso degli stessi dati per finalità diverse sta da un lato nella **segmentazione della tavola** nelle sue componenti (testata, titolo, note e corpo) dall'altro nell'**utilizzo di tag (caratteri di controllo)**. Ognuna delle componenti necessita, infatti, di un diverso trattamento, difficilmente gestibile interpretando la tavola come un tutt'uno. Inoltre, la scomposizione delle tavole in componenti permette di riutilizzare ogni componente più volte (una testata, ad esempio, anche se utilizzata da più tavole viene realizzata una sola volta). I caratteri di controllo, inseriti nella fiancata della tavola, possono essere gestiti dalle routine informatiche per caratterizzare i diversi prodotti editoriali. L'insieme delle *tag* è definito in modo che ciascuna routine ne utilizzi una parte o al limite tutte (figura 10).

Le modalità di funzionamento devono essere **parametrizzate e gestite esternamente** ai programmi: analogamente a quanto descritto per il programma di impaginazione per le pubblicazioni cartacee (capitolo 4). Per poter raggiungere l'obiettivo prefissato, cioè gestire in modo integrato le diverse modalità di diffusione, è necessario che anche i file contenenti gli elementi che compongono la tavola e i parametri, così come il file di input, possano essere utilizzati da tutte le applicazioni che realizzano i differenti prodotti editoriali.

In una prospettiva di evoluzione ulteriore, applicazioni diverse che utilizzano gli stessi file di dati, gli stessi oggetti e gli stessi parametri per realizzare i diversi prodotti editoriali, potrebbero diventare parti di **un'unica applicazione complessiva** (figura 11) da cui sia possibile scegliere la modalità di diffusione desiderata.

Figura 11 - Componenti di un unico programma di gestione dei prodotti editoriali



Questa potrebbe, inoltre, gestire anche alcune caratteristiche di contorno; ad esempio, per le pubblicazioni cartacee, potrebbe gestire la numerazione di pagina, le intestazioni e piè di pagina, realizzando l'impaginazione completa. Mentre per la realizzazione su CD ROM, potrebbero essere gestiti indici e collegamenti ipertestuali per facilitarne la navigazione. Ovviamente questa è un'ipotesi ancora ambiziosa e di portata notevole su cui vale la pena di investire.

Appendice

È qui riportato un breve questionario che è stato rivolto ai referenti Eurostat presso gli INS chiedendo loro di interessare all'occorrenza responsabili della produzione o della diffusione. Attraverso le risposte si voleva acquisire una conoscenza, anche se superficiale, circa le modalità di trattamento dei dati per le pubblicazioni presso istituzioni che hanno in comune con l'Istat metodologie e finalità simili.

1 - Which is the format of data for editing paper volumes (Excel, Access, Sas, Oracle, others...)?

Sweden: Word and Excel for simpler publications. Pagemaker for more complex publications.

Finland: PDF , Ventura 8 - ideal process. All processes are not yet digitalized, some documents are scanned from paper from many different sources. (Excel tables, Word texts, HG graphs).

Denmark: Excel and Word

Germany: Excel and Word

Spain: Mainly Excel. Corel Draw is also used for some special volumes.

2 - Who prepares the layout of paper publications (external editors, internal specialized offices, others...)?

Sweden: Simple publications like Statistical Reports: The person responsible for the contents of the publication. Yearbooks and other more complex publications: Our own graphic design studio.

Finland: We have two internal graphic designers and about 10 layout makers. Some of our experts prepare their materials themselves using ready-made templates. Some of our work is done by external editors. (due to lack of resources).

Denmark: Text and figures are put into a Word-template, originally developed by internal specialists.

Germany: Internal specialized office

Spain: There's a manual of Institutional image in force and lots of courses have been delivered among the people involved so that every production unit is able to provide originals almost ready to be printed. All printing process is made within INE. There is a unit (Unidad de Composición) within the Dissemination Direction responsible of layout that deals with some publications and help producers. In the case of graphics and maps all of them are made by this specialized unit.

3 - Generally, do the statistical offices, which produce data, also carry out statistical tables on their own, or do they provide elementary data sets that are accessible by query programs?

Sweden: The statistical departments at Statistics Sweden usually produce their own tables. For the moment the Unit for Information and Publishing is working on a new tool for table production together with Statistics Denmark. to be used when preparing Internet publications.

Finland: Depending on the process: if Ventura is used, the tables need to be edited a slightly. If the experts prepare the materials themselves, the tables are ready to be scanned, but need paging and logo setting.

The statistical units derive their data from databases.

We are streamlining our process, with an aim to produce services that can be edited from the same electronic archive. (printing, the internet).

Denmark: The tables are produced by the producers in the statistical offices.

Germany: The departments of our office carry out the statistical tables on their own

Spain: They produce always statistical tables in many different formats.

The tabulation plan of a survey is never a responsibility of Dissemination Unit.

4 - Is it more important for your publications to release the information quickly or to give them a pleasant and coloured feature?

Sweden: Emphasis is put on quick publication in Statistical Reports which have a very standardized and simple layout. Next year these will be published as PDF.files on our web site.

We also produce a few thematic annuals with a more elaborate design. We seldom use more than two colours (black + colour), in the publications but the cover can be in four colours.

Finland: We have fast production where the appearance is not the main issue but we spend more time and effort with publications intended to be marketed

Denmark: The Word-template can give you an acceptable layout, which does not delay the dissemination.

Germany: We have so-called "subject matter series" for releasing the information quickly and others (for example the series "In the spotlight" for which a pleasant feature and layout is considered important

Spain: Of course quickness is our obsession. We've made all efforts to automatise the procedures (Excel Macros, others) so that as soon as new data are available the new original for printing is ready. In some cases (Labour Force), the original is sent to printing the same day the data are delivered.

In some other cases (Spain in Figures, abstracts) more time is dedicated to design aspects. The Graphic Design Unit produces this special material.