



manuale per la progettazione concettuale dei dati statistici

istat
istituto centrale
di statistica

**metodi e norme
serie B n. 25**
edizione 1989

*Al fine di favorire la diffusione e l'utilizzazione dei dati e delle informazioni
l'ISTAT autorizza la riproduzione parziale o totale del contenuto del presente
volume con la citazione della fonte.*

Supplemento all'Annuario Statistico Italiano

ISSN: 0075-1847

INDICE

Introduzione	1
1. I sistemi informativi statistici	2
1.1. Il processo produttivo di informazioni statistiche.	2
1.2. I sistemi informativi statistici e l'integrazione del patrimonio informativo	3
1.3. La struttura del manuale.	6
2. La progettazione dei sistemi informativi	9
2.1. Le fasi della progettazione dei sistemi informativi.	9
2.1.1. L'analisi dei requisiti.	9
2.1.2. La progettazione concettuale.	10
2.1.3. La progettazione logica.	12
2.1.4. La progettazione fisica.	12
2.2. I vantaggi della progettazione concettuale.	12
2.3. Gli strumenti automatici per la progettazione concettuale.	14
3. Gli scenari di applicazione della metodologia integrata MIDAS	15
3.1. Primo scenario di applicazione.	15
3.2. Secondo scenario di applicazione.	16
3.3. Terzo scenario di applicazione.	18
4. Il modello Entita'/Relazione (E/R)	19
4.1. I modelli concettuali	19
4.2. Le strutture di rappresentazione di un modello concettuale	20
4.3. Il modello Entita'/Relazione	22
4.3.1. Strutture di classificazione	22
4.3.2. Vincoli di integrita'	23
5. La metodologia per la costruzione degli schemi dei dati disaggregati	27
5.1. Analisi del questionario	27
5.2. Disegno concettuale dei dati disaggregati	31
5.3. Integrazione degli schemi parziali	35
5.4. Verifiche finali.	36
6. Il Modello Concettuale Statistico (MCS)	39
6.1. Strutture di rappresentazione del Modello Concettuale Statistico	39
6.1.1. Classe di oggetti	41
6.1.2. Attributo di categoria	41
6.1.3. Classificazione statistica	41
6.1.4. Elaborazione statistica / Classe di dati	43
6.1.5. Generalizzazione tra classi di dati	49
6.1.6. Aggregato di attributi di categoria	49
6.1.7. Raggruppamento di attributi di categoria	49
6.2. Regole di costruzione di schemi MCS	55
6.3. Alcuni esempi di schemi MCS e loro integrazione	55

7. La metodologia complessiva per la progettazione dei dati disaggregati e di quelli aggregati	61
7.1. Analisi dei requisiti	61
7.2. Disegno concettuale dei dati disaggregati.	65
7.3. Disegno concettuale dei dati aggregati.	65
7.4. Integrazione dei sottoschemi	67
7.5. Verifiche finali	68
APPENDICE A: LA DESCRIZIONE LINGUISTICA DI SCHEMI E/R	75
APPENDICE B: LE OPERAZIONI SUI DATI DISAGGREGATI	79
APPENDICE C: LA DESCRIZIONE LINGUISTICA DI SCHEMI MCS	83
GLOSSARIO	85

Introduzione

Negli ultimi anni l'ISTAT ha avviato, in collaborazione con il Dipartimento di Informatica e Sistemistica dell'Università di Roma, un'attività di ricerca volta a definire una metodologia per la progettazione concettuale e la documentazione dell'attività statistica dell'Istituto e a creare gli appropriati strumenti informatici.

Nel passato, gli stati di avanzamento di tale attività sono stati documentati da una serie di pubblicazioni dell'Istituto, tra le quali citiamo:

- E. Fortunato, G. Manicardi, C. Vaccari: Estratto da: Progetto concettuale di basi di dati territoriali (Relazione finale);
- C. Batini, E. Fortunato: Metodi per la progettazione di basi di dati statistici.

Il presente manuale costituisce il risultato finale dell'attività di ricerca, relativamente agli aspetti metodologici sopra citati.

Il manuale è stato concepito con l'obiettivo di facilitare l'applicazione alle indagini statistiche degli strumenti metodologici proposti. Per tale motivo si è preferito fornire un prodotto comprensivo di tutte le soluzioni metodologiche via via acquisite nel corso dell'attività: alcune di tali soluzioni, e in particolare quelle relative alla progettazione dei dati disaggregati, sono già delineate nelle pubblicazioni precedenti.

Ringraziamo in modo particolare il prof. Carlo Batini e l'ing. Giuseppe Di Battista del Dipartimento di Informatica e Sistemistica dell'Università "La Sapienza" di Roma per i suggerimenti relativi all'impostazione ed al contenuto del manuale.

1. I sistemi informativi statistici

1.1. Il processo produttivo di informazioni statistiche.

Il processo produttivo di informazioni statistiche è caratterizzato da tre elementi fondamentali:

- i dati in ingresso;
- l'attività di trasformazione;
- i dati in uscita.

I dati in ingresso sono informazioni generalmente disponibili ad un elevato livello di disaggregazione e vengono di norma acquisiti all'esterno dall'Ente che gestisce il processo produttivo.

L'effettivo livello di disaggregazione dei dati in ingresso è variabile. Ad esempio, nel caso in cui le informazioni sono acquisite mediante rilevazione diretta, si dispone dei cosiddetti **dati elementari**, cioè di dati al massimo livello di disaggregazione possibile; nel caso in cui si ricorre a fonti amministrative (ad esempio Comuni, USL, ecc.), più frequentemente saranno disponibili informazioni ad un qualche livello di aggregazione. Tuttavia, nel contesto di questa trattazione il livello di aggregazione di un dato non è rilevante in assoluto ma, piuttosto, in relazione a quello dei dati complessivamente gestiti dal processo; per questo motivo, nel riferirci ai dati in ingresso, nel seguito sarà sempre utilizzato il termine **dati disaggregati**.

L'attività di trasformazione rappresenta l'attività strettamente statistica dell'Ente e consiste, in sostanza, nell'applicazione di un insieme prestabilito di **funzioni di aggregazione** ai dati disaggregati.

L'attività di trasformazione è un elemento fondamentale del processo produttivo. Infatti, le funzioni di aggregazione che caratterizzano tale attività hanno l'effetto di ridurre il potenziale informativo di partenza: una volta effettuate le aggregazioni, gli utenti saranno in

grado di procedere ad ulteriori aggregazioni dei dati ottenuti ma in nessun caso potranno, sulla base di essi, risalire al patrimonio di informazioni originario ovvero ottenere una diversa qualità delle informazioni. Così, ad esempio, avendo a disposizione le aggregazioni: "popolazione residente per sesso" e "popolazione residente per età", non sarà possibile ottenere da esse l'aggregazione: "popolazione residente per sesso ed età".

Una efficace attività di trasformazione richiede quindi una adeguata, preventiva conoscenza delle necessità informative dell'utenza, della evoluzione dei fenomeni di interesse e della importanza che essi assumono per gli utenti, a seconda del contesto di analisi in cui essi operano.

I dati in uscita sono in generale informazioni ad elevato livello di aggregazione generate dall'attività di trasformazione e destinate all'utenza; una parte di tali dati può essere destinata all'utenza interna e costituire quindi l'ingresso di un nuovo processo produttivo.

Le considerazioni svolte a proposito del livello di aggregazione dei dati in ingresso valgono anche per quelli in uscita i quali verranno indicati nel seguito con il termine **dati aggregati**.

In generale, il contesto organizzativo nel quale vengono gestiti i processi produttivi delle indagini statistiche assume una grande importanza. Nel caso dell'Istat, in particolare, la grande quantità di informazioni raccolte rende necessario un impianto organizzativo complesso, strutturato in unità operative (servizi e reparti), ognuna delle quali ha la responsabilità di una o più indagini statistiche.

Individuare i fenomeni da prendere in considerazione per ognuna delle indagini e ripartire queste ultime tra le varie unità operative sono nodi cruciali dell'attività dell'Istituto, su cui pesa in modo decisivo la necessità di contenere costi e tempi di esecuzione.

Il risultato è che ogni reparto tende a gestire i processi di produzione di indagini affini non tanto dal punto di vista del loro contenuto, quanto da quello delle connesse soluzioni tec-

niche e operative. Si consideri ad esempio il caso dei censimenti della popolazione e delle abitazioni i quali, pur riguardando fenomeni non omogenei, vengono organizzati e gestiti dallo stesso servizio in quanto richiedono la stessa organizzazione per la rilevazione.

Per contro, le informazioni complessivamente disponibili per un dato fenomeno risultano generalmente disperse tra i vari reparti e subiscono trasformazioni disomogenee, dettate dalla sensibilità e dalle specifiche necessità di analisi delle varie unità. Si veda in proposito lo schema di fig. 1.1.

1.2. I sistemi informativi statistici e l'integrazione del patrimonio informativo

La situazione descritta nel paragrafo precedente pone problemi nella costruzione dei sistemi informativi statistici (nel seguito, sistemi informativi).

Infatti, l'elemento che caratterizza un sistema informativo è la possibilità di mettere in connessione le informazioni disponibili sui fenomeni del mondo reale che si riferiscono ad una stessa area informativa. Così, ad esempio, il Sistema Informativo del Lavoro comprenderà le informazioni disponibili su fenomeni quali l'occupazione, la disoccupazione, il salario, la mobilità del lavoro, ecc. In questo modo, consentendo di trattare in modo integrato i dati relativi ad una stessa area informativa, le organizzazioni statistiche sono in grado di soddisfare in modo adeguato le esigenze di conoscenza e di analisi degli utenti. Si veda in proposito lo schema di fig. 1.2. Il concetto di sistema informativo rimanda quindi a quello di integrazione del patrimonio informativo.

Nella situazione organizzativa descritta, tale integrazione può essere valutata rispetto a due coordinate (si veda in proposito la fig. 1.3). La prima coordinata va nel verso dei processi produttivi ed è relativa alla cosiddetta **integrazione verticale** del patrimonio informativo. In questo caso, l'esigenza generale è quella di verificare la coerenza delle scelte operate nell'ambito dei singoli processi produttivi, scelte che riguardano sia i dati disaggregati che quelli aggregati, oltre alla attività di trasformazione.

In effetti, nelle unità operative le responsabilità di tali scelte risultano spesso ripartite tra diversi soggetti, peraltro caratterizzati da una notevole mobilità. Per tale motivo risulta tal-

volta difficile individuare con esattezza il contenuto dei dati e dell'attività statistica relativi ai singoli processi produttivi.

La seconda coordinata va nel verso delle classi di dati omogenei ed è relativa alla cosiddetta **integrazione orizzontale** del patrimonio informativo.

Gli insiemi di dati omogenei rilevanti ai fini dell'integrazione orizzontale sono quelli dei dati disaggregati e dei dati aggregati. In questo caso, quindi, l'esigenza è duplice:

- da una parte, si tratta di ricostruire il patrimonio informativo di base (dati disaggregati) complessivamente a disposizione dell'Ente, indipendentemente dalla sua distribuzione organizzativa;

- dall'altra, è necessario ricomporre il patrimonio informativo finale (dati aggregati). Tale patrimonio rappresenta, sul versante esterno, l'offerta complessiva di informazioni all'utenza, mentre, sul versante interno, documenta l'attività statistica globalmente svolta dall'Ente.

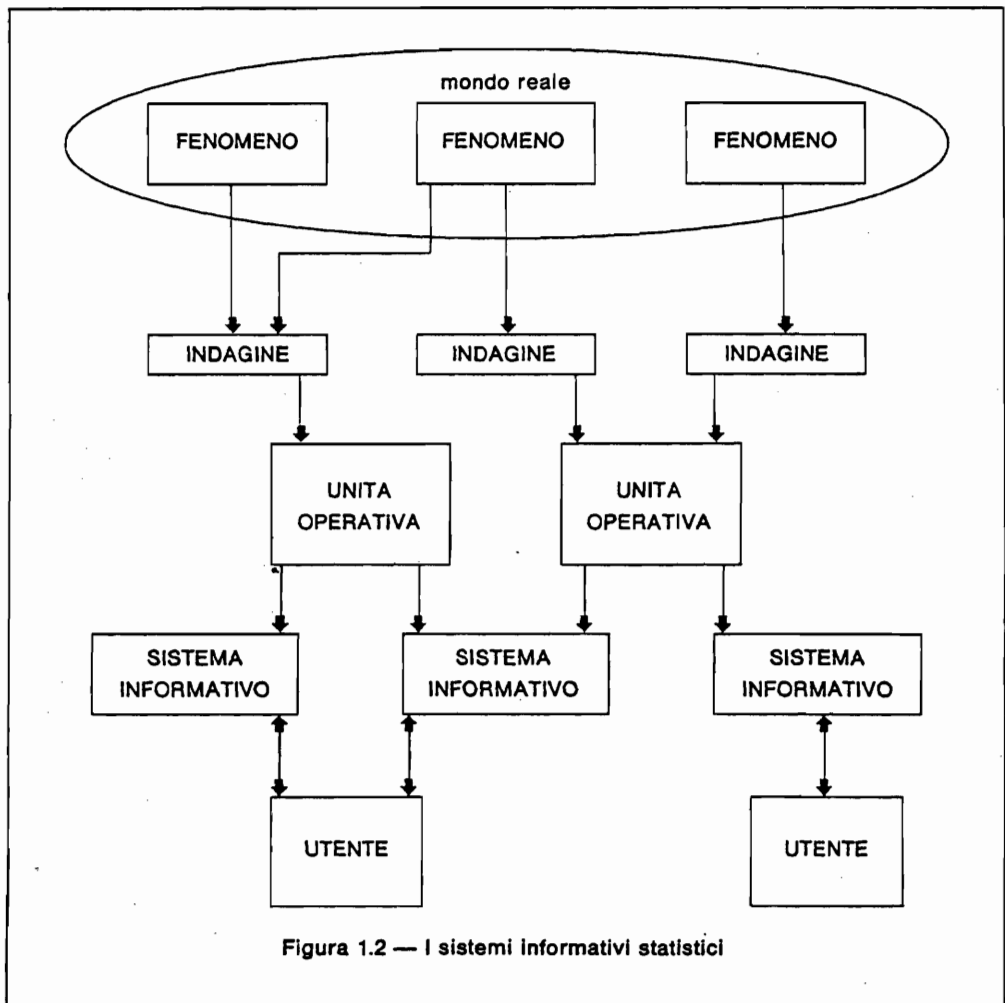
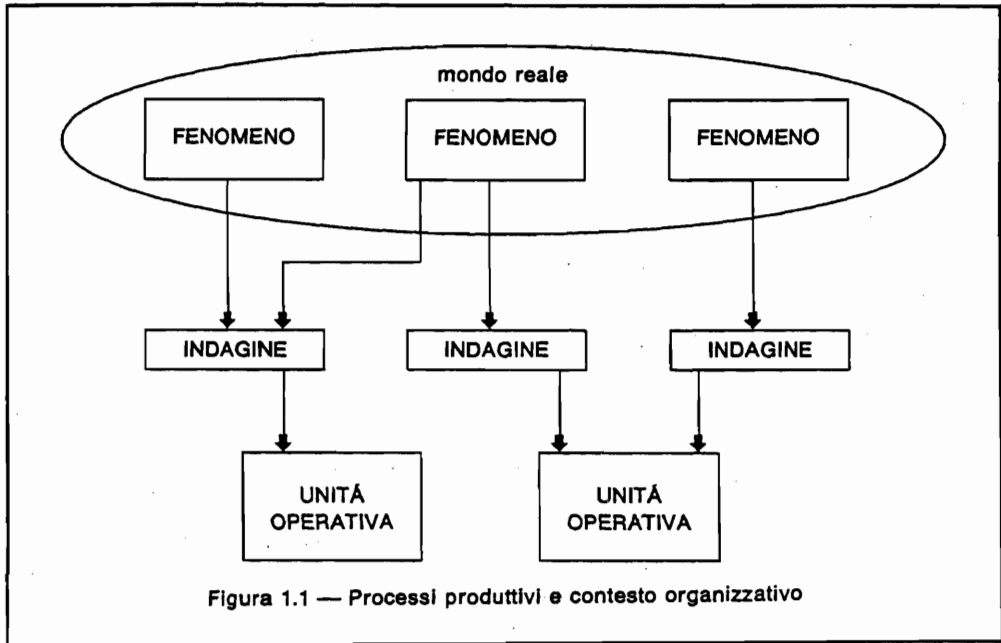
Nell'attività di integrazione orizzontale, l'analisi delle informazioni di interesse è caratterizzata da alcune situazioni tipiche, che richiedono specifici interventi. Le più rilevanti appaiono le seguenti:

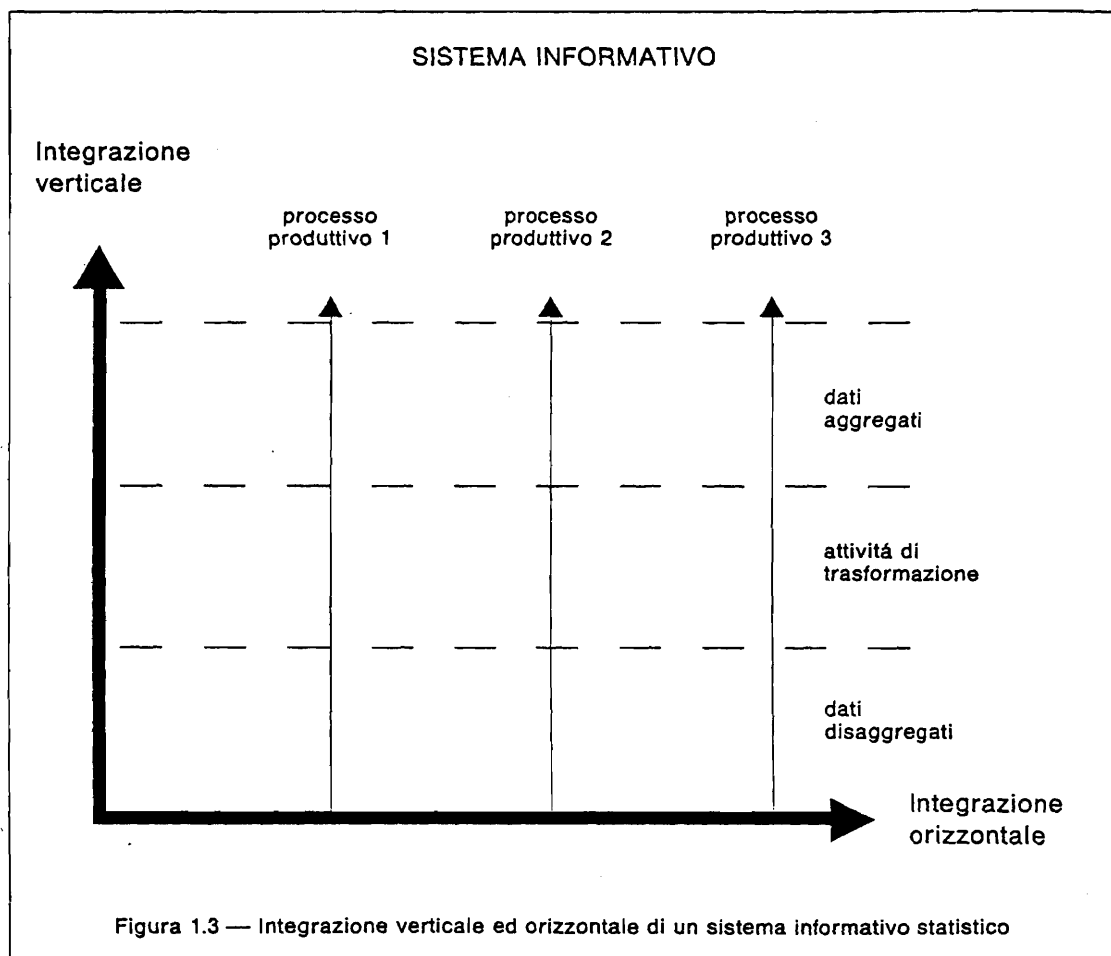
- talvolta è difficile individuare, nelle articolazioni dell'organizzazione, i frammenti di conoscenza di interesse per il sistema informativo; in particolare, risulta difficile riconoscere ed interpretare i legami informativi (correlazioni) esistenti tra tali frammenti;

- esiste in genere una notevole disomogeneità nelle definizioni e nelle classificazioni dei fenomeni di interesse, derivante dalla eterogeneità degli approcci e dalla autonomia di gestione delle unità operative preposte alle indagini;

- si riscontrano generalmente casi di ridondanza e/o incongruenza delle informazioni, determinati dalla autonomia suddetta, a fronte delle quali emergono peraltro lacune informative, valutabili solo con un'analisi integrata delle informazioni di interesse.

Da quanto discusso, risulta evidente che l'integrazione del patrimonio informativo si basa sul confronto tra i dati di interesse e, quindi, sulla possibilità di descrivere in modo





formale e non ambiguo tali dati (si veda in proposito la fig. 1.4). Più in generale, la descrizione formale del patrimonio informativo a disposizione consente di verificare, in qualsiasi stadio dell'attività di produzione, la coerenza degli sviluppi raggiunti con gli obiettivi finali e con il patrimonio informativo preesistente.

Le metodologie di progettazione concettuale hanno appunto l'obiettivo di fornire una descrizione formale, integrata ed indipendente dall'ambiente tecnologico, dei contenuti e delle caratteristiche di un sistema informativo.

Con il termine **formale** si intende che la descrizione deve essere espressa con un linguaggio non ambiguo; ciò rende tale descrizione interpretabile in modo univoco anche da chi non l'ha prodotta.

Con il termine **integrata** si intende che la descrizione deve essere in grado di rappresentare in modo globale tutta la realtà di interesse.

Con il termine **Indipendente dall'ambiente tecnologico** si intende infine che la descrizione deve riferirsi unicamente agli aspetti concettuali della realtà di interesse per il sistema informativo e non anche a quelli della realizzazione concreta del sistema.

Come sarà meglio specificato in seguito, la progettazione concettuale si avvale di un insieme di strutture di rappresentazione detto **modello concettuale**; la particolare descrizione ottenuta applicando un modello concettuale ad una determinata realtà di interesse viene invece chiamata **schema concettuale**.

In questo volume viene presentata una metodologia di progettazione concettuale denominata **MIDAS** (Metodologia Integrata per la progettazione dei Dati Statistici).

MIDAS è il risultato dell'attività di ricerca condotta dall'ISTAT e dal Dipartimento di Informatica e Sistemistica dell'Università "La Sapienza" di Roma. MIDAS è costituita dalle seguenti due parti:

- una metodologia per la progettazione dei dati disaggregati. L'applicazione di tale metodologia fornisce come risultato gli schemi concettuali dei dati disaggregati. In questa fase vengono utilizzate le strutture di rappresentazione del modello concettuale **Entità/Relazione** (E/R);

- una metodologia per la progettazione dei dati aggregati. In questo passo, a partire dagli schemi concettuali ottenuti nella fase precedente, vengono prodotti schemi concettuali che rappresentano sia i dati aggregati che le attività di trasformazione effettuate per la loro produzione. Il modello concettuale utilizzato è **MCS** (Modello Concettuale Statistico), appositamente ideato per la rappresentazione dei dati statistici.

1.3. La struttura del manuale.

Il manuale è strutturato nel modo seguente.

Nel capitolo 2 si riportano schematicamente le fasi della progettazione di un sistema informativo.

Nel capitolo 3 vengono illustrate le modalità di applicazione di MIDAS in vari scenari organizzativi; tali scenari rappresentano le tipiche situazioni nelle quali vengono costruiti i sistemi informativi.

Nel capitolo 4, dopo una trattazione delle caratteristiche generali dei modelli concettuali, viene presentato il modello concettuale Entità/Relazione che, come detto, è il modello utilizzato da MIDAS per la progettazione dei dati disaggregati.

Nel capitolo 5 viene illustrata la metodologia per la progettazione dei dati disaggregati. La presentazione viene effettuata attraverso esempi di utilizzazione che si riferiscono al Sistema Informativo del Lavoro (SIL).

Il capitolo 6 è dedicato al modello MCS, cioè al modello per la progettazione dei dati aggregati.

Nel capitolo 7 viene esposta la metodologia per la progettazione dei dati aggregati, anche in questo caso attraverso esempi di utilizzazione relativi al SIL.

Il manuale comprende inoltre un glossario dei termini utilizzati o di particolare rilevanza in tema di progettazione concettuale dei sistemi informativi.

Nelle appendici sono illustrate la descrizione linguistica dei modelli E/R ed MCS e le operazioni su dati disaggregati.¹

¹ Il capitolo 1 è dovuto a E.Fortunato, il capitolo 2 a G.Manicardi, il capitolo 3 a E.Fortunato, il capitolo 4 a A.Magalotti, il capitolo 5 a C.Vaccari, il capitolo 6 a G.Barcaroli, il capitolo 7 a G.Barcaroli, l'appendice A a E.Fortunato, l'appendice B a A.Magalotti, l'appendice C a G.Manicardi.

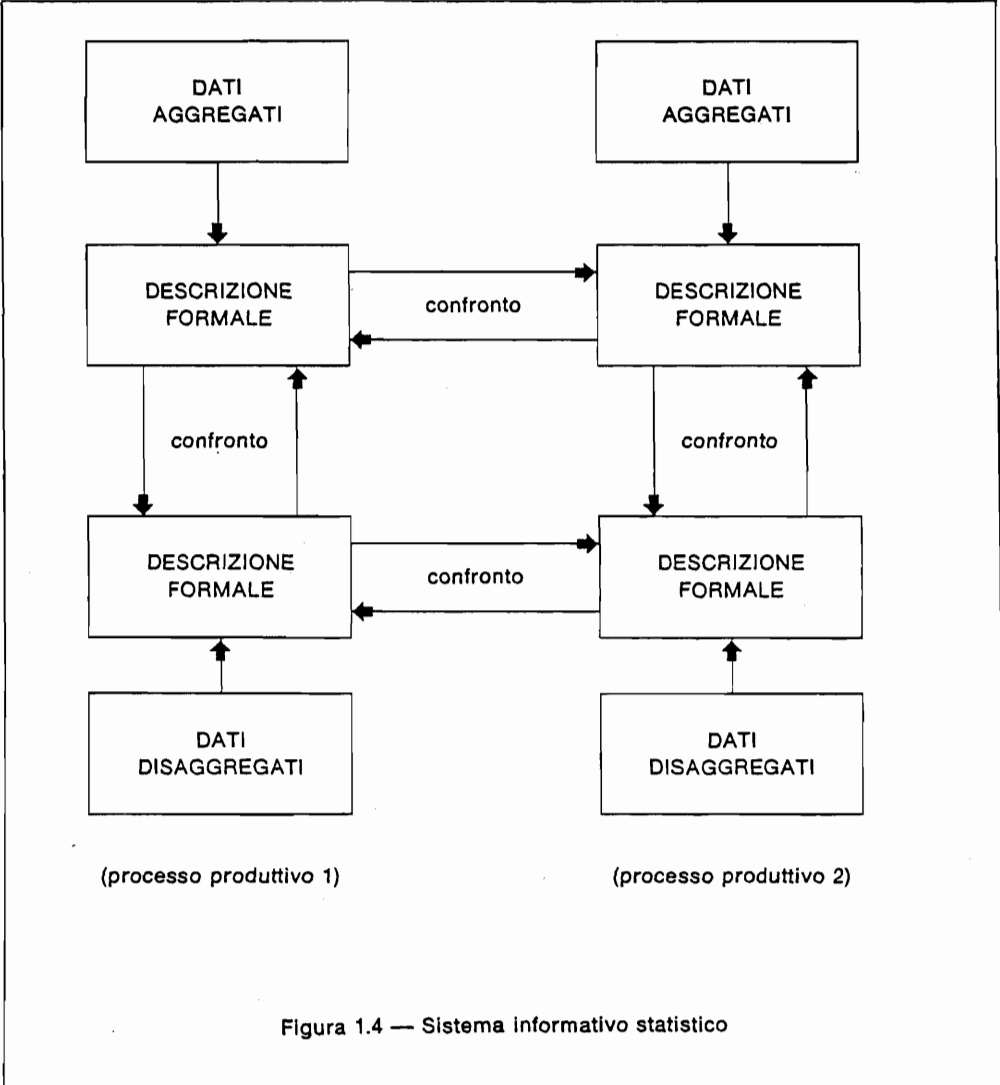


Figura 1.4 — Sistema informativo statistico

2. La progettazione dei sistemi informativi

2.1. Le fasi della progettazione dei sistemi informativi.

La progettazione di un sistema informativo puo' essere suddivisa nelle seguenti quattro fasi:

- analisi dei requisiti;
- progettazione concettuale;
- progettazione logica;
- progettazione fisica.

Attraverso tali fasi, vengono progressivamente definiti gli elementi necessari alla realizzazione di un sistema informativo, dapprima focalizzando l'attenzione sui contenuti e gli obiettivi (livello informativo) e, successivamente, operando le scelte necessarie per la concreta realizzazione del sistema (livello informatico). In questo processo vengono via via coinvolte le diverse figure professionali dell'organizzazione: da quelle statistiche (distribuite tra le varie unita' operative) a quelle informatiche (che fanno capo al Centro Elaborazione Dati dell'Ente). Le fasi suddette saranno trattate piu' diffusamente nei prossimi paragrafi.

2.1.1. L'analisi dei requisiti.

Per **requisiti di un sistema informativo** si intendono le informazioni necessarie per descrivere in modo corretto e completo la realta' di interesse per il sistema informativo.

Tali informazioni sono relative ai dati, alle operazioni previste sui dati e agli eventi che regolano l'ordine e la causalita' delle operazioni. I dati sono i **requisiti statici** del sistema informativo; le operazioni e gli eventi rappresentano invece i **requisiti dinamici**. L'analisi dei requi-

siti ha come obiettivo la raccolta e l'analisi delle informazioni suddette.

Le modalita' con le quali viene effettuata l'analisi dei requisiti dipendono strettamente dalla forma, cioe' dal **linguaggio**, utilizzato per descrivere i requisiti. Generalmente, nell'organizzazione vengono utilizzati linguaggi molto diversi, che possono descrivere le informazioni in modo piu' o meno strutturato, cioe' secondo regole non ambigue e descrivibili formalmente.

Un modo usuale di rappresentare i requisiti e' il **linguaggio naturale**, con il quale vengono in genere espressi requisiti quali:

- interviste agli utenti finali effettuate per ottenere le informazioni necessarie alla progettazione;
- leggi, norme, regolamenti, circolari e note esplicative rilevanti ai fini dei contenuti e delle caratteristiche del sistema;
- verbali delle riunioni e delle commissioni in cui si definiscono i fenomeni di interesse del sistema informativo;
- liste a diverso livello di dettaglio delle informazioni da acquisire attraverso i questionari di rilevazione o da ottenere come aggregazioni.

Il linguaggio naturale e' una forma di descrizione assai poco strutturata e risulta quella prevalentemente disponibile quando nell'organizzazione si ha una idea di massima del sistema informativo da costruire e i relativi contenuti non sono quindi definiti con precisione.

Un'altra modalita' di descrizione dei requisiti e' rappresentata dai **questionari di rilevazione** o, piu' in generale, dai **moduli**. Questa forma di rappresentazione, molto piu' strutturata della precedente, e' tipica nelle situazioni in cui nell'organizzazione esiste un'idea precisa, al-

meno rispetto ai dati disaggregati, dei contenuti e delle caratteristiche del sistema informativo da realizzare.

In virtú dell'alto grado di strutturazione, i questionari di rilevazione rendono in genere semplice l'interpretazione e l'utilizzazione delle informazioni in essi riportate. In questo senso, nei questionari si possono distinguere quattro parti con funzioni diverse, e cioè (si veda in proposito la fig. 2.1):

- **parte certificativa**, che contiene informazioni quali data di emissione, bolli, firme, ecc. in genere non significative ai fini della progettazione;
- **parte estensionale**, cioè l'insieme delle zone del questionario che devono essere riempite all'atto della compilazione;
- **parte intensionale**, cioè la parte del questionario che contiene riferimenti espliciti o impliciti a concetti cui vengono associati valori nella parte estensionale;
- **parte descrittiva**, cioè la parte del questionario che contiene istruzioni o regole che devono essere rispettate nella compilazione.

Infine, i requisiti possono essere descritti sotto forma di **tavole statistiche**. Questa terza forma di rappresentazione, anch'essa notevolmente strutturata, e' in genere quella disponibile quando le decisioni sui contenuti del sistema informativo, relative ai dati aggregati, sono abbastanza consolidati nell'organizzazione.

Va notato che il grado di strutturazione dei requisiti rende piu' o meno complessa l'analisi dei requisiti. Questionari e moduli consentono evidentemente di operare in condizioni migliori, riducendo i rischi di errata interpretazione dei requisiti; descrizioni in linguaggio naturale possono invece prestarsi a diverse interpretazioni.

Dal punto di vista delle figure tecnico-professionali coinvolte, l'analisi dei requisiti richiede l'interazione tra le competenze statistiche delle unita' operative interessate al sistema informativo e quelle informatiche, che dovranno assicurare la corrispondenza tra la realizzazione informatica e gli obiettivi fissati. Tale interazione deve essere tanto piu' stretta quanto piu' i requisiti sono disponibili in forme non strutturate.

2.1.2. La progettazione concettuale.

Come gia' ricordato nel capitolo 1, l'obiettivo della progettazione concettuale e' quello di fornire una descrizione formale, integrata e indipendente dall'ambiente tecnologico, dei contenuti di un sistema informativo. Concretamente cio' avviene traducendo i requisiti del sistema informativo, raccolti e analizzati nella fase precedente. Tali requisiti costituiscono quindi l'input della progettazione concettuale.

Per descrivere la realta' di interesse del sistema informativo, la progettazione concettuale si avvale di strutture di rappresentazione; l'insieme delle strutture di rappresentazione utilizzate e' detto **modello concettuale**. Il risultato dell'applicazione di un modello concettuale ad una determinata realta' di interesse e' un insieme di **schemi concettuali** i quali costituiscono quindi l'output della progettazione concettuale.

La costruzione dello schema concettuale relativo ad un sistema informativo avviene in modo incrementale: partendo dai requisiti raccolti nella fase precedente, si procede per passi successivi alla attribuzione di appropriate strutture di rappresentazione ai concetti del mondo reale di interesse. Questo processo ha termine quando tutta la realta' di interesse e' rappresentata in modo completo e non ambiguo nello schema, mediante le strutture di rappresentazione del modello adottato.

Le classiche strategie con cui possono essere generate le strutture di uno schema concettuale sono note in letteratura con i termini **top-down** (TD) e **bottom-up** (BU):

- la strategia **top-down** produce lo schema concettuale finale attraverso una serie di raffinamenti successivi, partendo da uno schema iniziale con pochi concetti molto astratti, e raffinando via via lo schema con trasformazioni che aumentano il grado di dettaglio nella rappresentazione della realta' di interesse. L'attivitá di progettazione che corrisponde a tale strategia e' detta **derivazione**;

- la strategia **bottom-up** procede suddividendo la descrizione della realta' di interesse in frammenti via via piu' piccoli, fin quando ogni frammento descrive una parte dei requisiti di dimensioni ridotte. Successivamente, ogni frammento viene rappresentato per mezzo di un semplice

PARTE INTENSIONALE

Segue QUADRO 1 - Notizie sulle persone della famiglia

	6	7	8	9	10
COGNOME E NOME					
RELAZIONE CON IL CAPO FAMIGLIA					
SESSO	M <input type="checkbox"/> F <input type="checkbox"/>	M <input type="checkbox"/> F <input type="checkbox"/>	M <input type="checkbox"/> F <input type="checkbox"/>	M <input type="checkbox"/> F <input type="checkbox"/>	M <input type="checkbox"/> F <input type="checkbox"/>
LUOGO DI NASCITA					
DATA DI NASCITA					
STATO CIVILE					
CITTADINANZA					
TITOLO DI STUDIO (a)					
PROFESSIONE O CONDIZIONE NON PROFESSIONALE (b)					
INDIRIZZO DEL LUOGO DI STUDIO O DI LAVORO (c)					
PRESENZA NEL COMUNE DI DIMORA ABITUALE (d)	<input type="checkbox"/> si <input type="checkbox"/> no	<input type="checkbox"/> si <input type="checkbox"/> no	<input type="checkbox"/> si <input type="checkbox"/> no	<input type="checkbox"/> si <input type="checkbox"/> no	<input type="checkbox"/> si <input type="checkbox"/> no

PARTE ESTENSIONALE

(a) Riportare il titolo di studio indicato al punto 9.1; per chi ne è privo indicare «nessuno». (b) Riportare la risposta fornita al punto 10.3; per le persone in condizione non professionale riportare la risposta fornita al punto 10.1; apporre un tratto per le persone in età inferiore a 14 anni. (c) Riportare l'indirizzo (via, numero civico, Comune, Provincia o Stato estero) indicato al punto 12.1. (d) Desumere la notizia dal punto 7.1.

QUADRO 2 - Notizie sulle persone temporaneamente assenti dal Comune di dimora abituale
(persone per le quali è stato indicato «NO» all'ultima riga del precedente Quadro 1)

	1	2	3
COGNOME E NOME			
LUOGO DI PRESENZA (a)			
MOTIVO DELL'ASSENZA (b)			

(a) Indicare l'indirizzo (via e numero civico), il Comune, la Provincia o lo Stato estero.
(b) Trascrivere la risposta fornita al punto 7.2 b).

PARTE DESCRITTIVA

QUADRO 3 - Notizie sulle persone temporaneamente presenti presso la famiglia

	1	2
COGNOME E NOME		
SESSO	M <input type="checkbox"/> F <input type="checkbox"/>	M <input type="checkbox"/> F <input type="checkbox"/>
DATA DI NASCITA		
LUOGO DI RESIDENZA (a)		

(a) Indicare l'indirizzo (via e numero civico), il Comune, la Provincia o lo Stato estero.

PARTE CERTIFICATIVA

Data 1981

IL RILEVATORE

IL CAPO FAMIGLIA
(o chi per esso)

Figura 2.1 — Parti certificativa, descrittiva, estensionale e intensionale in un modulo del censimento della popolazione 1981

schema concettuale. Infine, gli schemi così prodotti vengono progressivamente fusi nello schema concettuale finale. L'attività di progettazione che corrisponde a tale strategia è detta **astrazione**.

Nei casi concreti, difficilmente sarà possibile adottare, nell'attività di progettazione, esclusivamente l'uno o l'altro dei metodi descritti. Frequentemente si presentano situazioni complesse, nelle quali può risultare più utile una strategia che chiameremo mista **top-down/bottom-up** (TD/BU).

La strategia mista TD/BU prevede sia passi di astrazione che passi di derivazione. In questo modo, l'attività di progettazione converge allo schema finale partendo dalla individuazione sia di concetti elementari che di concetti ad alto livello di astrazione.

2.1.3. La progettazione logica.

La progettazione logica viene realizzata dal settore informatico che, a questo punto del progetto, ha a disposizione le informazioni necessarie per la realizzazione del sistema informativo. Tali informazioni sono fornite dallo schema concettuale prodotto nella fase di progettazione concettuale.

La progettazione logica consiste nel tradurre gli schemi concettuali nello **schema logico** del sistema informativo, cioè in una rappresentazione formale in cui si tiene conto del particolare sistema di gestione (chiamato DBMS - Data Base Management System) adottato per la realizzazione informatica.

Lo schema logico di un sistema informativo risulta pertanto diverso a seconda che venga utilizzato, ad esempio, un DBMS di tipo relazionale o gerarchico.

Lo schema logico ha l'obiettivo di consentire la maggiore efficienza possibile rispetto alla esecuzione delle operazioni; a questo scopo verranno valutati aspetti quali la frequenza di attivazione delle transazioni, i volumi dei dati utilizzati, le caratteristiche di accesso, il tipo di transazione (batch o online). Inoltre, saranno impostati i parametri che assicurano la sicurezza e la riservatezza dei dati.

2.1.4. La progettazione fisica.

La fase di progettazione fisica, anch'essa curata dal settore informatico, consiste nella traduzione dello schema logico nel cosiddetto **schema interno**, cioè in una rappresentazione che tenga conto delle particolari caratteristiche hardware e software del sistema informatico esistente.

In questa attività vengono scelti i supporti e i parametri fisici di memorizzazione; definite le aree di memorizzazione e le dimensioni dei blocchi e dei buffer, risolti i problemi relativi alla compressione fisica dei dati. L'obiettivo primario della progettazione fisica è quindi l'efficienza della rappresentazione fisica.

Nella fig. 2.2 vengono illustrate le fasi del processo di progettazione dei sistemi informativi, con i relativi input e output.

2.2. I vantaggi della progettazione concettuale.

L'introduzione di un passo di progettazione concettuale nella costruzione di un sistema informativo offre una serie di vantaggi rilevanti. L'uso degli schemi concettuali, che come abbiamo visto ne costituiscono i prodotti tipici, consente infatti di ottenere:

- **il coinvolgimento e il controllo da parte del committente.** Gli schemi concettuali risultano facilmente comprensibili anche da utenti non esperti. Ciò consente al committente di interagire con il progettista a tutto vantaggio della qualità del progetto. L'uso di un linguaggio comune all'interno dell'Ente fra committenti e progettisti favorisce una maggiore chiarezza progettuale e quindi un più facile e rapido raggiungimento degli obiettivi fissati.

- **una maggiore chiarezza degli obiettivi.** Molte ambiguità e conflitti che nascono tra i committenti di una applicazione e i realizzatori derivano dalla scarsa chiarezza e precisione delle specifiche di progetto. Lo schema concettuale è una specifica formale e non ambigua che garantisce utente e progettista e costituisce un documento comune da utilizzare ogni

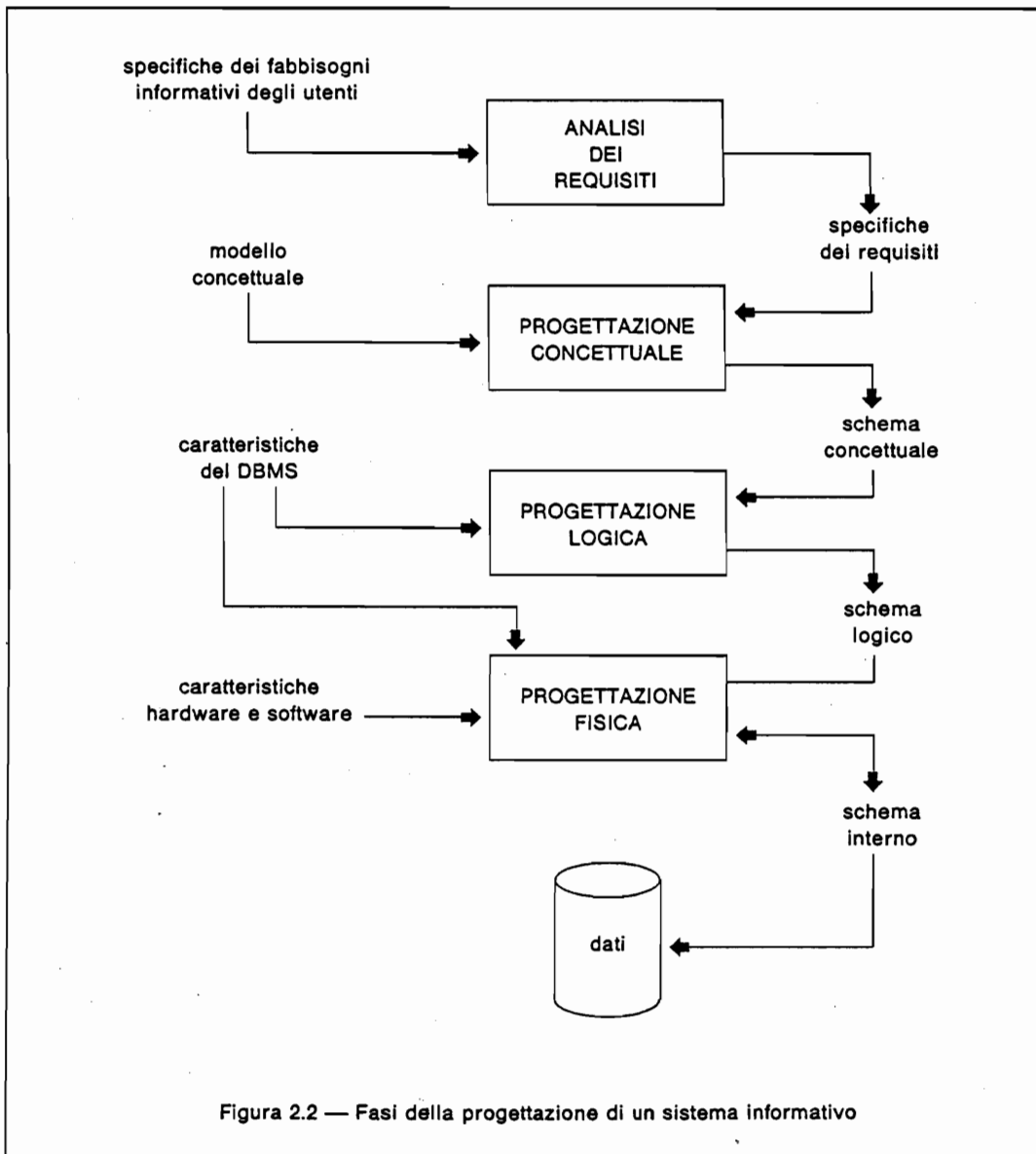


Figura 2.2 — Fasi della progettazione di un sistema informativo

qualvolta nascono conflitti o ripensamenti in fase di produzione.

- **la stabilita' della progettazione.** Nella fase di progettazione concettuale il dato viene analizzato come risorsa indipendente dalle applicazioni e dalla struttura organizzativa dell'azienda. Il modello concettuale e' quindi un prodotto relativamente stabile, non influenzato dalle applicazioni esistenti e invariante rispetto alle modifiche tecnologiche (cambiamenti nella dotazione hardware e software dell'Ente) che possono intervenire in un momento successivo alla formulazione del progetto. Cio' consente la trasferibilita' del progetto sulle nuove tecnologie e, quindi, un maggior potere contrattuale del committente dell'applicazione nei confronti del produttore di hardware e software.

- **una migliore qualita' della documentazione.** Lo schema concettuale e' infatti uno strumento di documentazione non ambiguo, di facile comprensione e, soprattutto se supportato da strumenti automatici grafici, anche di facile formulazione.

- **una maggiore facilita' di integrazione di nuovi dati.** L'uso di schemi concettuali consente di valutare con rapidita' l'impatto dell'inserimento di nuovi dati nel sistema e di individuare eventuali ridondanze.

- **migliori possibilita' di confronto fra sistemi informativi esistenti.** Spesso accade di dover confrontare, prescindendo dalle tecnologie, il contenuto informativo di sistemi gia' in esercizio e progettati indipendentemente da diversi Enti. Cio' e' particolarmente vero nella Pubblica Amministrazione dove, a fronte di una generale unitarieta' di intenti, esiste una articolazione in organizzazioni (Ministeri, EE.LL., etc.) le quali, in virtu' della loro indipendenza di gestione, tendono a sviluppare i propri sistemi informativi in modo completamente autonomo. L'esistenza degli schemi concettuali dei dati gestiti dai vari sistemi puo' consentire un piu' agevole confronto del loro contenuto informativo e, qualora lo si desideri, anche una loro piu' affidabile integrazione, rendendo cosi' attuabile una politica di razionalizzazione delle attivita' dei singoli Enti e, nel

complesso, di tutta la Pubblica Amministrazione.

2.3. Gli strumenti automatici per la progettazione concettuale.

La possibilita' di disporre di strumenti automatici di supporto all'attivita' di progettazione concettuale e' di grande importanza nell'economia generale delle organizzazioni produttive. Infatti, tali strumenti consentono un piu' efficace svolgimento di due funzioni tipiche delle organizzazioni e cioe' la documentazione e la progettazione dei sistemi informativi.

Riguardo al primo aspetto, va notato che la gestione della documentazione e' necessaria sia ai progettisti sia agli utenti del sistema informativo.

Per i primi, la documentazione risulta indispensabile soprattutto nel caso di sistemi di grandi dimensioni, in cui il numero di persone coinvolte nel processo di progettazione e' notevole e percio' frequenti sono le comunicazioni "orizzontali" tra progettisti interessati alla stessa fase del progetto e le comunicazioni "verticali" tra progettisti coinvolti in fasi differenti.

Per i secondi, la documentazione fornisce una concreta possibilita' di controllo della rispondenza del progetto ai requisiti.

Riguardo al secondo aspetto, un adeguato strumento automatico puo' essere di grande ausilio al progettista nelle scelte progettuali. In particolare, il progettista puo' essere assistito nella assegnazione di adeguate strutture di rappresentazione ai concetti espressi dai requisiti del sistema e, quindi, nella costruzione dei relativi schemi concettuali.

Nel prossimo futuro, l'obiettivo del gruppo di ricerca ISTAT-DIS e' quello di rendere disponibile uno strumento software per la progettazione e la documentazione dei sistemi informativi.

In attesa di tale realizzazione, in questo manuale viene proposto uno standard di documentazione, adottato per la progettazione concettuale dei sistemi informativi fin qui realizzati in ISTAT (vedi Appendice A e Appendice C).

3. Gli scenari di applicazione della metodologia integrata MIDAS

Abbiamo già valutato in precedenza, a proposito del problema dell'integrazione, come l'ambiente organizzativo influenzi la progettazione di un sistema informativo; ed abbiamo già descritto come MIDAS consenta tale integrazione, attraverso il confronto tra rappresentazioni formali degli insiemi di dati gestite in un sistema informativo. In questa sezione, vogliamo approfondire come MIDAS può essere utilizzata nelle diverse situazioni in cui può essere condotta la progettazione concettuale di un sistema informativo.

In genere, esiste una notevole varietà di situazioni di progetto: in alcuni casi, la progettazione interviene quando sono già noti gli obiettivi e i contenuti del sistema informativo; in altri, quando si ha solo un'idea di massima di tali obiettivi, i quali dovranno essere compiutamente definiti nell'ambito dell'attività di progetto.

Chiameremo **scenari di applicazione** le diverse situazioni in cui può essere sviluppata la progettazione concettuale di un sistema informativo. Nei prossimi paragrafi mostreremo come MIDAS può essere utilizzata in tre tipici scenari di applicazione.

3.1. Primo scenario di applicazione.

Nello scenario di applicazione illustrato in fig. 3.1, la progettazione concettuale si colloca a monte del processo produttivo. Lo scenario è caratterizzato dai seguenti aspetti:

- si ha un'idea di massima dei dati disaggregati di interesse; i relativi requisiti possono essere costituiti da schemi provvisori dei questionari di rilevazione ovvero da li-

ste indicative dei dati, espressi in linguaggio naturale;

- si ha un'idea di massima dei dati aggregati di interesse; i relativi requisiti sono in genere costituiti da liste indicative e provvisorie delle aggregazioni, descritte in linguaggio naturale.

L'applicazione della metodologia avviene in tre fasi.

PRIMA FASE

Nella prima fase vengono progettati i dati disaggregati. La progettazione si sviluppa in modo incrementale, come segue:

- a partire dai requisiti a disposizione, viene prodotto uno schema concettuale dei dati disaggregati, che descrive la realtà di interesse allo stato attuale di conoscenza;
- analizzando tale schema, si individuano nuovi aspetti della realtà e si definiscono in modo più preciso i requisiti del sistema;
- sulla base dei nuovi requisiti, si raffina lo schema di partenza, producendo un nuovo schema concettuale.

Il processo termina quando sono precisati tutti i requisiti del sistema e il relativo schema concettuale non li rappresenti in modo esauriente.

SECONDA FASE

Nella seconda fase vengono progettati i dati aggregati, utilizzando i requisiti disponibili per i dati aggregati e lo schema concettuale dei dati disaggregati, prodotto nella fase precedente.

Inizialmente, l'attività di progettazione ha lo scopo di:

- verificare se le aggregazioni desiderate sono compatibili con lo schema dei dati disaggregati;
- produrre lo schema concettuale delle aggregazioni compatibili;
- individuare eventuali aggregazioni di interesse non previste nei requisiti di partenza.

Sulla base dei risultati dell'attività precedente, sarà possibile individuare più precisamente le aggregazioni di interesse ed effettuare i conseguenti interventi sui requisiti.

In alcuni casi, si potrà anche intervenire a livello dei dati disaggregati operando le modifiche necessarie per rendere compatibili le aggregazioni.

A partire dai nuovi requisiti e dallo schema concettuale dei dati disaggregati (eventualmente modificato) si procederà quindi al raffinamento dello schema concettuale dei dati aggregati.

Il processo termina quando sono completamente individuate le aggregazioni di interesse, e lo schema concettuale è una rappresentazione esauriente di tali aggregazioni. Gli schemi definitivi prodotti nelle due fasi precedenti saranno chiamati rispettivamente **schema dei dati disaggregati** e **schema dei dati aggregati**.

TERZA FASE

Nella terza fase, infine, gli schemi concettuali finali potranno essere utilizzati per produrre gli schemi definitivi del questionario di rilevazione e delle tavole statistiche di interesse.

3.2. Secondo scenario di applicazione.

Nel secondo scenario di applicazione, illustrato in fig. 3.2, la progettazione concettuale si colloca in una fase intermedia del processo produttivo. Lo scenario è caratterizzato dai seguenti aspetti:

- i questionari di rilevazione sono definitivi;
- si ha un'idea di massima delle tavole statistiche da produrre, e dei dati aggregati di interesse per il sistema; i requisiti

sono costituiti da liste indicative e provvisorie di aggregazioni, espresse in linguaggio naturale.

Anche in questo scenario, l'applicazione della metodologia avviene in tre fasi.

PRIMA FASE

Nella prima fase viene prodotto uno schema concettuale dei dati disaggregati, che descrive il contenuto informativo dei questionari di rilevazione. In questo caso la metodologia svolge sostanzialmente un ruolo documentativo e il problema è quello di produrre uno schema concettuale che rappresenti in modo corretto tali contenuti; in genere la produzione dello schema avviene in modo iterativo, secondo una delle classiche strategie di progettazione (top-down, bottom-up, mista), confrontando di volta in volta gli schemi prodotti con i questionari di rilevazione. Lo schema finale prodotto in tale attività è chiamato **schema dei dati disaggregati**.

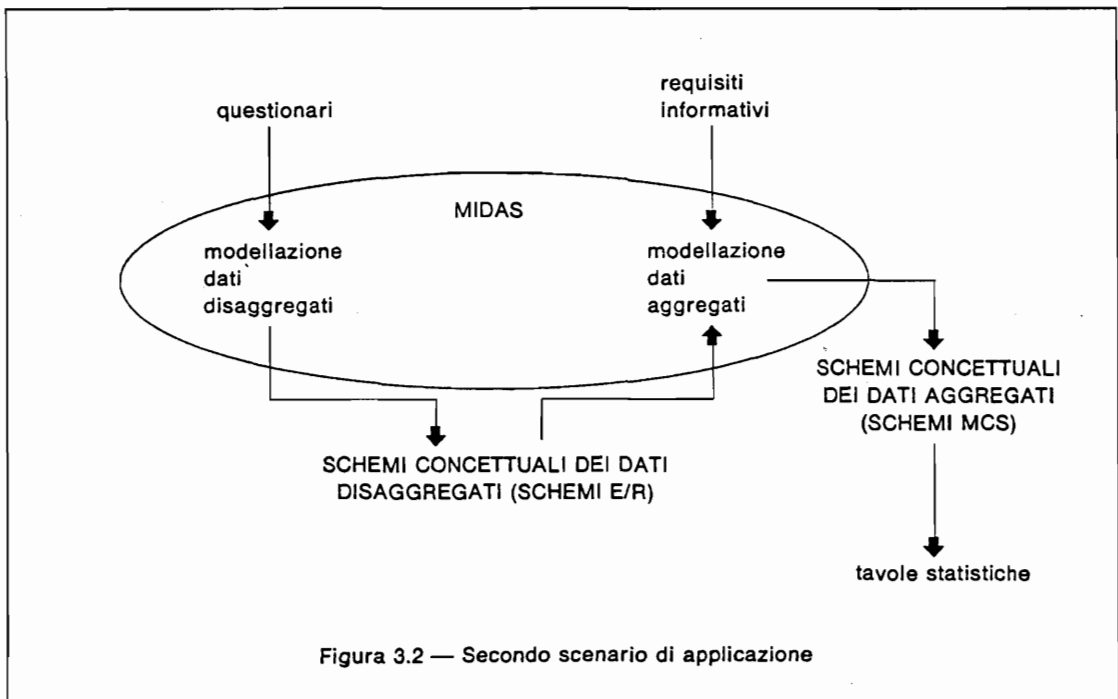
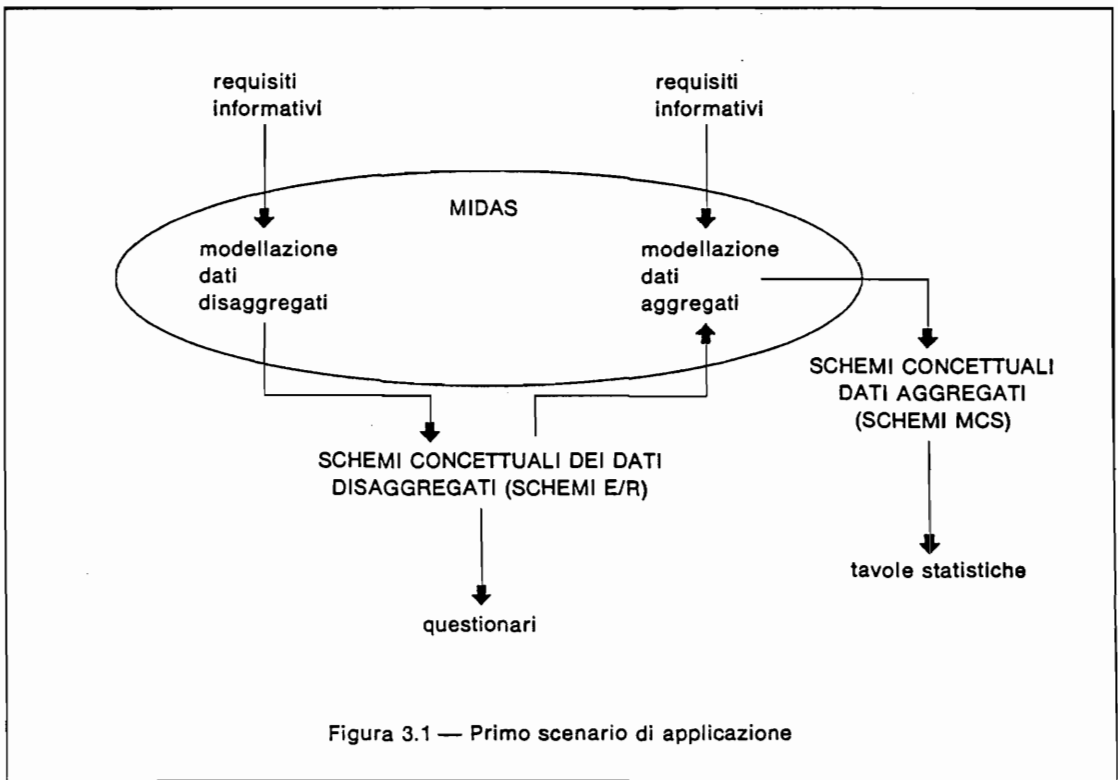
SECONDA FASE

La seconda fase è analoga a quella descritta nel primo scenario di applicazione. In essa vengono progettati i dati aggregati, utilizzando i requisiti relativi e lo schema dei dati disaggregati.

L'attività di progettazione ha inizialmente lo scopo di:

- verificare se le aggregazioni desiderate possono essere ottenute dai questionari di rilevazione;
- produrre lo schema concettuale delle aggregazioni ammissibili;
- individuare eventuali aggregazioni di interesse non previste nei requisiti.

Sulla base dei risultati dell'attività precedente, sarà possibile individuare con maggiore precisione le aggregazioni di interesse ed effettuare i conseguenti interventi sui requisiti. Nello scenario in esame, gli interventi sullo schema dei dati disaggregati mirano unicamente a ristrutturazioni dello stesso, al fine di ottenere una descrizione della realtà più funzionale alla progettazione dei dati aggregati; in nessun caso sarà quindi possibile ottenere le aggregazioni incompatibili con i questionari. La natura e i contenuti di tali ristrutturazioni saranno trattate più diffusamente nel capitolo 7.



A partire dai nuovi requisiti e dallo schema dei dati disaggregati, si procederà quindi al raffinamento dello schema dei dati aggregati.

Il processo termina quando sono completamente individuate le aggregazioni di interesse, e il relativo schema concettuale rappresenta esaurientemente tali aggregazioni. Lo schema definitivo prodotto durante tale attività è chiamato **schema finale dei dati aggregati**.

TERZA FASE

Nella terza fase, infine, gli schemi concettuali potranno essere utilizzati per produrre gli schemi definitivi delle tavole statistiche di interesse.

Lo scenario appena descritto è quello in cui tipicamente sono stati progettati nel passato i sistemi informativi dell'ISTAT; pertanto, quando nel seguito tratteremo più diffusamente la metodologia MIDAS, si farà riferimento ad esso.

3.3. Terzo scenario di applicazione.

In questo scenario di applicazione, illustrato in fig. 3.3, la progettazione concettuale si colloca alla fine del processo produttivo, quando i questionari di rilevazione e le tavole statistiche sono definiti; questionari e tavole rappresentano i requisiti del sistema informativo, relativamente ai dati disaggregati e ai dati aggregati.

La progettazione avviene in due fasi.

PRIMA FASE

Nella prima fase, l'attività da svolgere è la stessa descritta nello scenario precedente e consiste nella progettazione dello schema concettuale dei dati disaggregati. Anche in questo caso, la progettazione avviene in modo iterativo, secondo una delle strategie illustrate in precedenza.

SECONDA FASE

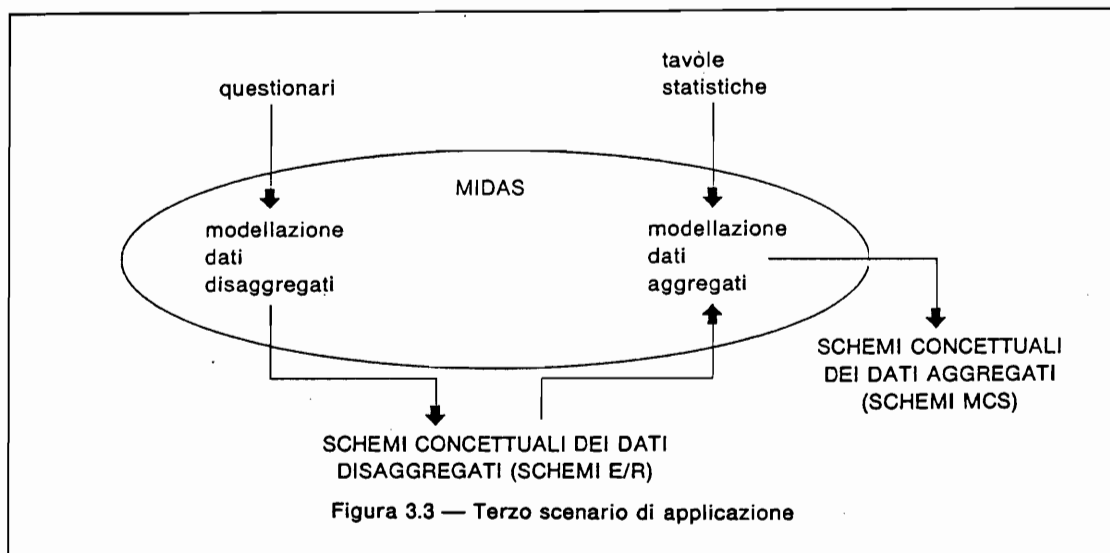
Nella seconda fase, che è relativa alla progettazione dei dati aggregati, gli input metodologici utilizzati sono le tavole statistiche e lo schema finale dei dati disaggregati.

L'attività ha lo scopo di:

- verificare se le tavole statistiche previste possono essere ottenute dai questionari di rilevazione;
- produrre lo schema concettuale delle tavole statistiche ammissibili.

La produzione dello schema concettuale avviene in due passi metodologici:

- inizialmente, vengono costruiti schemi concettuali parziali, relativi cioè alle singole tavole statistiche o ad insiemi limitati di esse;
- successivamente, si integrano gli schemi parziali nello schema finale dei dati aggregati; in tale attività è possibile, tra l'altro, valutare eventuali ridondanze nei dati.



4. Il modello Entita'/Relazione (E/R)

4.1. I modelli concettuali

Un **modello di dati** permette di descrivere i dati relativi ad un qualsiasi sistema informativo, e le relazioni intercorrenti tra essi, mediante determinate **strutture di rappresentazione**, il cui utilizzo e' sottoposto a regole definite nel modello stesso. Se la descrizione risultante e' indipendente dall'ambiente tecnologico nel quale i dati saranno trattati, il modello e' di tipo **concettuale**. In un modello concettuale i dati oggetto di descrizione possono essere dati elementari o disaggregati oppure dati aggregati (risultato di elaborazioni sui dati elementari o disaggregati). Le strutture di rappresentazione di un modello per i dati elementari o disaggregati sono in linea di principio notevolmente differenti da quelle di un modello per i dati aggregati, in quanto mentre nel primo caso le rappresentazioni sono direttamente riferibili al mondo reale (o meglio, agli oggetti del mondo reale rispetto ai quali i dati sono rilevati), nel caso di dati aggregati, oltre alle caratteristiche degli oggetti fonte dei dati, occorre poter anche rappresentare quelle delle elaborazioni eseguite per ottenerli (fig. 4.1).

L'uso dei modelli concettuali nelle metodologie di progettazione risponde all'esigenza di avere a disposizione strumenti di rappresentazione intermedi tra la realta' e la rappresentazione logico-fisica dei dati. Infatti, come gia' visto in precedenza, quest'ultima e' fortemente orientata alla struttura di memorizzazione utilizzata, e non possiede pertanto la capacita' di descrivere quegli aspetti della realta' non direttamente riconducibili ai dati. Classici modelli per la rappresentazione logico-fisica dei dati sono quelli gerarchico, reticolare e relazionale.

Come vedremo meglio in seguito, le strutture di rappresentazione si basano sul concetto di

astrazione. Un processo di astrazione e' un procedimento mentale attraverso il quale a partire da un insieme di oggetti considerati omogenei rispetto ad alcune proprieta' e' possibile giungere alla definizione di un nuovo oggetto. Partendo, ad esempio, dalle singole abitazioni, considerandone le proprieta' comuni rilevanti e tralasciando quelle non significative, e' possibile giungere alla definizione dell'oggetto ABITAZIONE: tale nuovo oggetto rappresenta l'insieme degli oggetti di partenza (le singole abitazioni).

I modelli concettuali possono essere valutati rispetto ad una serie di caratteristiche. Le piu' rilevanti appaiono:

- **l'espressivita'**, intesa come capacita' del modello di rappresentare esaurientemente le proprieta' degli oggetti di interesse;
- **la chiarezza**, cioe' la capacita' del modello di descrivere le proprieta' degli oggetti in modo familiare per l'utente e quindi a lui facilmente comprensibile;
- **l'univocita'**, cioe' la possibilita' di rappresentare gli oggetti simili della realta' di interesse con un unico tipo di struttura, cosi' che ogni struttura del modello abbia nello schema concettuale un suo specifico ruolo;
- **la formalita'**, cioe' la possibilita' di descrivere formalmente la sintassi e la semantica del modello in modo da non lasciare ambiguita' di interpretazione degli schemi concettuali;
- **la completezza**, intesa come capacita' del modello di esprimere tutte le proprieta' di interesse senza dover ricorrere a linguaggi non formali o al linguaggio naturale.

4.2. Le strutture di rappresentazione di un modello concettuale

In generale nei modelli concettuali identifichiamo tre tipi di strutture di rappresentazione:

1. strutture di classificazione;
2. strutture per la rappresentazione dei vincoli;
3. strutture per la rappresentazione delle operazioni.

Le strutture di classificazione permettono di rappresentare insiemi di oggetti della realta' attraverso l'individuazione delle proprieta' comuni per giungere alla definizione di **CLASSI (DI OGGETTI)**. Si possono individuare differenti procedimenti mentali (astrazioni) attraverso i quali insiemi di oggetti vengono riassunti in un concetto unificante.

La modalita' piu' semplice per definire una CLASSE e' quella di mettere in rilievo le proprieta' in comune dell'insieme degli oggetti (**astrazione di classificazione**). E' il caso di insiemi di persone che sono riassumibili nel concetto di persona, da cui la definizione della classe PERSONA. Vale la pena sottolineare che, come esiste una corrispondenza tra il concetto di persona e la relativa classe, e' possibile stabilire una corrispondenza tra gli individui che formano l'insieme delle persone (da cui il concetto di PERSONA) e gli oggetti rappresentati attraverso la classe che chiameremo **istanze della classe** e che nel nostro esempio saranno le istanze della classe persona (fig. 4.2).

Un secondo procedimento di astrazione e' quello mediante il quale si giunge alla definizione di un concetto (e quindi di una classe) a partire da altri concetti (altre classi) visti come componenti del primo (**astrazione di aggregazione**). Ad esempio, a partire dai concetti di MATRICOLA, NOME, COGNOME e SESSO si puo' definire la classe DIPENDENTE.

Un terzo meccanismo e' quello mediante il quale si puo' giungere alla definizione di una nuova classe sulla base delle caratteristiche comuni individuabili tra altre classi (**astrazione di generalizzazione**). E' il caso della definizione della classe PERSONA come generalizzazione delle classi UOMO e DONNA.

E' importante, a questo punto, sottolineare che lo SCHEMA CONCETTUALE, con le relative CLASSI, rappresenta una descrizione della realta' che si puo' considerare stabile nel tempo. I cambiamenti che avvengono nella realta' sono registrati solo dagli insiemi di istanze delle classi dello schema, rimanendo quindi invariata la concettualizzazione della realta' descritta attraverso le CLASSI. Ad esempio in un' azienda si puo' considerare, perlomeno in un periodo ragionevole, immutabile la concettualizzazione descritta dalla classe IMPIEGATO; cio' che variera' saranno gli individui che ne fanno parte e le relative occorrenze di una eventuale base di dati che li registra insieme alle loro caratteristiche.

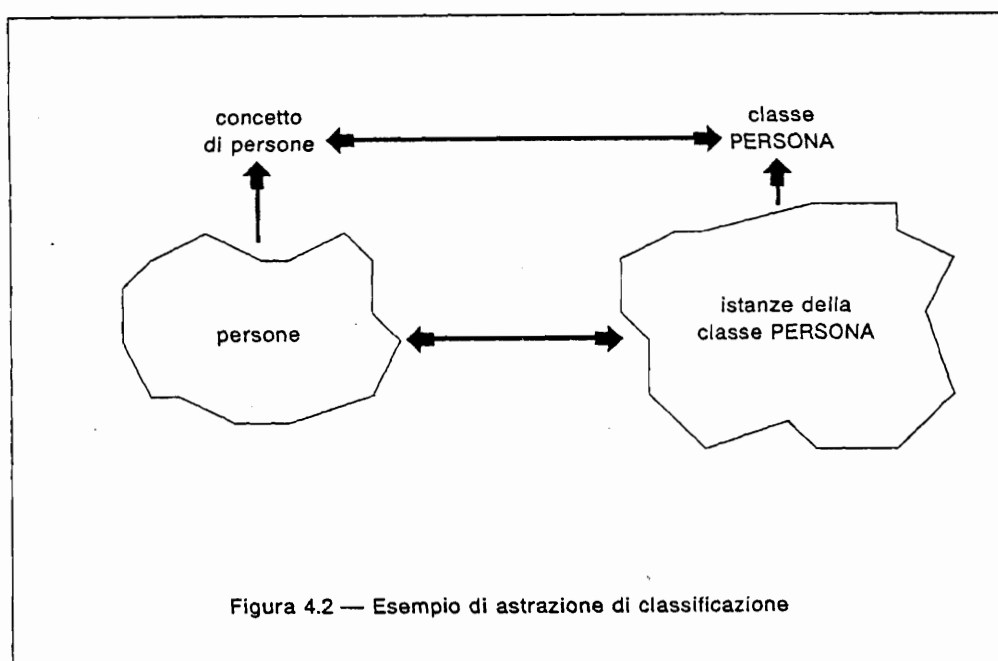
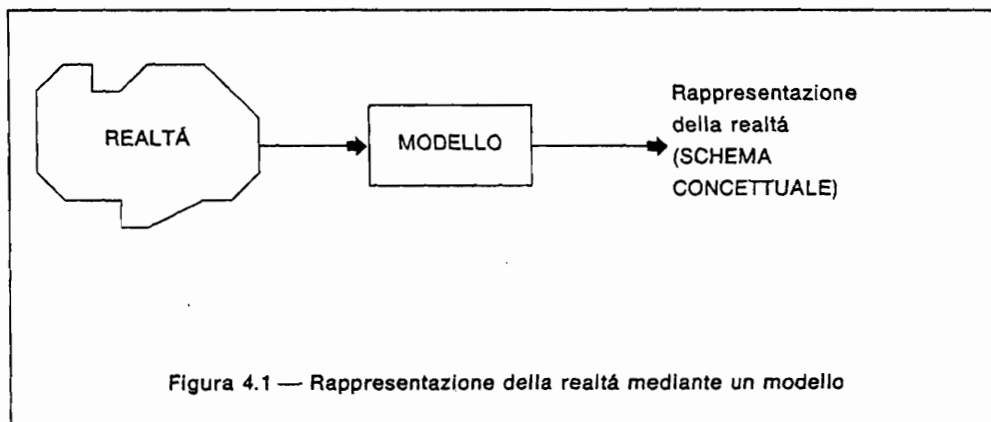
Le regole che gli oggetti definiti nelle classi debbono soddisfare per rappresentare in modo corretto la realta' d'interesse sono espresse, nello schema concettuale, mediante **VINCOLI DI INTEGRITA'**. Distinguiamo tra vincoli impliciti, propri delle strutture di classificazione di ogni modello, e vincoli espliciti che vengono espressi nel modello da specifiche strutture di rappresentazione simboliche o testuali.

L'utilizzo di strutture di rappresentazione delle **OPERAZIONI** permette di descrivere i cambiamenti che avvengono nella realta' tramite le modalita' con cui cambiano le occorrenze della base di dati che ne rappresentano la descrizione. Esempi di operazioni di questo tipo sono le operazioni di "inserimento", "variazione" e "cancellazione" che vanno sotto il nome di "operazioni di manipolazione".

Oltre al tipo di operazioni e' utile descrivere le strutture di controllo che permettono di definire il modo con cui si interviene sulle occorrenze della base di dati. Le strutture di controllo tipiche sono:

- la **sequenza**, l'esecuzione una dopo l'altra delle operazioni che seguono;
- l'**alternativa**, l'esecuzione di una o piu' operazioni condizionate dal verificarsi o meno di una certa condizione;
- il **ciclo**, la ripetizione ciclica, sotto condizione, di una certa operazione.

Il presente manuale non ha, tra le proprie finalita', quella approfondire le tematiche relative a tale argomento, che e' proprio piu' dell'analisi funzionale che di quella dei dati. Di conseguenza in questa sede la trattazione delle operazioni sara' limitata a cenni sulla loro



tipologia, senza approfondimenti particolari; l'argomento viene comunque trattato nell'Appendice B.

4.3. Il modello Entita'/Relazione

Il **MODELLO ENTITA'/RELAZIONE** (Entity Relationship Model), E/R nel seguito, viene utilizzato, nella metodologia MIDAS per la rappresentazione concettuale dei dati disaggregati. La sua presentazione si deve a P.P. Chen nel 1976, mentre la versione qui descritta risulta essere un arricchimento di quella prima versione.

L'esposizione che segue mostrerà come le strutture prima introdotte vengono utilizzate all'interno del modello E/R. Tale modello viene frequentemente presentato facendo riferimento ai simboli utilizzati per la **rappresentazione grafica**, che sono riportati nella fig. 4.3; nella Appendice A è definita una **descrizione linguistica**, che può essere utilizzata per la documentazione della progettazione concettuale.

4.3.1. Strutture di classificazione

Le strutture di classificazione mediante le quali vengono rappresentati gli oggetti del mondo reale nel MODELLO E/R sono:

- l'**entita'** ;
- la **relazione** .

Per entita' si intendono classi di oggetti con caratteristiche comuni; è il caso dell'entita' DIPENDENTE che individua tutte le persone che hanno un rapporto di lavoro all'interno di un'azienda (fig. 4.4). I singoli dipendenti rappresenteranno le istanze di quella entita'.

Le proprietà elementari di una entita' sono dette **attributi**; alla precedente entita' DIPENDENTE si possono associare come attributi la MATRICOLA, il NOME, il COGNOME, il SESSO. I singoli attributi possono assumere diversi valori entro un dominio predefinito: il numero di matricola potrà essere un codice alfanumerico definito dall'azienda, il sesso avrà valori MASCHIO o FEMMINA, ecc. (fig. 4.4).

Chiameremo **attributo multiplo** quell'attributo al quale per ogni istanza dell'entita' è possi-

bile associare più valori. Ad esempio l'attributo LINGUE STRANIERE è multiplo in quanto ogni dipendente può conoscere più di una lingua straniera (fig. 4.4).

Chiameremo inoltre **attributo aggregato** un attributo composto da due o più attributi semplici. Ad esempio INDIRIZZO è attributo aggregato, in quanto composto dagli attributi semplici VIA, NUMERO, CITTA', CAP.

Il concetto di attributo permette di giungere alla definizione dell'entita' attraverso un processo di aggregazione delle sue proprietà secondo il meccanismo precedentemente descritto come astrazione di aggregazione; nel nostro caso la aggregazione delle proprietà MATRICOLA, NOME, COGNOME, SESSO individua la classe DIPENDENTE.

Una **relazione** rappresenta il legame logico tra due o più entita'; l'appartenenza dei dipendenti ad un determinato reparto potrà essere rappresentato dalla relazione LAVORA IN tra l'entita' DIPENDENTE e l'entita' REPARTO (fig. 4.5). Le istanze della classe descritta come relazione saranno combinazioni di istanze delle entita' che essa connette; nell'esempio una generica istanza della relazione LAVORA IN sarà individuata dalla coppia (dipendente, reparto di appartenenza).

Anche la relazione ha propri attributi, nel nostro esempio potrebbe essere PERIODO DI APPARTENENZA, cioè una proprietà non direttamente riferibile né all'entita' DIPENDENTE, né a quella REPARTO, ma alla loro relazione LAVORA IN. È da notare che l'attributo usato in questo caso è definibile come attributo aggregato essendo composto dai due attributi DATA INIZIO DI APPARTENENZA e DATA FINE DI APPARTENENZA (fig. 4.5).

Due entita' si trovano in una **relazione sottoinsieme** se ogni istanza dell'entita' dipendente (detta entita' "figlia") è anche istanza dell'entita' superiore (detta entita' "padre"), mentre non è vero il contrario: cioè non tutte le istanze dell'entita' "padre" sono anche istanze dell'entita' "figlia". Ne consegue che tutte le proprietà (attributi, relazioni, vincoli, ecc.) della entita' "padre" sono ereditate da quella definita "figlia". Quanto detto è esemplificato dalla figura 4.6 in cui è rappresentata la relazione tra l'entita' DIPENDENTE e l'entita' IMPIEGATO: è evidente che non tutti i dipendenti sono identificabili nella sola classe IMPIEGATO, che potrà quindi avere delle sue proprietà non generalizzabili nella entita' di

ordine superiore; e' il caso, ad esempio, dell'attributo LIVELLO IMPIEGATIZIO.

Un caso particolare della relazione di sottoinsieme e' la **gerarchia di generalizzazione**, che si ha nel caso in cui e' possibile definire, relativamente ad una stessa entita' "padre", piu' entita' "figlie" che godono della seguente proprieta': ogni istanza di un'entita' "figlia" e' anche istanza dell'entita' "padre" (come nel caso piu' generale); in piu' ogni istanza dell'entita' "padre" e' anche istanza di una ed una sola entita' "figlia". In pratica l'insieme delle istanze delle entita' figlie coincide con l'insieme delle istanze dell'entita' padre; inoltre le entita' figlie non hanno tra di loro istanze comuni.

Un esempio e' descritto nella fig. 4.7: e' definita una gerarchia di generalizzazione tra l'entita' DIPENDENTE e le entita' OPERAIO, IMPIEGATO e DIRIGENTE. Ricorrono le proprieta' enunciate: un operaio (o un impiegato o un dirigente) e' anche un dipendente; ogni dipendente e' sicuramente un operaio, oppure un impiegato, oppure un dirigente, senza possibilita' di essere due o tre cose insieme. Anche in questo caso le proprieta' dell'entita' "padre" sono ereditate dalle entita' "figlie". Le entita' OPERAIO, IMPIEGATO, DIRIGENTE avranno come attributi, oltre ai propri, gli attributi dell'entita' DIPENDENTE: MATRICOLA, COGNOME, NOME e LINGUE CONOSCIUTE.

4.3.2. Vincoli di integrita'

Nella rappresentazione dei **VINCOLI ESPLICITI** del MODELLO E/R distinguiamo:

- vincoli di identificazione ;

- vincoli di cardinalita' .

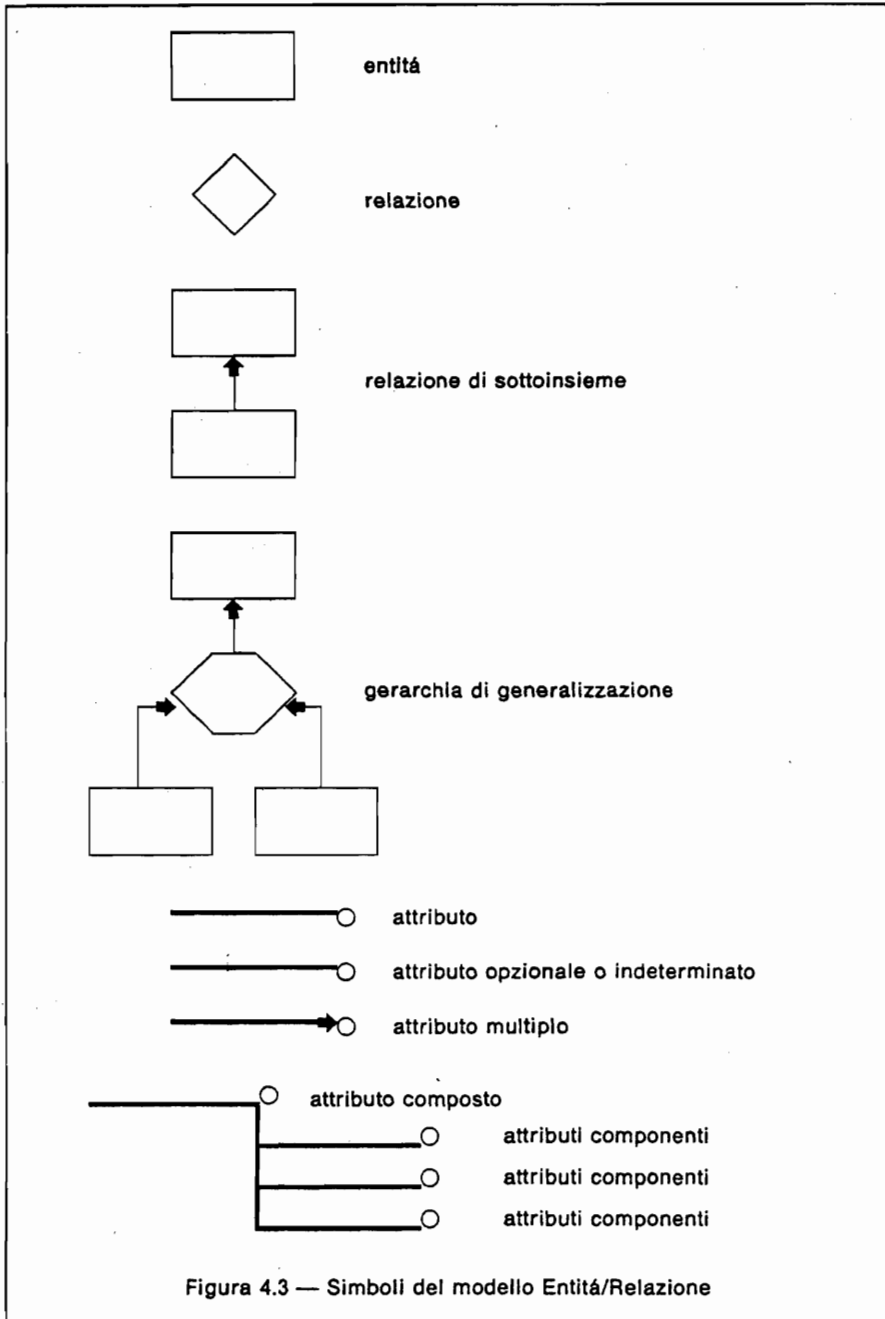
Chiameremo **identificatore** l'insieme di attributi mediante il quale e' possibile identificare univocamente le singole istanze dell'entita'. Ad esempio per l'entita' DIPENDENTE l'identificatore e' individuato dall'attributo MATRICOLA (fig. 4.4).

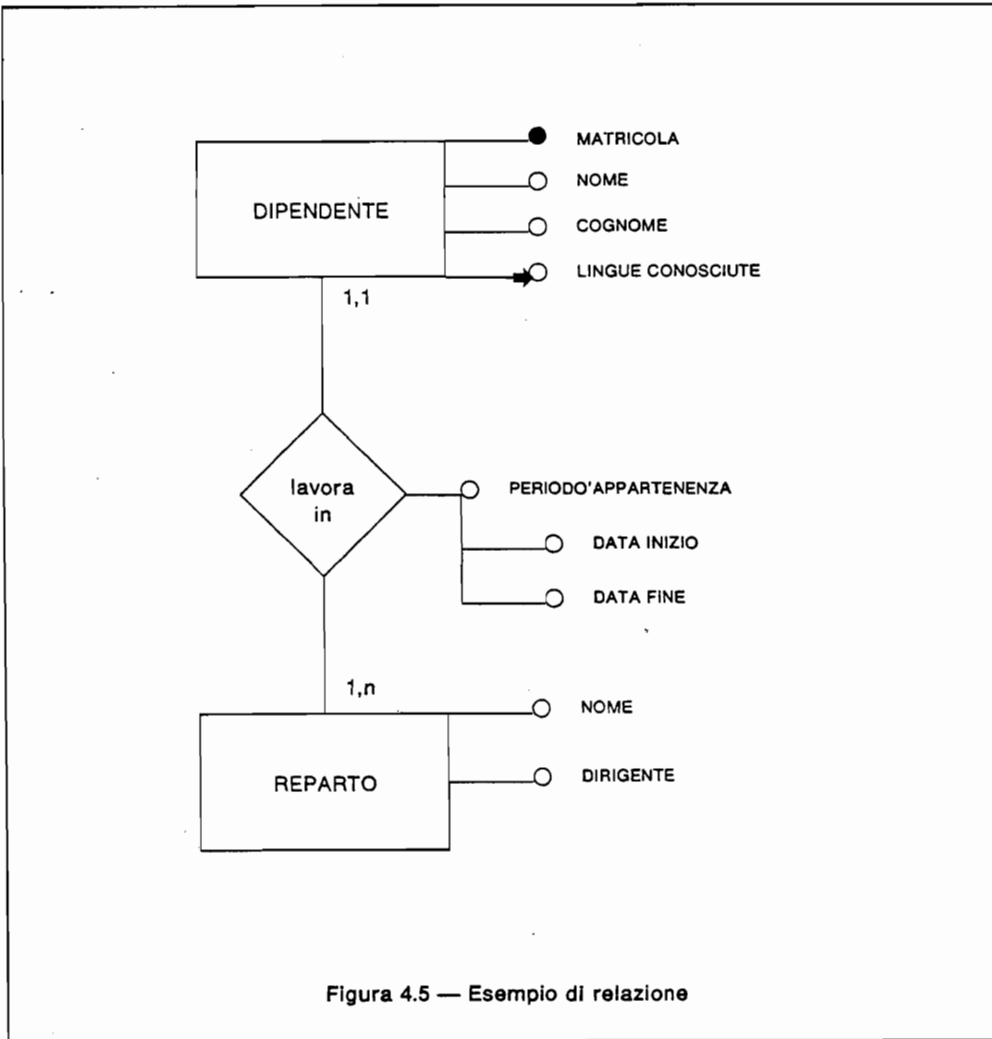
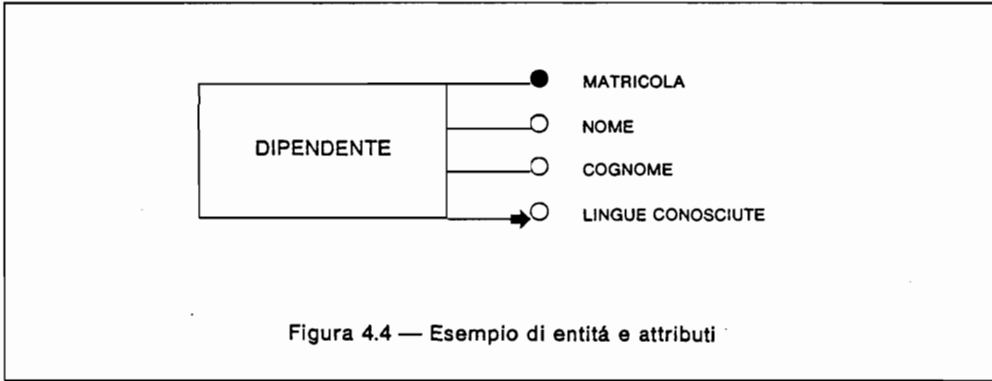
Per vincoli di identificazione si intendono le regole che permettono di individuare univocamente le istanze di una classe: in ogni istanza di una entita' deve essere presente un diverso valore dell'identificatore.

Nel caso dell'entita' IMPIEGATO cio' significa che l'attributo identificatore MATRICOLA deve assumere valori differenti per ogni impiegato dell'azienda.

Per **cardinalita' minima e massima** di una entita' coinvolta in una relazione si intende il minimo ed il massimo numero di volte che ogni sua istanza puo' essere associata con istanze delle altre entita' coinvolte nella relazione. La cardinalita' 0 indichera' che puo' esistere una istanza della entita' non coinvolta nella relazione. Valori superiori, da 1 a n, indicano che una o piu' istanze dell'entita' sono coinvolte nella relazione. Nell'esempio della fig. 4.5, l'espressione "1,1" associata a DIPENDENTE indica che ogni istanza di tale entita' (cioe' ogni dipendente) e' associata ad uno ed un solo reparto (la cardinalita' minima in questo caso coincide con quella massima), mentre l'espressione "1,n" associata a REPARTO esprime il fatto che ogni reparto ha almeno un dipendente, e puo' averne piu' di uno.

Attribuire dei vincoli di cardinalita' significa definire in maniera precisa il numero di istanze, minimo e massimo, che legano una relazione con un'entita'.





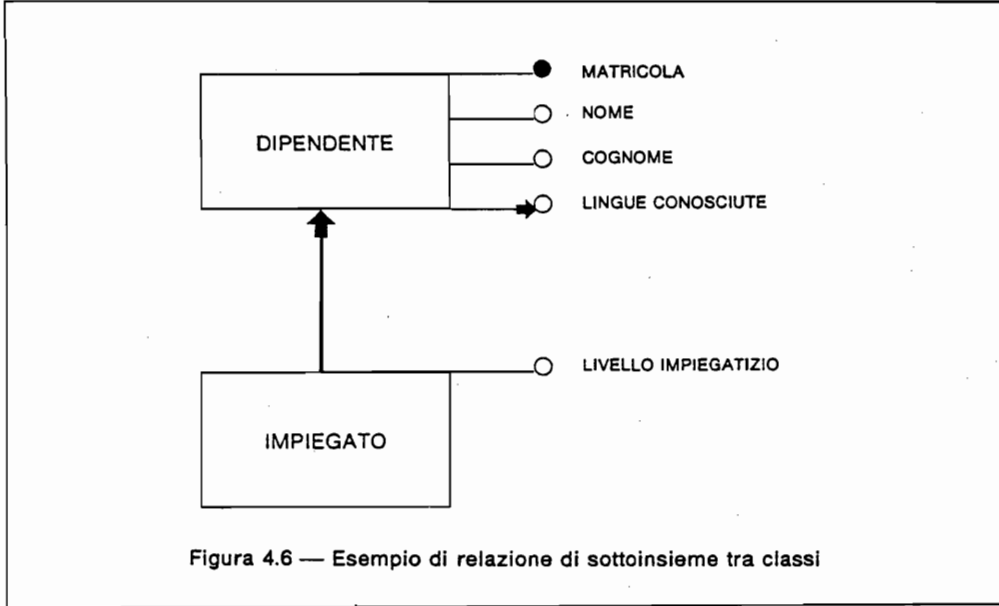


Figura 4.6 — Esempio di relazione di sottoinsieme tra classi

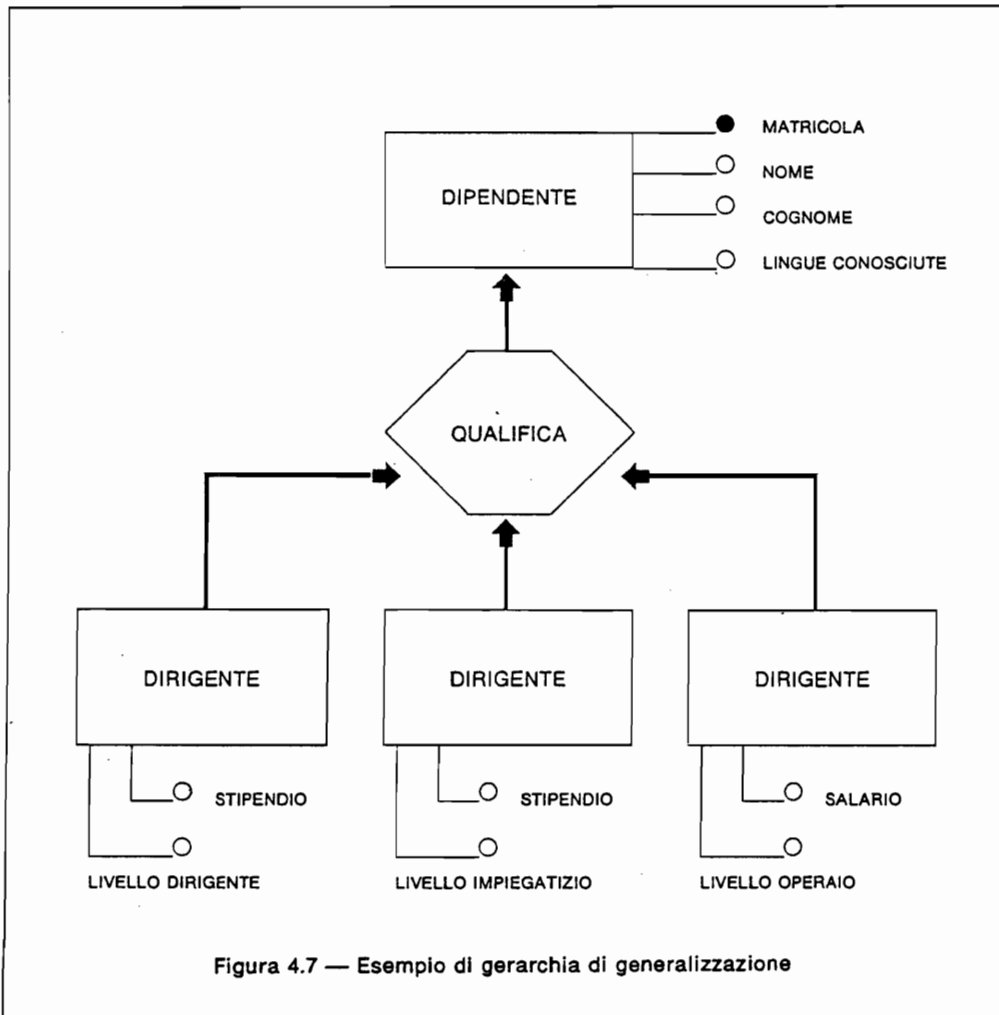


Figura 4.7 — Esempio di gerarchia di generalizzazione

5. La metodologia per la costruzione degli schemi dei dati disaggregati

In questa sezione viene mostrata con un esempio concreto, relativo al SIL - Sistema Informativo del Lavoro, la metodologia per la progettazione di dati disaggregati; in questa applicazione assumiamo come requisiti a disposizione i questionari di rilevazione.²

La metodologia complessiva per la progettazione dei dati disaggregati e' composta di quattro fasi, ognuna delle quali richiede una serie di attivita'; nel seguito, dopo una schematica presentazione dei passi e delle azioni della metodologia, questi verranno analizzati in dettaglio.

La strategia da seguire consiste nel progettare uno schema concettuale per ogni area del questionario di rilevazione, e successivamente nell'integrare tra di loro i diversi schemi.

La metodologia puo' essere cosi' schematizzata:

1. **ANALISI DEL QUESTIONARIO:** per ogni questionario vengono selezionate le aree omogenee e viene costruito l'**albero delle aree**. Vengono quindi analizzati i concetti contenuti in ogni area.
2. **DISEGNO CONCETTUALE:** ai concetti definiti in ogni area nella fase precedente vengono assegnate le strutture di rappresentazione del modello E/R, producendo in tal modo gli schemi concettuali di area.

3. **INTEGRAZIONE:** viene scelta una prima coppia di schemi concettuali e ne viene eseguita l'integrazione, verificando e risolvendo eventuali conflitti. Il nuovo schema prodotto (unificazione dei due schemi originari) viene quindi a sua volta integrato con un altro tra gli schemi ottenuti nella fase precedente. Il processo viene ripetuto finche' non vengono considerati tutti gli schemi: a questo punto tutti i concetti sono contenuti nello schema finale.

4. **VERIFICHE FINALI:** vengono effettuati, se possibile, eventuali arricchimenti e ristrutturazioni dello schema globale ottenuto nel passo precedente, verificando ed eliminando le eventuali ridondanze.

Una rappresentazione dei passi, coi relativi ingressi ed uscite, e' riportata nelle figg. 5.1a, 5.1b e 5.1c.

Analizziamo ora separatamente le quattro fasi, servendoci dell'esempio del SIL.

5.1. Analisi del questionario

Ogni questionario e' divisibile in aree, in zone cioe' che riportano informazioni relative ad uno stesso argomento. Consideriamo ad esempio un frammento del questionario della

² Il SIL e' un Sistema Informativo che ha l'obiettivo di unificare tutti i dati disponibili sul fenomeno "lavoro" presso vari operatori istituzionali, come l'Istat, il Ministero del Lavoro, l'INPS etc.; il Sistema, progettato e realizzato all'Istat sulla base delle indicazioni di numerose Commissioni, dovra' in futuro essere 'trasportato' anche presso il Centro Elettronico del Ministero del Lavoro. Il primo nucleo del Sistema Informativo del Lavoro comprendera' i dati della Rilevazione trimestrale dell'Istat sulle Forze Lavoro e i dati che riguardano il lavoro rilevati nel Censimento della Popolazione del 1981; ad essi si riferiranno gli esempi forniti in questo capitolo.

ANALISI DEL QUESTIONARIO

Per ogni questionario:

1. seleziona aree omogenee e costruisci l'albero delle aree
2. per ogni area estrai ed analizza i concetti elementari

DISEGNO CONCETTUALE

Per ogni area costruisci lo schema concettuale di area, assegnando le strutture di rappresentazione del modello E/R ai concetti, finché questi sono tutti espressi nello schema

INTEGRAZIONE

Scegli uno schema concettuale di area

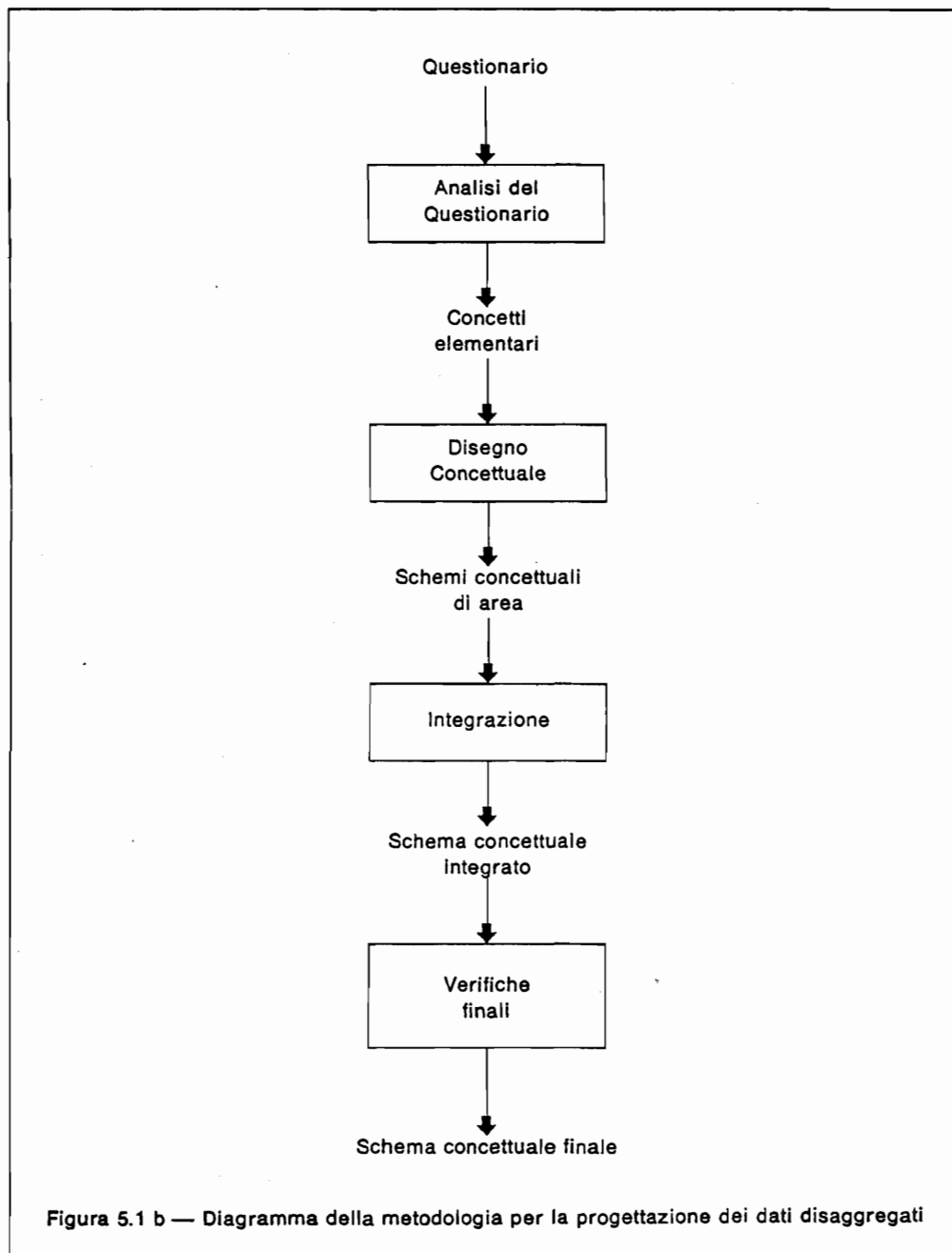
1. scegli un nuovo schema concettuale di area
2. verifica e risolvi i conflitti di nome e di tipo tra i due schemi
3. unifica i due schemi in un unico schema concettuale

Considera lo schema prodotto al punto 3. e ripeti i passi 1—3 finché tutti i concetti sono descritti nello schema

VERIFICHE FINALI

**Effettua eventuali arricchimenti e ristrutturazioni
Verifica e rimuovi ridondanze**

Figura 5.1 a — Sequenza della metodologia per la progettazione dei dati disaggregati



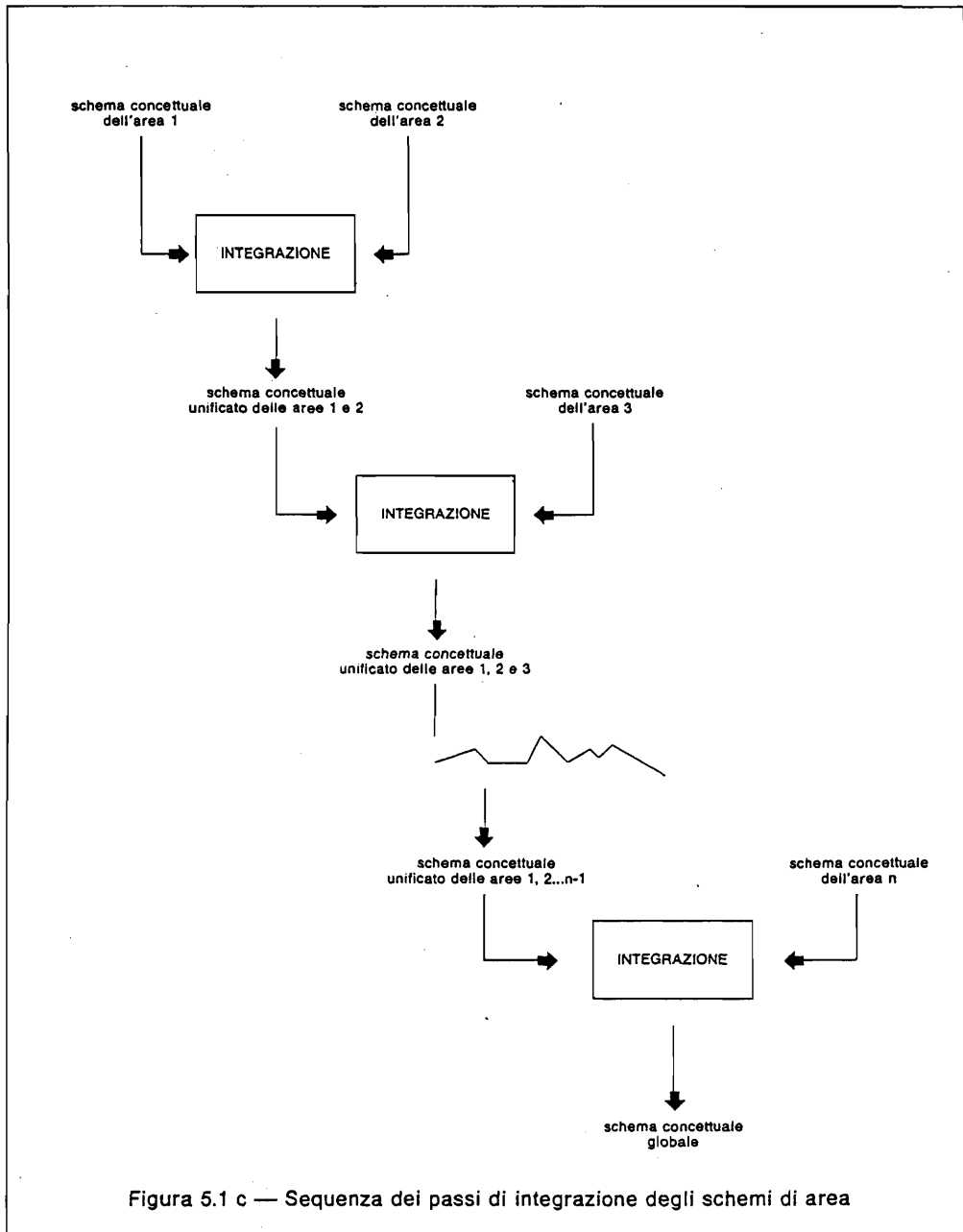


Figura 5.1 c — Sequenza dei passi di integrazione degli schemi di area

Rilevazione trimestrale delle Forze Lavoro (fig. 5.2).

Le informazioni rilevate si riferiscono a tre diversi aspetti del mondo reale, che corrispondono a diverse aree del modello:

1. Persone
2. Occupati (persone che hanno lavorato nella settimana di riferimento)
3. Non occupati

Le diverse aree possono a loro volta essere suddivise in sottoaree. Ad esempio nell'area degli occupati si possono distinguere le sottoaree:

1. persone con seconda attivita';
2. persone con attivita' temporanea;
3. persone che lavorano part-time;

mentre nell'area dei non occupati abbiamo le sottoaree:

1. non occupati con precedenti attivita' lavorative;
2. non occupati senza precedenti attivita' lavorative.

Dopo aver scomposto il questionario di interesse in aree e sottoaree, si passa ad una prima provvisoria specificazione dei principali concetti presenti nel questionario e dei legami che sussistono tra di essi. Questa attivita' si concretizza nella costruzione dell'**albero delle aree** riportato nella fig. 5.3 (tra parentesi vengono riportati i quesiti del questionario corrispondenti ad ogni nodo dell'albero).

Vedremo come questa prima strutturazione dei concetti influenzerà in modo determinante la costruzione dello schema concettuale complessivo relativo al frammento di questionario considerato.

5.2. Disegno concettuale dei dati disaggregati

Questa fase ha come ingresso la struttura per aree e sottoaree prodotta nella fase precedente. Per ogni area viene costruito lo schema concettuale di area assegnando ad ogni con-

cetto la relativa struttura di rappresentazione del modello E/R.

Così, ad esempio, nella fig. 5.4a è riportato lo schema E/R relativo alla prima area (dati sulle persone): al concetto di persona è assegnata l'entità PERSONA, ai concetti di stato civile e condizione prevalente sono assegnati gli attributi STATO CIVILE e CONDIZIONE PREVALENTE dell'entità PERSONA.

In modo analogo, nelle figg. 5.4b e 5.4c sono riportati gli schemi E/R relativi alle altre due aree: si noti come la presenza di sottoaree determini in questi casi l'uso delle relazioni di sottoinsieme nei corrispondenti schemi concettuali (ad esempio, nella fig. 5.4b le entità PERSONA CHE LAVORA PART-TIME, PERSONA CON ATTIVITA' TEMPORANEA e PERSONA CON SECONDA ATTIVITA' sono tutte legate da una relazione di sottoinsieme all'entità PERSONA CHE LAVORA).

Per quanto riguarda la scelta delle strutture di rappresentazione per i concetti elementari, essa va fatta ricordando le proprietà delle strutture del modello E/R; in generale si può osservare che:

1. Un concetto sarà tipizzato come entità tutte le volte che è dotato di proprietà elementari e di almeno una tra le seguenti caratteristiche:
 - si trova in relazione con altri concetti;
 - si trova in rapporto di generalizzazione con altri concetti.
2. Un concetto sarà tipizzato come attributo semplice quando compare solo come proprietà elementare di un altro concetto.
3. Un concetto sarà tipizzato come attributo aggregato quando da una parte è aggregato di altri concetti, e dall'altra questa è l'unica proprietà che lo caratterizza nel modello, non comparando in relazione con altri concetti: ad esempio nella fig. 5.4c INFORMAZIONI SULL'ULTIMA OCCUPAZIONE è attributo aggregato degli attributi semplici POSIZIONE NELLA PROFESSIONE e BRANCA DI ATTIVITA' ECONOMICA.
4. Un concetto sarà tipizzato come relazione quando esso compare nel questionario come legame logico tra due o più concetti, rappresentati nello schema da entità.

9 - STATO CIVILE

- Celibe o nubile 1
- Coniugato 2
- Vedovo 3
- Separato, divorziato, già coniugato 4

10 - CONDIZIONE UNICA O PREVALENTE

10.1 - Qual è attualmente la sua condizione?

- Occupato 1
- In cerca di nuova occupazione 2
- In cerca di prima occupazione 3
- In servizio di leva 4
- Casalinga 5
- Studente 6
- Inabile al lavoro 7
- Persona ritirata dal lavoro 8
- Altra condizione (benestante, anziano e simili) 9

10.2 - Ha effettuato ore di lavoro nella settimana di riferimento qualunque sia la condizione dichiarata al q. 10.1?

- Sì 1
- No 2

RISPONDERE SEMPRE PER GLI OCCUPATI (q. 10.1 = 1) E PER TUTTE LE ALTRE PERSONE CHE HANNO EFFETTUATO ORE DI LAVORO NELLA SETTIMANA DI RIFERIMENTO (q. 10.2 = 1)

11 - ATTIVITÀ LAVORATIVA PRINCIPALE O UNICA

11.1 - Ore di lavoro settimanali

- EFFETTIVAMENTE svolte nella settimana di riferimento

- svolte ABITUALMENTE

ATTENZIONE: se le ore effettive sono uguali alle ore abituali passare al q. 11.3

11.2 - Motivo per cui le ore effettive sono diverse dalle ore abituali

Le ore effettive sono superiori alle ore abituali:

- Orario di lavoro variabile 01
- Altri motivi 02

Le ore effettive sono inferiori alle ore abituali:

- Cattivo tempo 03
- Ridotta attività economica dell'azienda per motivi tecnici od economici 04
- Conflitti di lavoro 05
- Frequenza di corsi scolastici o di formazione professionale 06
- Orario di lavoro variabile 07
- Malattia, infortunio, inabilità al lavoro 08
- Motivi personali o familiari (compresa la maternità) 09
- Ferie annuali 10
- Festività infrasettimanali 11
- Inizio dell'attività nella settimana 12
- Cessazione dell'attività nella settimana 13
- Mancanza di occasioni di maggior lavoro, fasi stagionali, ecc. 14

11.3 - Posizione nella professione

- Imprenditore 1
- Libero professionista 2
- Lavoratore in proprio 3
- Coadiuvante 4
- Dirigente 5
- Impiegato o intermedio 6
- Operaio, subalterno e assimilati 7
- Apprendista 8
- Lavorante a domicilio per conto d'impresa 9

11.4 - Branchia di attività economica

- Agricoltura, foreste, caccia e pesca 01
- Energia ed acqua 02
- Estrazione e trasformazione di minerali non energetici e prodotti derivati, industria chimica 03
- Industria della trasformazione dei metalli, meccanica di precisione 04
- Officine e botteghe di riparazione di beni di consumo (auto, calzature, elettrodomestici, ecc) 05
- Altre industrie manifatturiere 06
- Costruzioni ed installazione impianti 07
- Commercio, alberghi e pubblici esercizi 08
- Trasporti e comunicazioni 09
- Credito, assicurazioni, servizi prestati alle imprese 10
- Pubblica amministrazione, forze armate, previdenza 11
- Altri servizi, istituzioni religiose, enti stranieri ed internazionali 12

11.5 - Tempi di lavoro

- L'attività è esercitata a tempo pieno 1

L'attività è esercitata a tempo parziale per uno dei seguenti motivi:

- frequenti corsi scolastici o di formazione professionale 2
- malattia o invalidità 3
- non ha potuto trovare un lavoro a tempo pieno 4
- non desidera un lavoro a tempo pieno 5
- altri motivi 6

11.6 - Carattere permanente o meno dell'occupazione

- Ha un'occupazione permanente 1

Ha un'occupazione temporanea perché:

- il contratto di lavoro riguarda un periodo di formazione (apprendistato, tirocinio, assistente di ricerca, ecc.) 2
- non ha potuto trovare un lavoro permanente 3
- non desidera un lavoro permanente 4
- altri motivi 5

12 - EVENTUALE SECONDA ATTIVITÀ LAVORATIVA

- Svolge soltanto un'attività 1

Possiede anche una seconda attività che:

- ha svolto nella settimana di riferimento 2
- ha svolto o prevede di svolgere nel corso dell'anno 3

Se è stato barrato il codice 2 Indicare anche:

- ore di lavoro svolta nella settimana di riferimento nella seconda attività.

13 - PRECEDENTI LAVORATIVI

RISPONDERE SOLO PER LE PERSONE CHE SI SONO DICHIARATE NON OCCUPATE (q. 10.1 = da 2 a 9) E CHE NELLA SETTIMANA DI RIFERIMENTO NON HANNO EFFETTUATO ORE DI LAVORO (q. 10.2 = 2)
PER TUTTI GLI ALTRI PASSARE AL Q. 14

13.1 - Ha mai lavorato in passato?

- Sì 1
- No 2

→ passare al q. 13.2 → passare al q. 14

13.2 - Da quanti mesi ha lasciato l'ultima occupazione?

- Indicare il numero dei mesi su due cifre antepo-
nendo zero se occorre
(per più di 35 mesi indicare sempre 36)

ATTENZIONE: per le persone che hanno lasciato l'occupazione da meno di 36 mesi (q. 13.2 = da 01 a 35) passare al q. 13.3
Per gli altri passare al q. 14

13.3 - Motivo per cui ha lasciato l'occupazione

- Licenziamento 1
- Fine di un lavoro a tempo determinato 2
- Dimissioni o cessazione di attività in proprio 3
- Pensionamento anticipato per motivi economici 4
- Ritiro dal lavoro per motivi di salute o maternità 5
- Pensionamento per raggiunti limiti di età o per altri motivi diversi da quelli economici o di salute 6
- Servizio di leva 7

13.4 - Informazioni sull'ultima occupazione posseduta

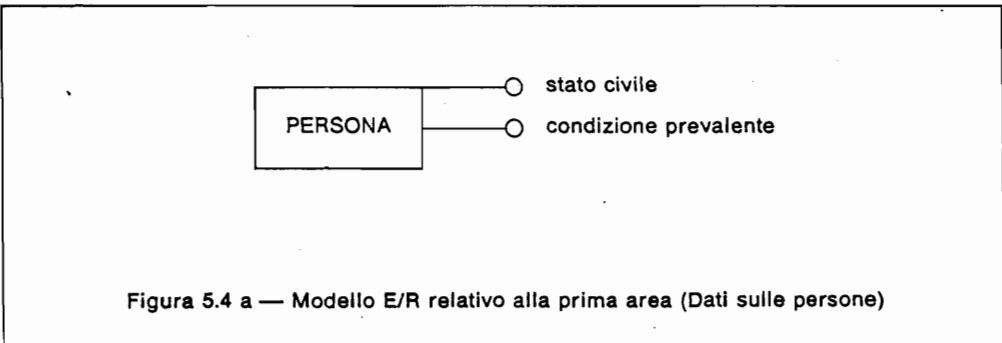
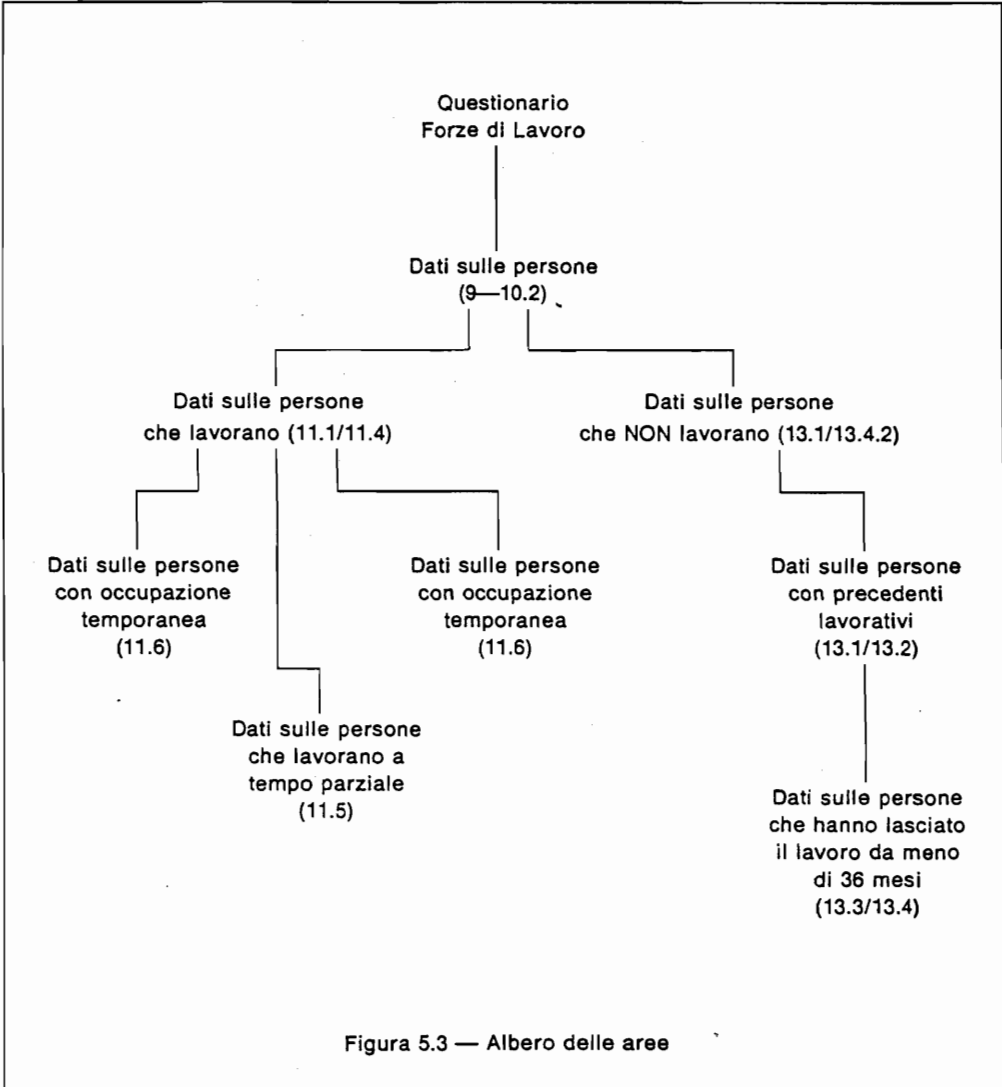
13.4.1 - POSIZIONE NELLA PROFESSIONE

- Usare i codici indicati al q. 11.3

13.4.2 - BRANCA DI ATTIVITÀ ECONOMICA

- Usare i codici indicati al q. 11.4

Figura 5.2 — Frammento del questionario relativo alla rilevazione trimestrale delle Forze Lavoro



concetti e' il fatto che i concetti abbiano nei due schemi le stesse proprieta'.

Ritornando al caso in esame, il passo di integrazione ha come ingressi gli schemi delle figg. 5.4a, 5.4b e 5.4c e produce, come risultato, lo schema di fig. 5.5, che contiene, al suo interno, tutti i concetti dello schema che si intendeva rappresentare.

5.4. Verifiche finali.

Scopo dell'ultimo passo e' di analizzare lo schema globale in modo tale da individuare possibili miglioramenti della sua qualita'. Le qualita' di uno schema concettuale cui possiamo essere interessati sono varie, ma in genere le piu' importanti sono:

1. la **correttezza**: nello schema ogni proprieta' del mondo reale e' riferita al concetto corretto;
2. la **completezza**: nello schema sono rappresentate tutte e solo le proprieta' del mondo reale di interesse;
3. la **leggibilita'**: la scelta delle strutture linguistiche e' stata fatta in modo da semplificare al massimo la comprensione dello schema;
4. la **minimalita'**: ogni concetto del mondo reale e' espresso una sola volta nel modello.

Analizzando lo schema di fig. 5.5 si nota che il concetto "lavoro" e' comune a molte delle entita' in esso contenute. E' allora conveniente introdurre l'entita' corrispondente (vedi fig. 5.6) che viene collegata alle entita' PERSONA CHE LAVORA (mediante la relazione SVOLGE) e PERSONA CON PRECEDENTI LAVORATIVI (mediante la relazione HA SVOLTO). L'attributo BRANCA DI ATTIVITA' ECONOMICA e' assegnato alla nuova entita' LAVORO, e viene nel contempo tolta da PERSONA CHE LAVORA e da PERSONA CON PRECEDENTI LAVORATIVI.

In quest'ultima entita' scompare l'attributo aggregato INFORMAZIONI SULL'ULTIMA OCCUPAZIONE, in quanto rimane il solo attributo componente POSIZIONE NELLA PROFESSIONE.

Un'ultima notazione sulle cardinalita'. In entrambe le relazioni SVOLGE e HA SVOLTO, l'entita' LAVORO ha sempre cardinalita' minima 1 (almeno una persona svolge o ha svolto un certo lavoro) e massima n (piu' persone possono svolgere o hanno svolto un certo lavoro). L'entita' PERSONA CHE LAVORA ha cardinalita' minima e massima rispettivamente 1 e 2 (i dati richiesti ad una persona che lavora sono relativi ad almeno una occupazione, ed eventualmente alla seconda). Infine in PERSONA CON PRECEDENTI LAVORATIVI la cardinalita' minima coincide con quella massima ed e' pari a 1 (vengono richiesti i dati relativi solo all'ultima occupazione svolta).

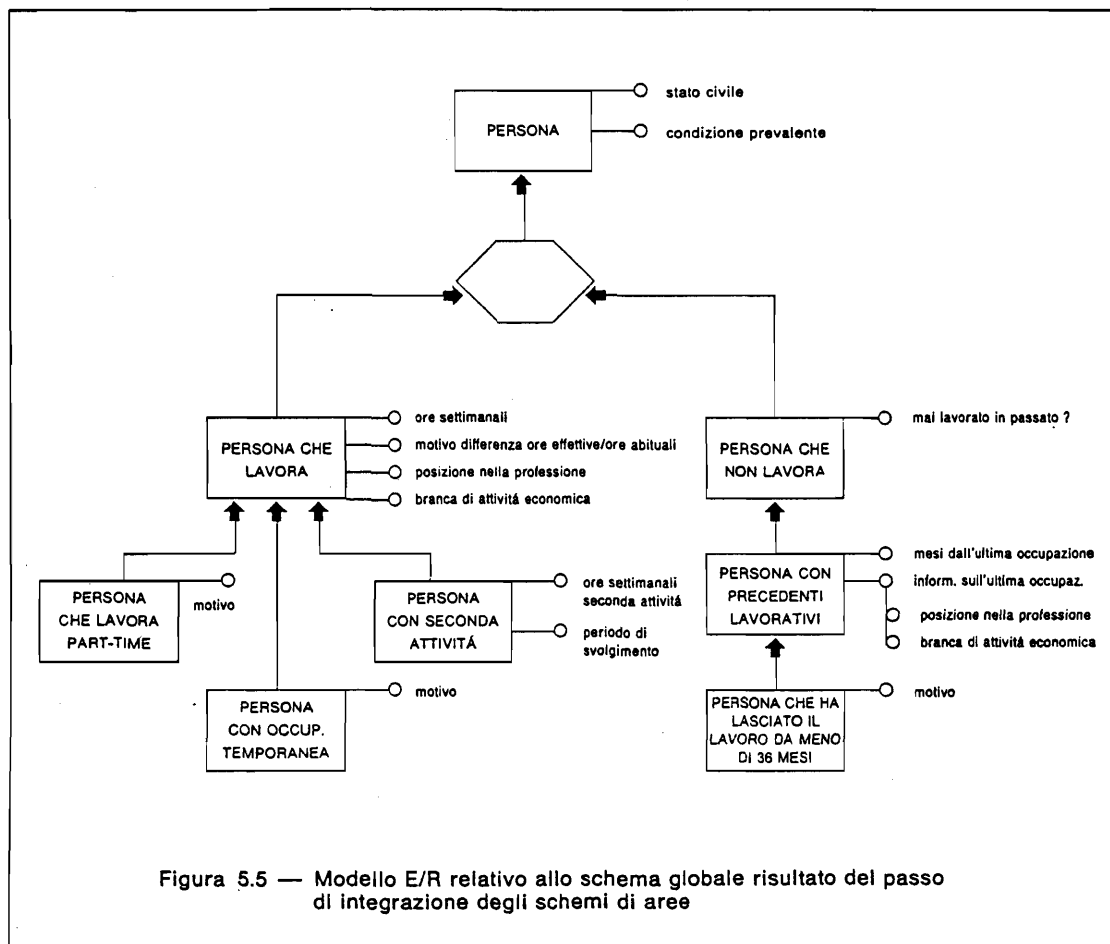


Figura 5.5 — Modello E/R relativo allo schema globale risultato del passo di integrazione degli schemi di aree

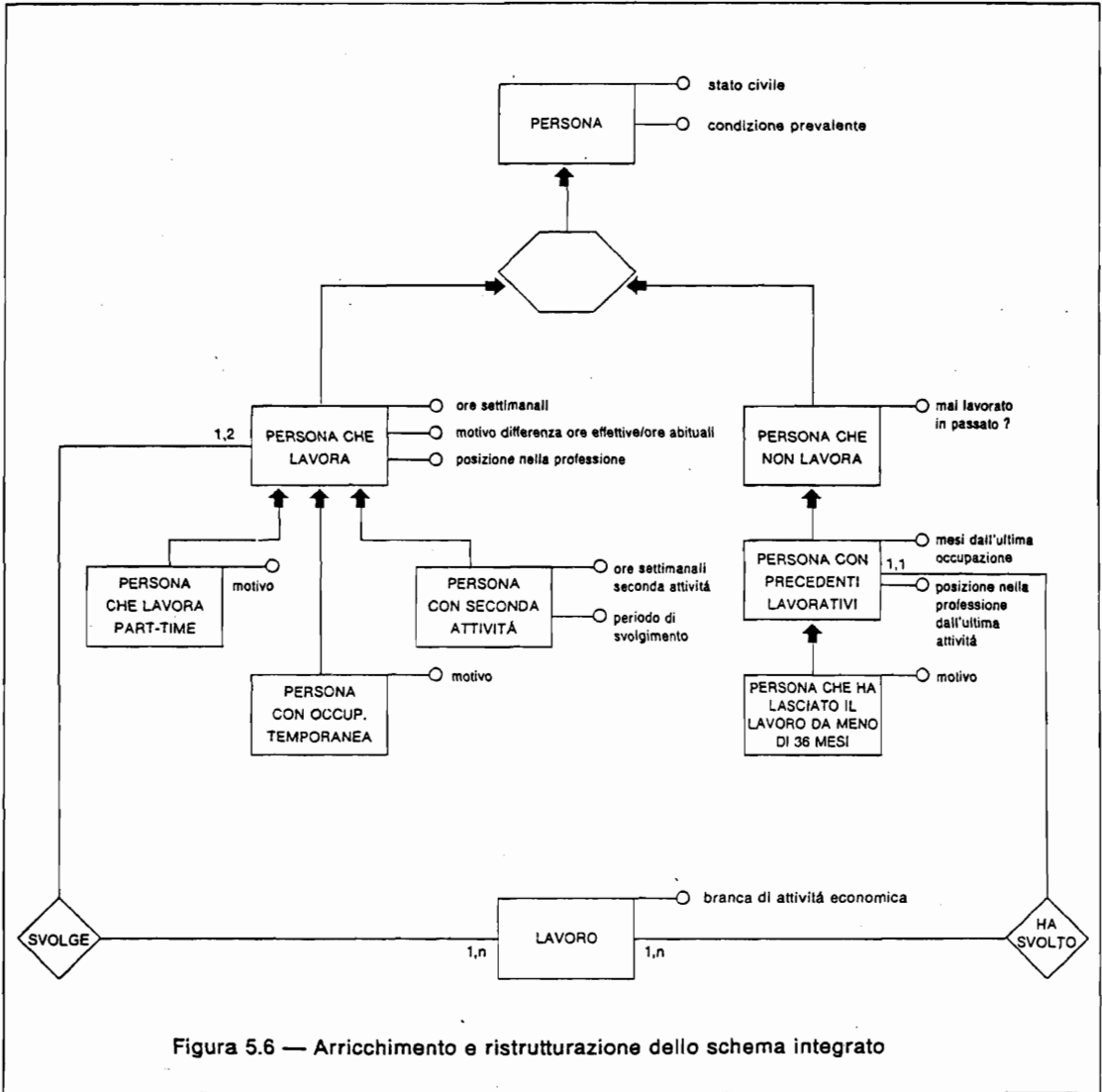


Figura 5.6 — Arricchimento e ristrutturazione dello schema integrato

6. Il Modello Concettuale Statistico (MCS)

Il **Modello Concettuale Statistico (MCS)** e' un modello concettuale per la rappresentazione delle proprieta' di dati aggregati. Esistono, nella letteratura e nella pratica corrente, altri modelli di rappresentazione di dati aggregati: citiamo il **SUBJECT**, il **GRASS**, il **SAM***. Tali modelli hanno pero' limiti intrinseci:

- nessun modello pone l'accento sulla necessita' di rappresentare le caratteristiche delle elaborazioni effettuate per ottenere i dati aggregati. Nel **SUBJECT**, nel **GRASS** e nel **SAM***, il tipo di elaborazione e' individuato unicamente dalla denominazione dei dati aggregati, e, a parte la limitatezza espressiva, e' impossibile ricostruire l'eventuale sequenza di elaborazioni diverse applicate per giungere ad un particolare dato statistico;

- il **SUBJECT** ed il **GRASS** si propongono come modelli **logici** dei dati aggregati, piu' che come modelli concettuali, tant'e' che nell'intenzione degli autori dovrebbero costituire la base per la realizzazione di veri e propri Sistemi di Gestione: la loro collocazione all'interno della fase di progettazione concettuale porrebbe dei problemi, dato che in questa fase il modello dei dati dovrebbe essere indipendente dalle modalita' della realizzazione concreta della corrispondente base dei dati;

- il **SAM*** e' invece un modello puramente concettuale, e di indubbia capacita' espressiva. Ma avendo l'obiettivo di rappresentare congiuntamente dati elementari e dati aggregati, fa uso di strutture di rappresentazione relativamente complesse, le cui modalita' di utilizzo per la costruzione degli schemi non sono di immediata comprensione. La nostra scelta, invece, e' stata quella di separare la rap-

presentazione dei dati elementari (per i quali vengono utilizzate le strutture e le regole del modello Entita'/Relazione) da quella dei dati aggregati, guadagnando in facilita' ed immediatezza di costruzione degli schemi, e risolvendo i problemi di collegamento tra i due ambienti mediante la metodologia integrata MIDAS.

La considerazione di tali limiti ha portato alla scelta di definire un nuovo modello, le cui caratteristiche rispondono alle esigenze progettuali e di documentazione di un ambiente quale quello proprio di un Istituto di statistica.

6.1. Strutture di rappresentazione del Modello Concettuale Statistico

Il modello MCS, al pari degli altri modelli, fa uso di **strutture di rappresentazione**, utilizzate per descrivere i dati aggregati, le loro caratteristiche e le relazioni tra essi intercorrenti. Uno schema MCS dei dati e' graficamente rappresentabile mediante un diagramma nel quale ogni struttura di rappresentazione e' identificata dai simboli riportati nella tabella 6.1.³ A tali simboli e' anche possibile associare delle descrizioni (dette **etichette**) che seguono una rigorosa sintassi e che sono utilizzabili per esprimere in modo dettagliato le caratteristiche dei concetti cui i simboli sono riferiti.

Nel seguito daremo una descrizione di ciascuna delle strutture ammesse dal modello, riportando per ognuna di esse la definizione, la sintassi dell'etichetta, ed un esempio di utilizzo.

³ Quasi tutte le strutture di rappresentazione sono graficamente espresse dai nodi del diagramma; solo il raggruppamento di attributi di categoria e' individuato da un arco contrassegnato da un asterisco

STRUTTURE DI RAPPRESENTAZIONE	SIMBOLI
Classe di oggetti	○ S
Attributo di categoria	○ C
Classificazione statistica	○ X
Elaborazione statistica / classe di dati	○ E
Generalizzazione	○ G
Attributo aggregato	○ A
Raggruppamento di attributi	*

Tabella 6.1 — Simboli grafici del Modello Concettuale Statistico

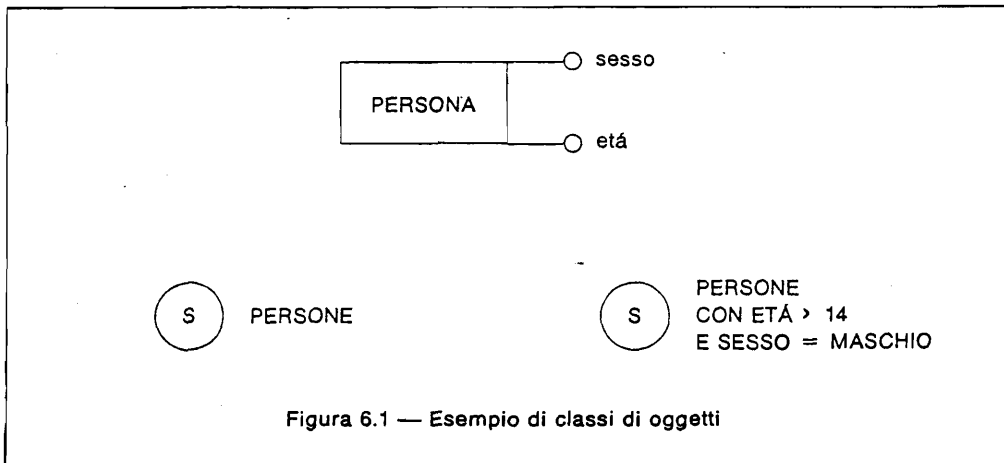


Figura 6.1 — Esempio di classi di oggetti

6.1.1. Classe di oggetti

Una classe di oggetti e' definita come un insieme di oggetti del mondo reale che hanno proprieta' comuni e che sono coinvolte in aggregazioni. Le classi che compaiono nello schema dei dati aggregati possono essere identiche a quelle presenti nello schema E/R, oppure possono essere derivate mediante una manipolazione, le cui caratteristiche compaiono nell'etichetta del corrispondente nodo dello schema MCS. Una classe di oggetti e' individuata, in uno schema MCS, da un **nodo S**.

L'**etichetta del nodo S** e' cosi' definibile:

< classe >
| WITH < parte qualificativa > |

dove < classe > indica il nome della classe di oggetti nello schema E/R di interesse per l'applicazione statistica, e < parte qualificativa > e' utilizzata opzionalmente per imporre eventuali restrizioni su tale classe (l'opzionalita' della parte qualificativa e' indicata dall'uso delle parentesi quadre, che nel seguito saranno utilizzate sempre con questa funzione).

Ad esempio, con riferimento alla fig. 6.1, nello schema E/R puo' comparire l'entita' PERSONA con gli attributi ETA' e SESSO. Al corrispondente nodo nello schema MCS puo' essere associata l'etichetta PERSONE, senza ulteriori specificazioni, oppure l'etichetta PERSONE CON ETA' > 14 E SESSO MASCHILE.

6.1.2. Attributo di categoria

Nel modello le variabili utilizzate per la classificazione di oggetti o di dati aggregati sono dette **attributi di categoria**. Tali attributi sono graficamente rappresentati dai **nodi C**. Ogni attributo di categoria presente in uno schema MCS deve essere presente nel corrispondente schema E/R, o come attributo dell'entita' relativa agli oggetti che debbono essere classificati, o come attributo di un'altra entita' collegata alla prima mediante una relazione, o, infine, come attributo di una relazione che collega l'entita' di interesse con una o piu' altre entita'.

L'**etichetta del nodo C** e' del tipo

< attributo >
| < parte descrittiva > |

dove < attributo > indica il nome dell'attributo di categoria e < parte descrittiva > (utilizzabile opzionalmente) serve a descrivere proprieta' particolari dell'attributo necessarie ad eliminare possibili ambiguita'. Nella fig. 6.2a e' riportato un esempio in cui compaiono gli attributi di categoria SESSO ed ETA' DELLA MADRE; lo schema E/R corrispondente e' quello riportato in alto nella figura. L'utilizzo della parte descrittiva nel caso di ETA' DELLA MADRE si rende necessario per distinguere i due attributi, nel caso si vogliono classificare gli oggetti dell'entita' NATO.

Nella fig. 6.2b e' riportato un ulteriore esempio che mostra la necessita' di qualificare con precisione l'attributo di categoria: se vogliamo classificare la classe di oggetti PERSONE secondo il NOME della REGIONE, dobbiamo anche indicare se si tratta della regione di residenza o di nascita, e dunque il tipo di relazione del corrispondente schema E/R che prendiamo in considerazione.

6.1.3. Classificazione statistica

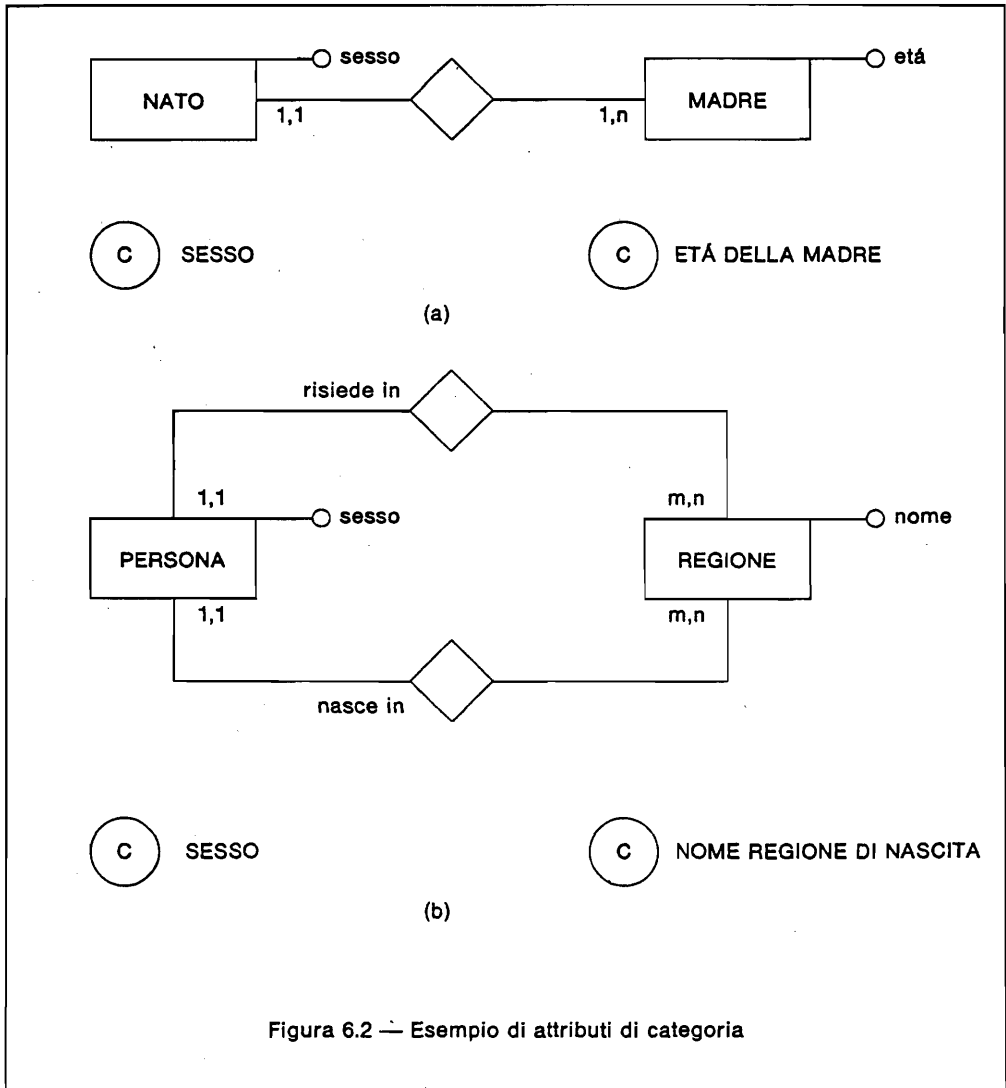
Una classificazione statistica esprime la relazione tra una classe di oggetti e gli attributi utilizzati per classificare tali oggetti. In sostanza, una classificazione statistica determina una particolare partizione dell'insieme di oggetti (o dell'insieme di dati) cui e' applicata; le caratteristiche di tale partizione (o, con un termine piu' statistico, distribuzione) sono fissate una volta noti i seguenti elementi:

- la classe di oggetti che si vuole classificare;
- gli attributi di categoria considerati nella classificazione;
- i domini relativi agli attributi di categoria.

Ogni oggetto e' inseribile in un dato sottoinsieme della partizione in base ai valori degli attributi di categoria riscontrabili in quell'oggetto.

In uno schema MCS, la classificazione statistica e' individuata graficamente dai **nodi X**.

Nella fig. 6.3 e' riportato un esempio di classificazione statistica. La classe di oggetti NATI e' classificata secondo gli attributi di categoria SESSO, PESO e ETA' DELLA MADRE. La cardinalita' della classificazione (il numero, cioe', di partizioni diverse individuabili) sara' data dalla cardinalita' dell'insieme prodotto



cartesiano dei domini degli attributi di categoria.⁴ Nell'esempio,

 D O M I N I I

SESSO : (MASCHIO, FEMMINA)
 PESO : (0-2000, 2001-3000,
 3001-oltre)
 ETA' : (14-25, 26-35,
 MADRE 36-oltre)

risulterà un numero di sottoinsiemi pari a 18 ($2 \times 3 \times 3$). Ogni sottoinsieme sarà individuato da una terna di valori, ognuno relativo al corrispondente attributo di categoria. Ad esempio, un nato maschio, di peso pari a 3100 grammi, con una madre di 28 anni sarà classificato nel sottoinsieme <MASCHIO, 3001-oltre, 25-36>

Gli attributi di categoria coinvolti in una classificazione possono partecipare anche ad altre classificazioni: in altre parole, un nodo C può essere collegato a più di un nodo X, come mostrato nella fig. 6.4.

La somma del numero di oggetti classificati in ogni partizione è normalmente pari alla numerosità della classe di oggetti: questo però può non essere vero nel caso che tra gli attributi di categoria ve ne sia uno o più di tipo multiplo, quando cioè per ogni oggetto che si voglia classificare vi siano più occorrenze di uno stesso attributo. Se, ad esempio (vedi fig. 6.5), vogliamo classificare le madri oltre che per attributi propri (CLASSE DI ETA'), anche per attributi dei figli posseduti, poiché questi possono essere più di uno per ogni madre, occorre tener conto che la somma complessiva delle madri classificate, ad esempio, per SESSO DEI FIGLI, sarà in generale maggiore del numero effettivo delle madri, in quanto ognuna di esse rientra tante volte nella distribuzione per quanti figli possiede. In questo caso, l'attributo multiplo è un nodo C contrassegnato da un asterisco.

In generale, è possibile riscontrare situazioni come quella riportata nella fig. 6.6, nella quale a livello di dati elementari esiste una relazione di tipo molti-a-molti tra PERSONA e AUTOMOBILE. Le possibili classificazioni coinvolgenti attributi dell'una e dell'altra entità sono ripor-

tate nella parte inferiore della figura. La soluzione (c) è concettualmente la migliore, in quanto non implica l'uso di attributi multipli.

6.1.4. Elaborazione statistica / Classe di dati

Una **elaborazione statistica** è una particolare applicazione di funzioni statistico-matematiche il cui risultato è una **classe di dati** aggregati, che nello schema è rappresentabile mediante i **nodi E**.

Tali classi di dati possono essere distinte in:

- **primitive** se sono il risultato di elaborazioni statistiche applicate a classi di oggetti: classici esempi sono le frequenze, mediante le quali si effettua il conteggio degli oggetti;
- **derivate** se sono il risultato di elaborazioni effettuate su classi di dati: ad esempio, la media o la varianza presuppongono che sia stato preventivamente eseguito il conteggio degli oggetti, e quindi vengono applicate a classi che non sono più di oggetti ma di dati.

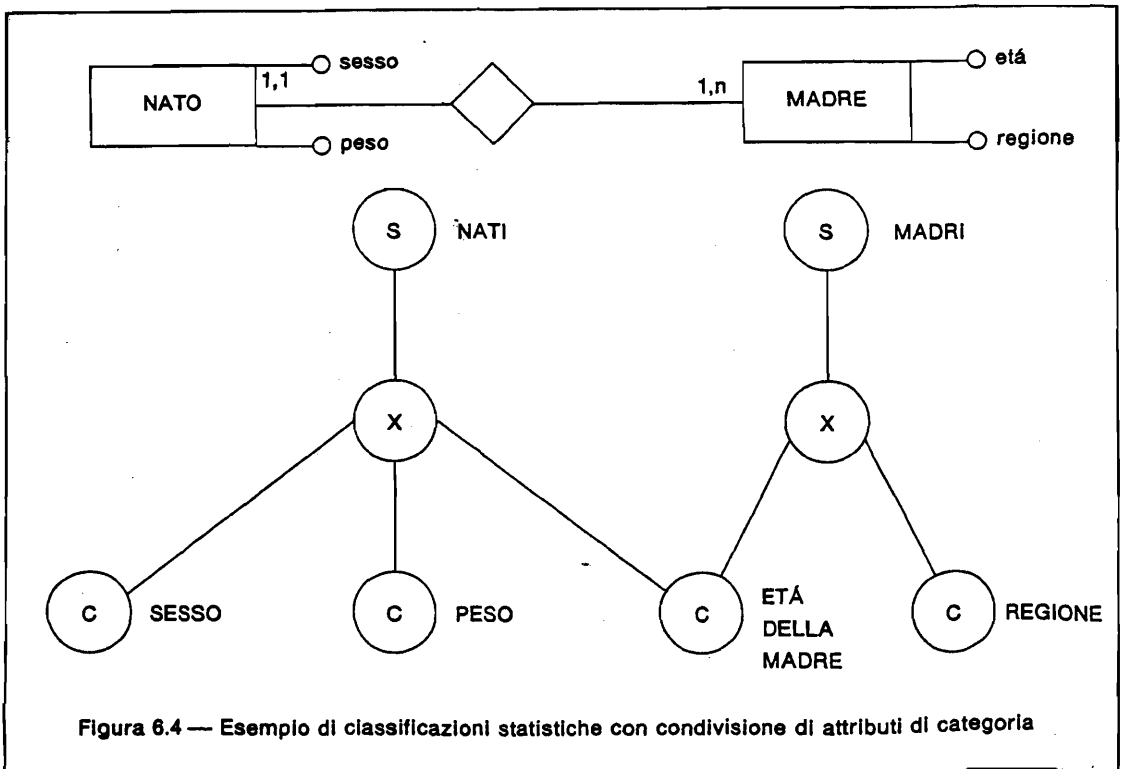
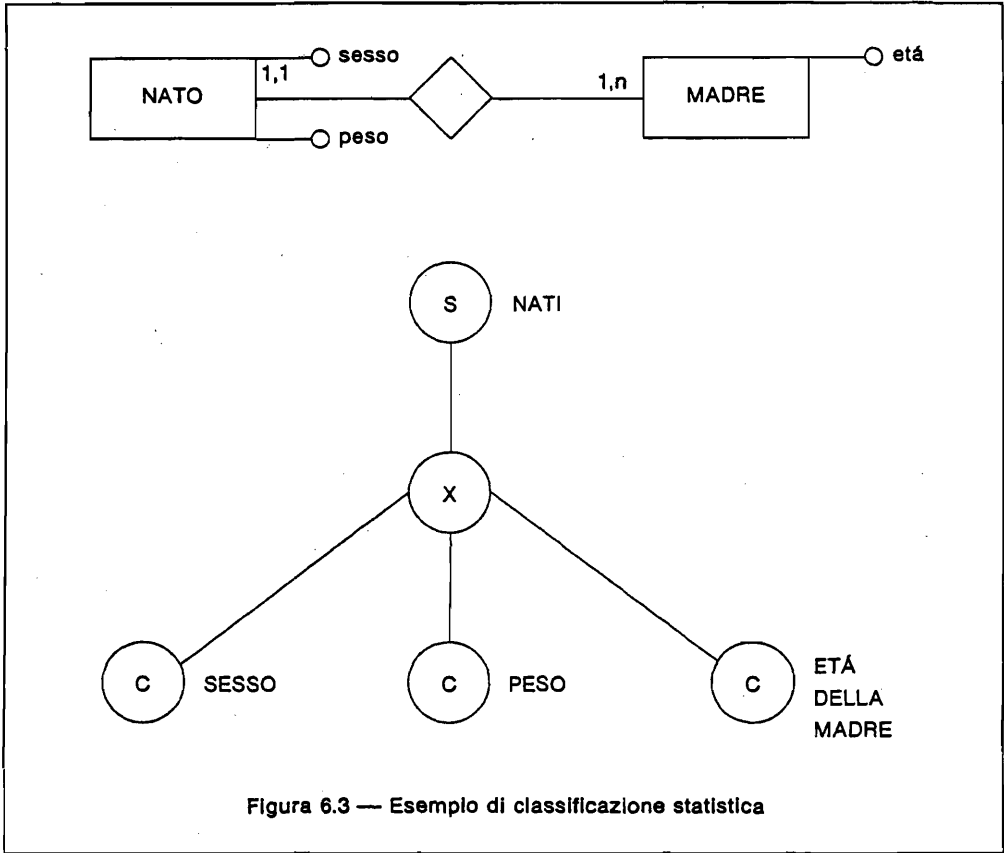
In generale, si avranno sequenze di nodi E applicate ad una classe di oggetti (nodo S). I nodi E di livello più alto saranno quelli corrispondenti alle elaborazioni più complesse, in quanto operanti su classi di dati che sono a loro volta risultato di elaborazioni; il nodo E di livello più basso è quello che opera direttamente sulla classe di oggetti, producendo i dati aggregati più semplici.

Classi di dati ottenute eseguendo elaborazioni su classi di oggetti (ad esempio conteggi) sono dette **classi di dati primitive**. Invece, classi di dati ottenute elaborando altre classi di dati sono dette **derivate**.

Una elaborazione statistica si intende completamente definita una volta che siano stati precisati i seguenti elementi:

1. la classe di oggetti o la classe di dati su cui viene applicata;
2. il tipo di elaborazione eseguita;

⁴ Intendiamo come prodotto cartesiano di due insiemi, l'insieme di tutte le coppie combinazioni di elementi del primo insieme con gli elementi del secondo insieme. Generalizzando, il prodotto cartesiano di n insiemi è l'insieme delle n-uple combinazioni degli elementi degli insiemi.



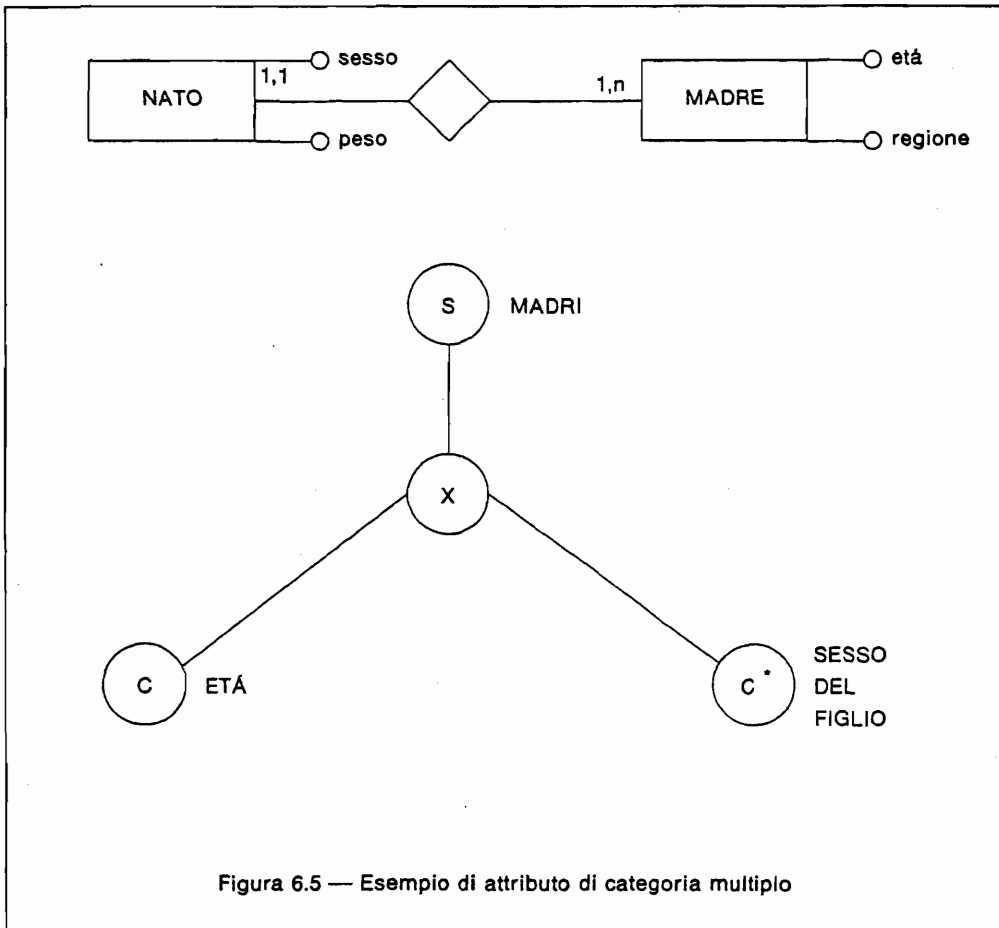


Figura 6.5 — Esempio di attributo di categoria multiplo

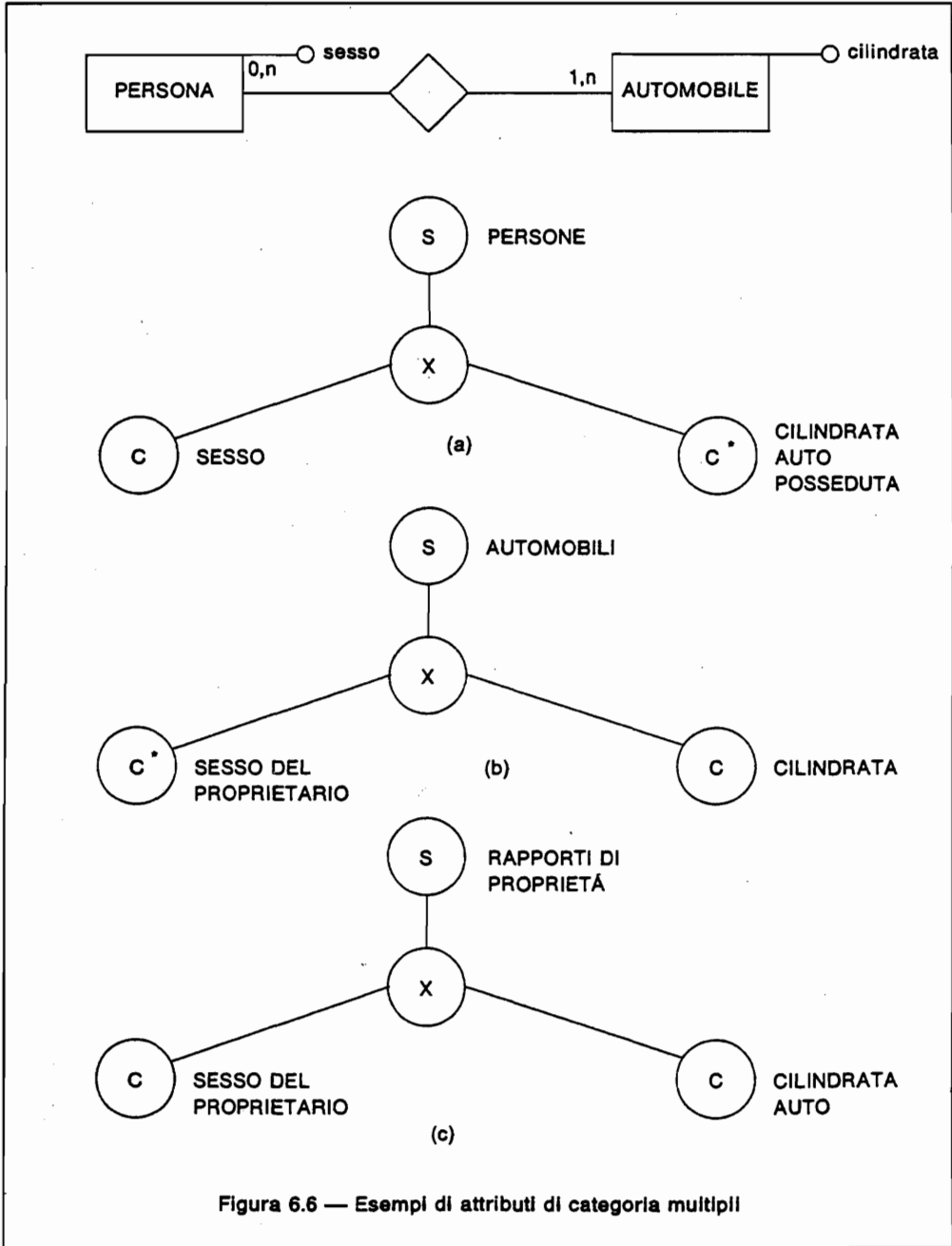


Figura 6.6 — Esempi di attributi di categoria multipli

3. la classificazione statistica applicata alla classe di dati prodotta dall'elaborazione in questione.

Il primo elemento e' definito dal nodo immediatamente discendente da quello corrente (nodi S nel caso di classe di oggetti, nodi E nel caso di classe di dati).

Il numero di nodi discendenti da un nodo E dipende dalle natura dell'operatore applicato, il quale puo' essere **monadico** o **poliadico**.

Nelle tabelle 6.2 e 6.3 riportiamo la tipologia degli operatori, distintamente per i monadici ed i poliadici, con la dizione inglese ed italiana.

OPERATORI MONADICI

COUNT	(NUMERO DI)
conteggio degli oggetti (produce frequenze)	
VALUE OF	(VALORE DI)
lista di valori degli attributi degli oggetti	
SUM	(SOMMA)
somma di valori di attributi numerici	
MIN	(MIN)
valore minimo assunto da un attributo	
MAX	(MAX)
valore massimo assunto da un attributo	
AVG	(MEDIA)
valore medio di un attributo	
%	(%)
frequenze percentuali	
VAR	(VAR)
varianza degli oggetti relativa ad un dato attributo	
STDEV	(SQM)
scostamento quadratico medio relativo ad un dato attributo	
INDEX	(INDICE)
rapporto % tra valori numerici di dati o di attributi di oggetti e valori di riferimento (base)	
CHANGE RATE	(TASSO VARIAZIONE)
rapporto % tra valori numerici di dati o di attributi di oggetti considerati in un periodo ed in uno precedente	

Tabella 6.2 - Operatori monadici

OPERATORI POLIADICI

PRODUCT	(PRODOTTO)
prodotto dei valori di due o piu' classi di dati	
QUOTIENT	(RAPPORTO)
divisione dei valori di una classe di dati coi valori	
ADDITION	(ADDIZIONE)
somma dei valori di due o piu' classi di dati	
SUBTRACTION	(SOTTRAZIONE)
differenza tra i valori di una classe di dati coi valori	

Tabella 6.3 - Operatori poliadici

Il tipo di operatore e' indicato nell'etichetta del nodo E, che possiede la seguente struttura:

| < nome classe dati > = |
 < tipo e parametr op. >
 | < specifiche di classificazione > |

Il nome della classe di dati e' opzionale: e' utile quando serve ad indicare dei dati la cui natura non e' immediatamente desumibile dal tipo dell'operatore, ma da un nome ad hoc. Si consideri l'esempio riportato nella fig. 6.7. Il nodo E di livello piu' alto rappresenta la classe di dati risultato dell'applicazione dell'operatore poliadico RAPPORTO alle classi di dati che esprimono le frequenze dei nati e delle donne in eta' feconda. E' conveniente, in questa situazione, dare il nome TASSO DI NATALITA' a tale nodo, per esprimerne direttamente il significato.

I **parametri degli operatori** variano da operatore ad operatore. Diamo di seguito i parametri che e' obbligatorio indicare per i diversi operatori:

1. gli operatori poliadici **RAPPORTO** e **DIFFERENZA** necessitano della definizione dell'ordine delle classi di dati cui vengono applicati. Sempre con riferimento all'esempio di fig. 6.7, nell'etichetta del nodo E, dopo l'indicazione del tipo di operatore (RAPPORTO), vengono riportate le classi di dati oggetto dell'elaborazione, cioe' le frequenze (NUMERO DI) di NATI e di DONNE CON ETA' > 14 E < 45: le prime costituiscono il numeratore, le seconde il denominatore del rapporto. Analogamente, nella DIFFERENZA occorre indicare il primo ed il secondo termine della sottrazione, mentre con gli operatori PRODOTTO e ADDIZIONE, le classi di dati sono indicate dai nodi discendenti, senza necessita' di indicarne l'ordine.
2. gli **operatori monadici**, con l'esclusione di NUMERO-DI, richiedono l'indicazione degli attributi degli oggetti sulla cui classe e' effettuata l'elaborazione secondo la forma:

< tipo operatore >
SU < nome-proprietà >

(nel caso di notazione inglese, al posto di SU occorre utilizzare ON). Ad esempio, se si vuole indicare l'eta' media delle madri, si scrivera':

MEDIA SU eta'
(AVG ON eta')

Non e' richiesta l'effettiva presenza di tali attributi come nodi C nello schema MCS: e' sufficiente che gli attributi indicati siano presenti nel modello E/R dei dati disaggregati.

3. l'operatore monadico **%** richiede l'indicazione dell'attributo (o degli attributi) rispetto al quale viene effettuata la totalizzazione a 100. A tal fine viene utilizzata la clausola RISPETTO A (RESPECT ON). Ad esempio:

% RISPETTO A sesso
(% RESPECT ON sesso)

e' riferibile alla tabella:

		SESSO		
ETA'		M	F	M+F
0-15		51	49	100
16->		33	67	100

4. l'operatore monadico **INDICE** dispone della clausola 'BASE = data' oppure 'BASE = riferimento geografico' per indicare il periodo o il territorio scelto come base per il calcolo degli indici. Ad esempio:

INDICE SU reddito BASE = 1980

INDICE SU reddito BASE = LAZIO

In entrambi i casi, la base indicata deve essere interna all'insieme di modalita' del dominio di uno degli attributi di categoria caratterizzanti la classificazione della classe di dati. Nel primo esempio, quindi, dovra' esistere un nodo C 'ANNO' (che possa assumere 1980 come modalita') nel secondo caso un nodo C 'REGIONE'.

Normalmente una classe di dati viene classificata secondo le stesse modalita' mediante le quali viene classificata la classe di oggetti indicata dal nodo S (quindi secondo gli attributi di categoria indicati nei nodi C discendenti dal corrispondente nodo X). E' pero' possibile

restringere l'insieme degli attributi utilizzati per classificare la singola classe di dati, utilizzando l'etichetta del nodo E, in due modi alternativi:

1. dichiarando esplicitamente tali attributi mediante la clausola **PER** (BY nella notazione inglese);
2. dichiarando gli attributi che **non** rientrano nella classificazione mediante la clausola **AGGREGANDO PER** (SUMMARIZING ON).

Verra' scelto l'uno o l'altro modo a seconda della convenienza: se, ad esempio, gli attributi sono pochi rispetto a quelli indicati nel nodo X, converra' utilizzare la clausola PER; se invece sono quasi tutti quelli del nodo X, converra' indicare quelli che scompaiono, utilizzando la clausola alternativa AGGREGANDO PER. Gli esempi riportati nella fig. 6.8 sono esattamente equivalenti tra di loro.

6.1.5. Generalizzazione tra classi di dati

Nel modello E/R e' possibile rappresentare relazioni di sottoinsieme tra le classi di oggetti (entita'), mediante il meccanismo della astrazione di generalizzazione. Nel modello MCS e' invece possibile rappresentare relazioni di sottoinsieme tra classi di dati, istituendo una sorta di struttura gerarchica tra i dati che vengono rappresentati.

In uno schema MCS le generalizzazioni sono espresse mediante **nodi G**.

Nell'esempio di fig. 6.9 la gerarchia tra i nodi G indica che i dati relativi al movimento della popolazione sono una generalizzazione dei dati sulla natalita' e sulla mortalita'. I nodi G di livello piu' basso insistono su nodi E, cioe' sulle effettive classi di dati.

6.1.6. Aggregato di attributi di categoria

Puo' essere necessario, od utile, introdurre strutture di attributi di categoria: ad esempio, DATA e' definibile come aggregato di ANNO e MESE, cosi' come INDIRIZZO lo e' di VIA, NUMERO CIVICO e CITTA'.

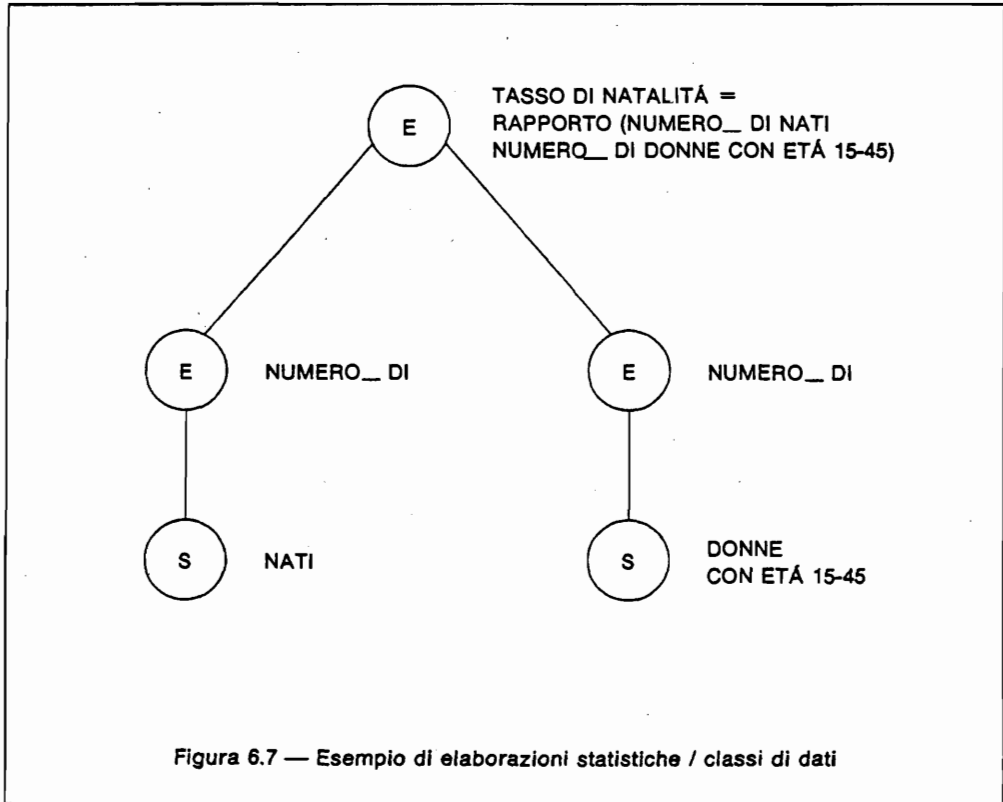
In uno schema MCS gli aggregati di attributi sono espressi da **nodi A**.

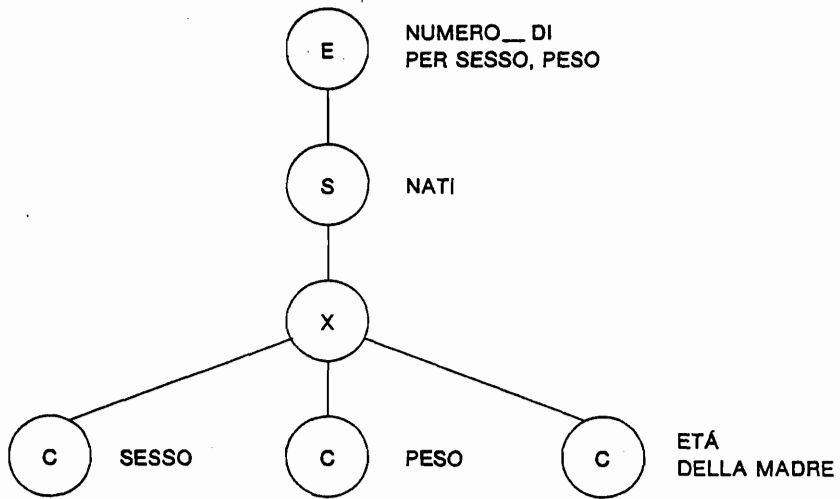
L'utilizzo di tali strutture e' di particolare importanza quando una classe di oggetti viene classificata secondo piu' attributi non direttamente riferibili a tali oggetti, ma ad altri posti in relazione con i primi. Nell'esempio di fig. 6.10, le persone sono classificate secondo la cilindrata e la potenza dell'automobile posseduta: associare tali attributi in una struttura e' sia utile (lo schema ne guadagna in chiarezza), sia necessario: si tratta di attributi multipli, i cui valori devono essere considerati congiuntamente per ogni automobile posseduta. E' l'attributo CARATTERISTICHE DELL'AUTOMOBILE che diviene un attributo multiplo, e che e' quindi connotato da un asterisco.

6.1.7. Raggruppamento di attributi di categoria

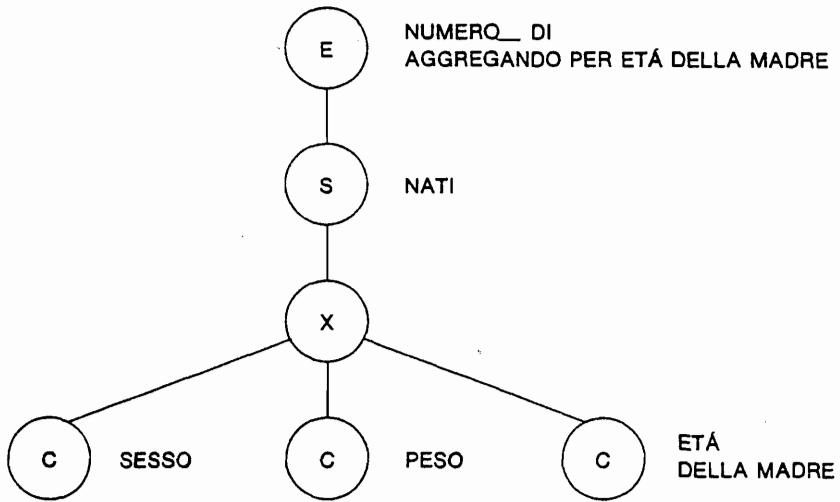
Un altro tipo di struttura definibile tra attributi di categoria e' ammissibile quando ricorrono le seguenti condizioni: i valori di un attributo determinano una partizione in sottoinsiemi del dominio di un altro attributo, e tali sottoinsiemi sono disgiunti. Ad esempio, l'attributo REGIONE si trova in tale rapporto con l'attributo PROVINCIA, dato che ogni regione determina un gruppo di province che non ha intersezioni con altri gruppi. Il raggruppamento (GROUPING) e' importante in quanto permette di sapere che una classificazione statistica disponibile per un attributo di categoria che si trova in una relazione (subordinata) di raggruppamento con un altro attributo, e' disponibile (previa una semplice elaborazione) anche rispetto a quest'ultimo. Ad esempio, una classificazione disponibile per PROVINCE, sara' disponibile anche per REGIONE, attraverso una semplice aggregazione dei dati secondo il rapporto PROVINCIA --> REGIONE. In uno schema MCS una relazione di grouping sara' individuata da **archi contrassegnati da un asterisco** colleganti due nodi C. Nella fig. 6.11 e' riportato un esempio di classificazione statistica coinvolgente attributi di categoria che si trovano in una relazione di grouping tra di essi. Come sappiamo, e' possibile esprimere, nelle etichette dei nodi E, variazioni della classificazione cui sono sottoposte le classi di oggetti (nodi S), mediante le clausole PER e AGGREGANDO PER. Abbiamo due casi:

1. se nel nodo E utilizziamo la clausola PER non vi sono mai problemi di ambiguita';





(a)



(b)

Figura 6.8 — Esempi di elaborazioni statistiche / classi di dati con restrizioni sulle classificazioni statistiche

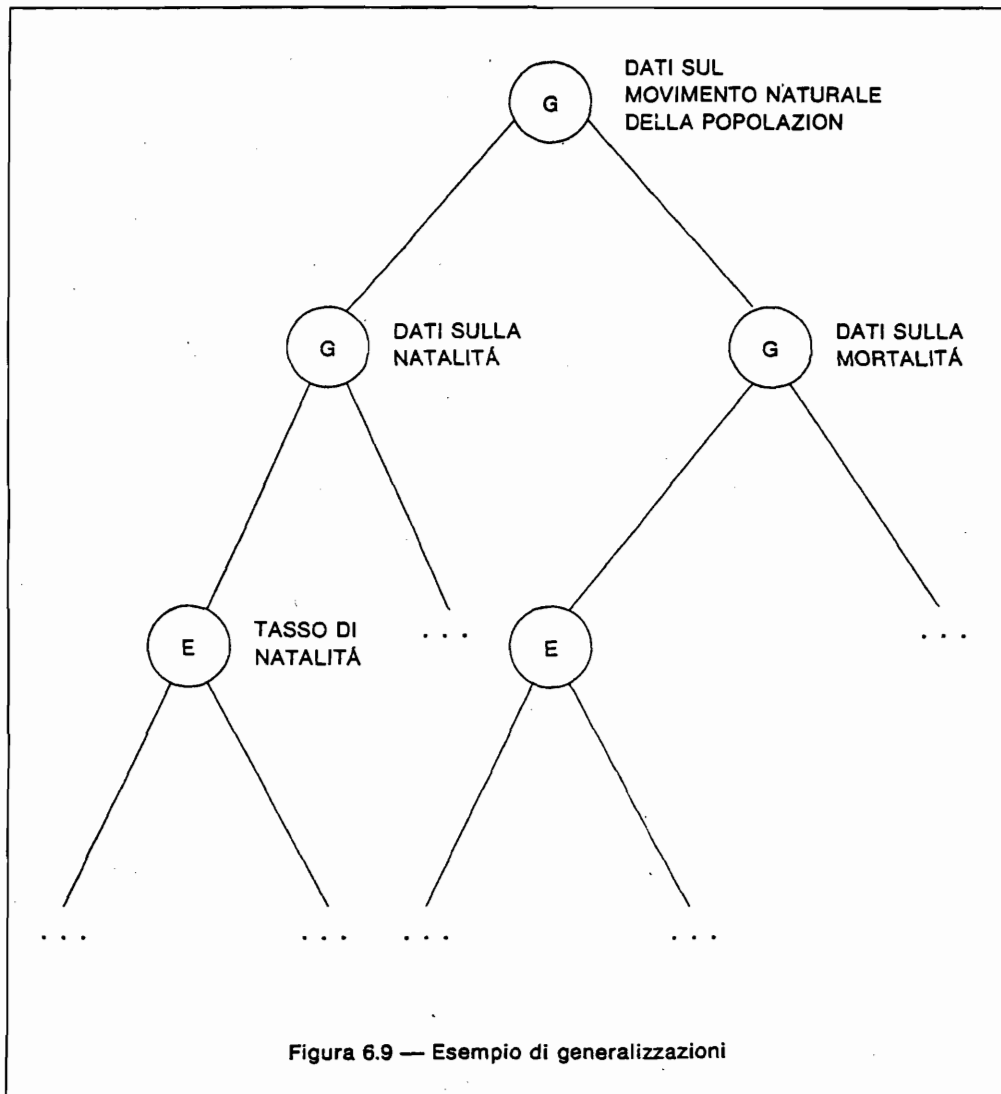
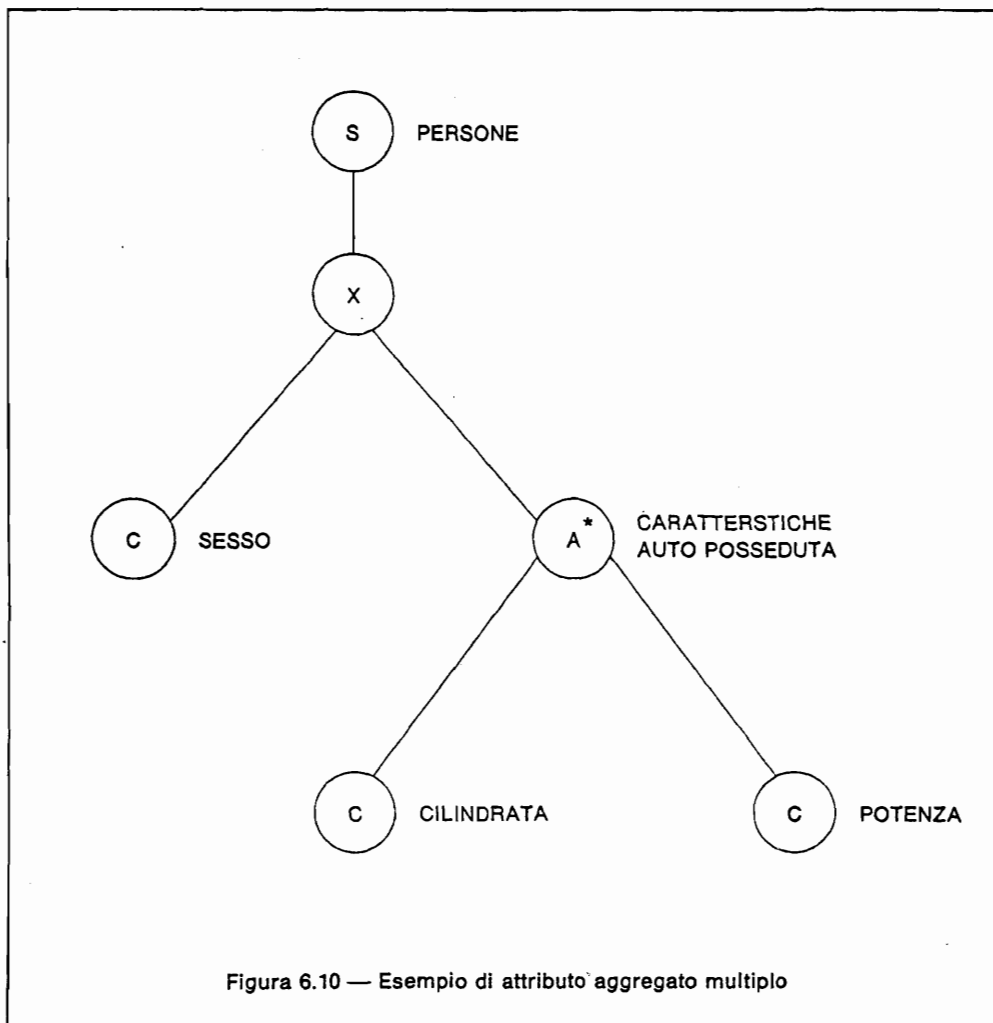
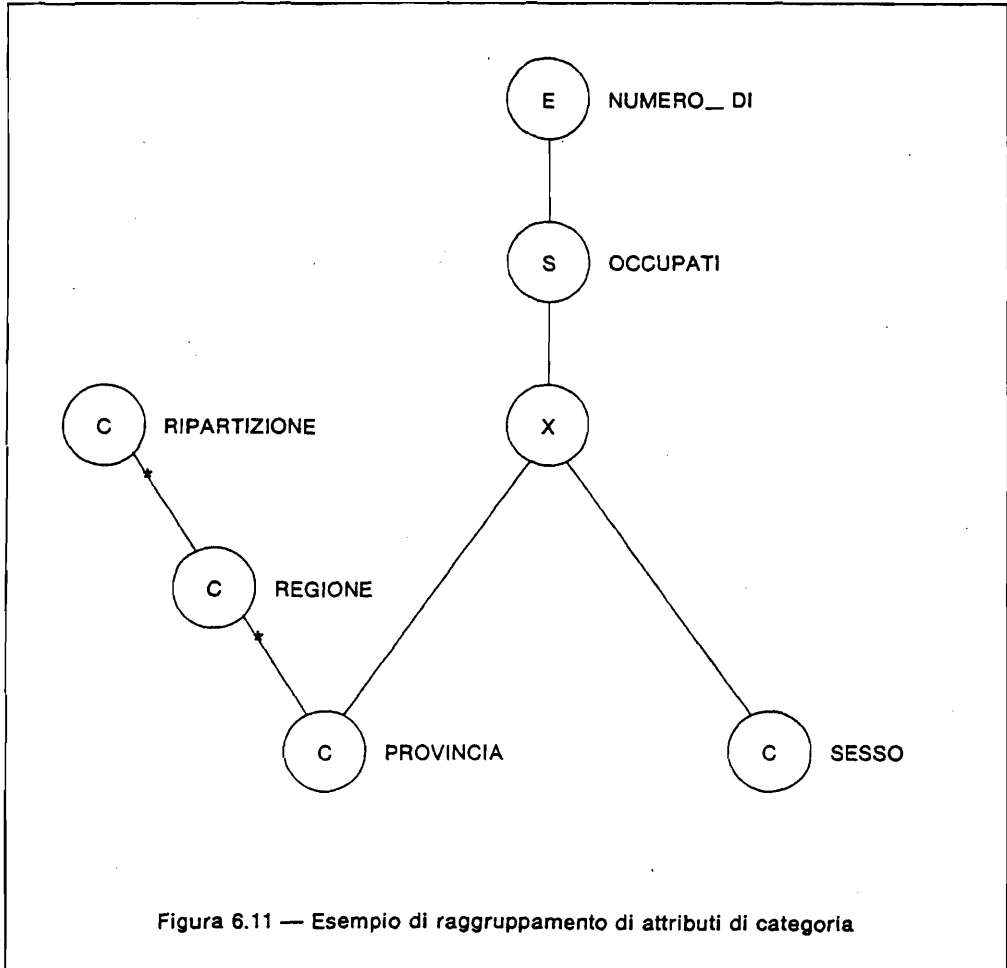


Figura 6.9 — Esempio di generalizzazioni





2. se utilizziamo la clausola AGGREGANDO PER assumiamo per convenzione che scompaiono dalla classificazione tutti gli attributi nella catena del grouping che si trovano sotto quello indicato, compreso quello indicato.

Nell'esempio, se nel nodo NUMERO DI (riferito al nodo S: OCCUPATI) indichiamo 'AGGREGANDO PER provincia', allora le frequenze degli occupati saranno disponibili per sesso, regione e ripartizione; se, invece, indichiamo 'AGGREGANDO PER regione', allora le frequenze saranno classificate per ripartizione e

sesso. Se, infine, indichiamo 'AGGREGANDO PER ripartizione', scompaiono tutti gli attributi nella relazione di grouping e la classificazione rimane solo relativamente al sesso.

6.2. Regole di costruzione di schemi MCS

Nella tabella 6.4 sono schematicamente riportate le regole di associazione tra i vari tipi di nodi.

NODO DI LIVELLO SUPERIORE	--->	NODI AMMISSIBILI DI LIVELLO INFERIORE
G		G, E, S
E		S
S		X, -
X		C, A
A		C, A
C		C, -
NODO DI LIVELLO INFERIORE	--->	NODI AMMISSIBILI DI LIVELLO SUPERIORE
G		G, -
E		G, E
S		G, E
X		S
A		X, A
C		A, C, X, -

Tabella 6.4 - Regole di associazione

6.3. Alcuni esempi di schemi MCS e loro integrazione

Diamo nel seguito un esempio di utilizzazione dell'MCS, con la costruzione di schemi parziali e di descrizioni linguistiche dei dati aggregati, e la successiva integrazione in un unico schema.

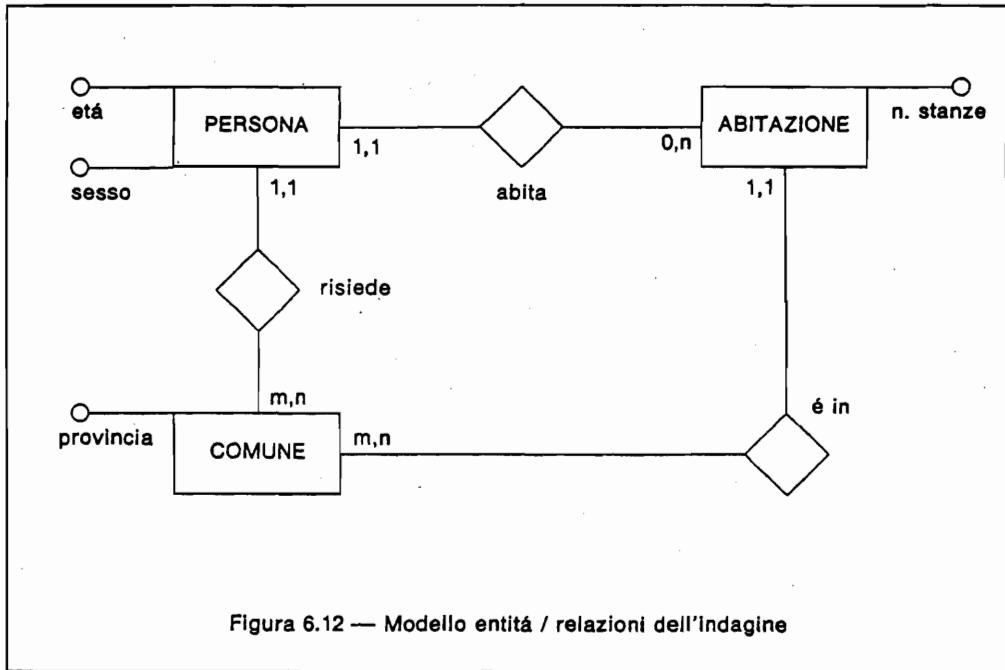
L'esempio e' basato su un'ipotetica indagine sulle persone e le abitazioni

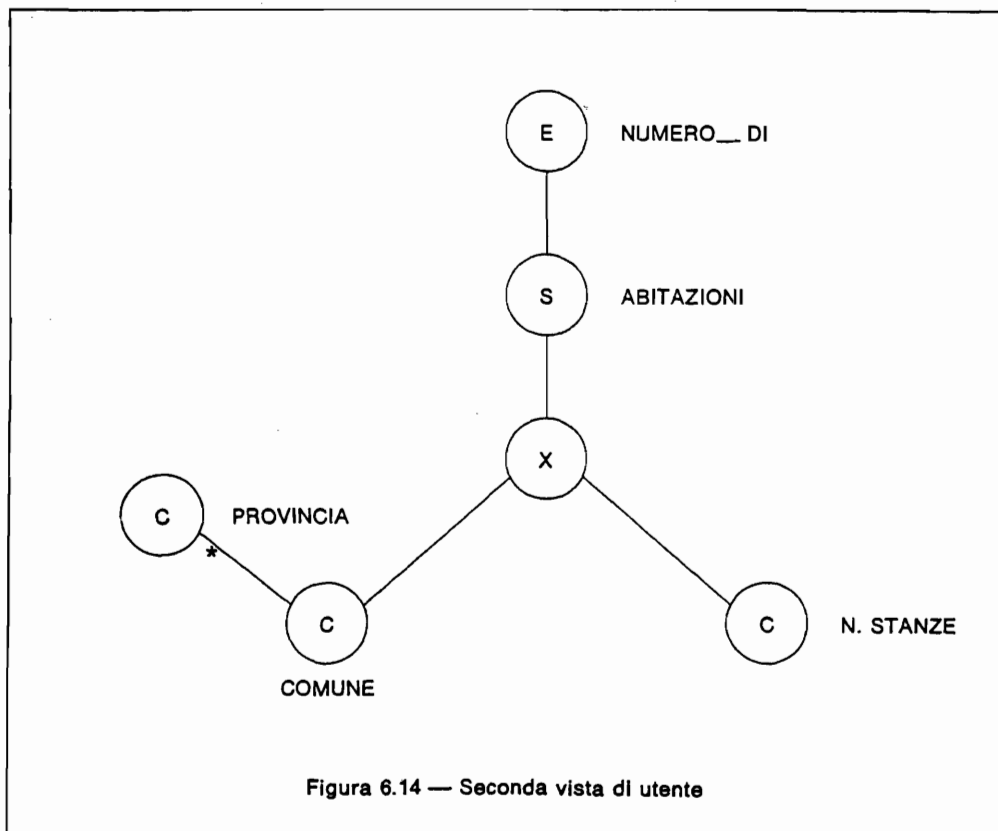
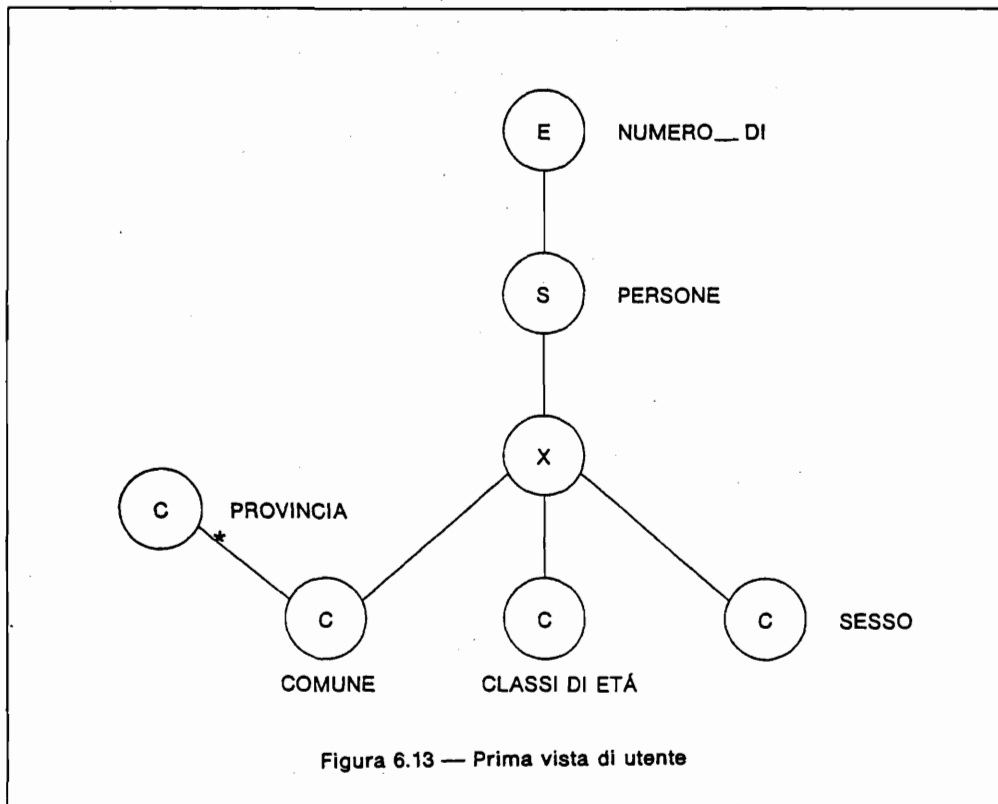
Il mondo reale coinvolto nell'indagine e' descritto, mediante le strutture di rappresentazione del modello E/R, nella fig. 6.12. Lo schema rappresenta i **dati elementari**, quelli cioe' direttamente rilevati presso le entita' del mondo reale. Supponiamo ora di voler descrivere i seguenti **dati aggregati** ottenuti attraverso elaborazioni dei dati disaggregati:

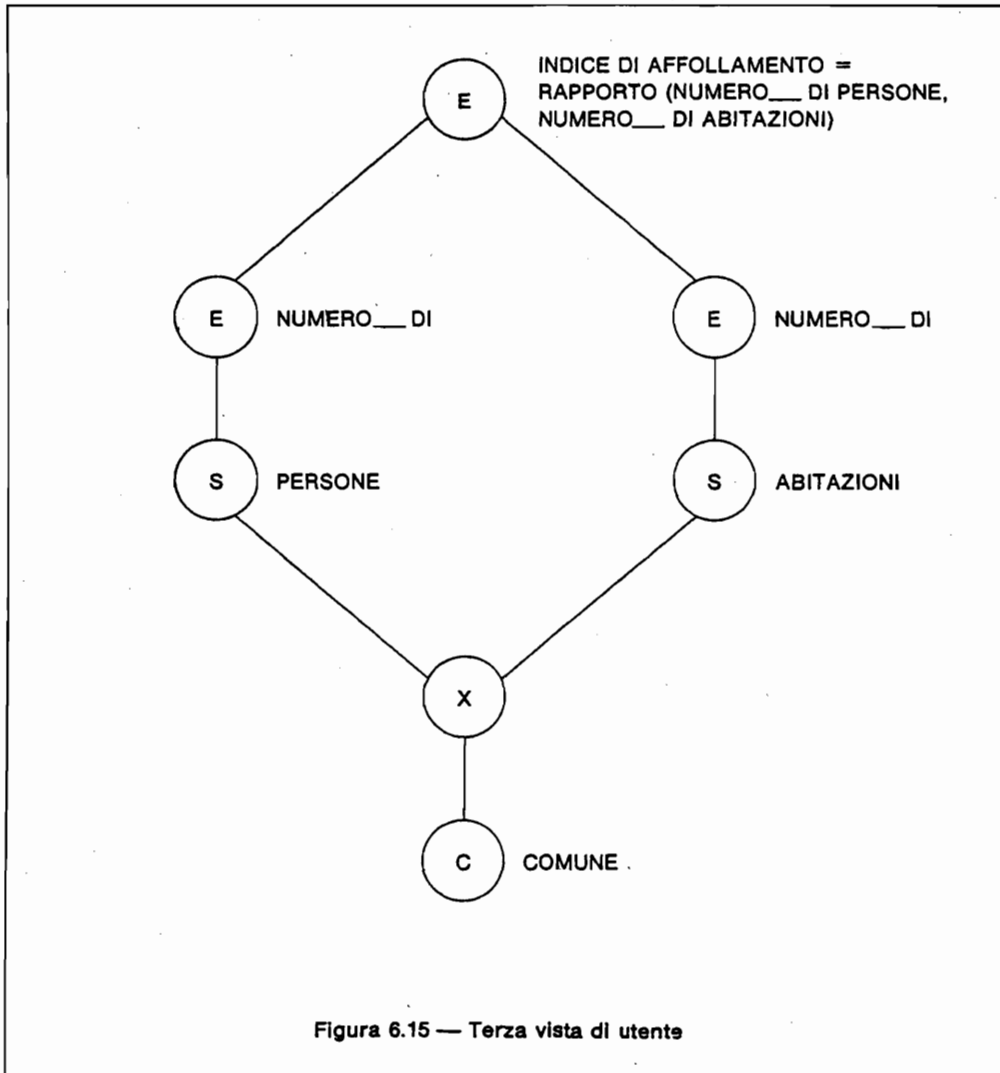
1. numero di persone (valori assoluti e percentuali per classi di eta') per sesso, classe d'eta' e comune;
2. numero di abitazioni per comune e numero di stanze;

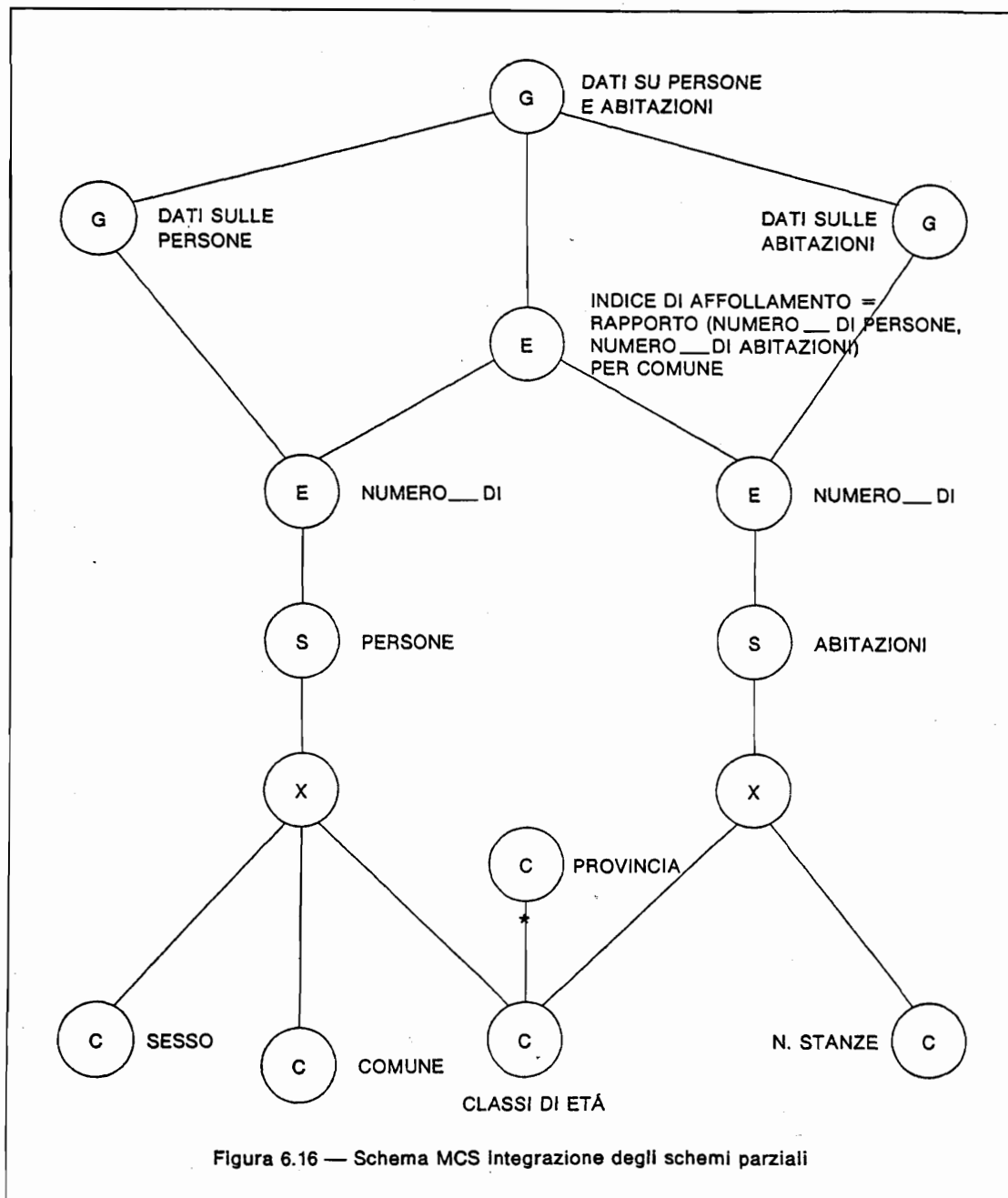
- 3 indice di affollamento, inteso come rapporto tra persone e abitazioni.

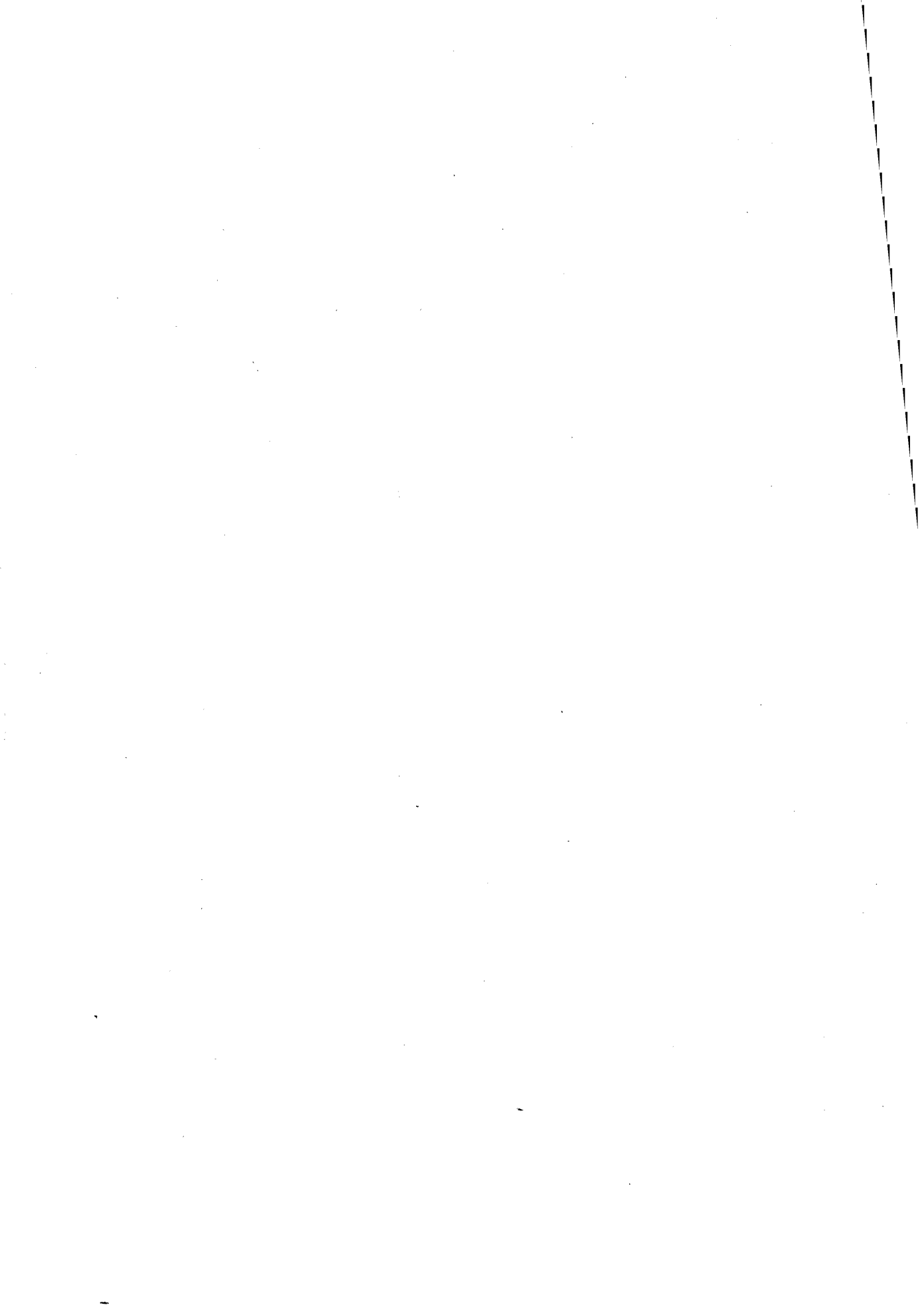
I relativi schemi sono riportati nelle figg. 6.13, 6.14 e 6.15. Lo schema globale, risultato dell'integrazione dei tre precedenti, e' invece riportato nella fig. 6.16.











7. La metodologia complessiva per la progettazione dei dati disaggregati e di quelli aggregati

La scelta di fondo della metodologia MIDAS e' di progettare lo schema dei dati aggregati solo dopo il progetto dello schema dei dati disaggregati; tale scelta attribuisce quindi importanza prioritaria all'analisi dei dati disaggregati e alla completa comprensione del loro significato prima di procedere alla progettazione delle aggregazioni possibili.

La metodologia complessiva per la progettazione dei dati aggregati, rappresentata nelle figg. 7.1a, 7.1b e 7.1c, puo' essere cosi' schematizzata:

1. **ANALISI DEI REQUISITI:** i requisiti relativi ai dati disaggregati (espressi in linguaggio naturale, oppure formalizzati in uno o piu' modelli di rilevazione) ed ai dati aggregati (anch'essi espressi in linguaggio naturale, oppure mediante le specifiche di tavole statistiche) vengono raccolti, raggruppati ed infine analizzati per eliminare eventuali ambiguita';
2. **DISEGNO CONCETTUALE DEI DATI DISAGGREGATI:** viene costruito lo schema concettuale globale dei dati disaggregati secondo le modalita' gia' riportate nel cap. 4;
3. **DISEGNO CONCETTUALE DEI DATI AGGREGATI:** per ogni aggregazione si effettuano due passi distinti:
 - a. viene anzitutto verificata la **compatibilita'** dell'aggregazione stessa con i dati disaggregati a disposizione; tale verifica e' fatta analizzando lo schema E/R globale dei dati disaggregati: se tale schema risulta non essere completo (nel senso che in esso non sono espressi alcuni concetti necessari per l'aggregazione), si puo' procedere ad un arricchimento

- / ristrutturazione dello schema E/R in modo da rendere compatibile l'aggregazione, oppure quest'ultima non viene ulteriormente considerata;
- b. si procede alla costruzione del sottoschema MCS relativo all'aggregazione;

4. **INTEGRAZIONE DEI SOTTOSCHEMI:** l'integrazione viene effettuata unificando coppie di sottoschemi di aggregazione fino ad ottenere lo schema globale che contenga tutti i sottoschemi;
5. **VERIFICHE FINALI:** vengono effettuate eventuali ristrutturazioni dello schema dei dati aggregati per migliorarne la leggibilita' e/o la minimalita'.

Analizziamo ora separatamente le varie fasi, servendoci anche in questo caso dell'esempio del Sistema Informativo del Lavoro introdotto nel cap.4.

7.1. Analisi dei requisiti

I requisiti sono suddivisi a seconda che riguardino i dati disaggregati o quelli aggregati. Nel seguito faremo riferimento ad una situazione quale quella descritta nel paragrafo 2.1 (primo scenario di applicazione), in cui sia il modello di rilevazione che le tavole statistiche sono un risultato dell'attivita' di progettazione; supporremo quindi che il frammento del modello di rilevazione sulle Forze Lavoro riportato nella fig. 5.2 ed utilizzato nel cap.5 sia solo una bozza di modello, suscettibile di modifiche.

I requisiti sulle aggregazioni sono espressi in linguaggio naturale; nel nostro esempio assumiamo che siano richieste le seguenti aggregazioni:

ANALISI DEI REQUISITI

**Distingui i tipi di requisiti
Analizza i diversi requisiti e rimuovi le ambiguità**

DISEGNO CONCETTUALE DEI DATI DISAGGREGATI

**Costruisci uno schema dei dati disaggregati a partire
dal requisiti che si riferiscono ai dati disaggregati**

DISEGNO CONCETTUALE DEI DATI AGGREGATI

Per ogni aggregazione definita nei requisiti:

1. controlla la completezza dello schema E/R dei dati disaggregati
2. se lo schema E/R non è completo rispetto alla aggregazione, procedi ad un arricchimento e/o ristrutturazione dello schema in modo da rendere ammissibile l'aggregazione
3. unifica i due sottoschemi in un unico sistema

INTEGRAZIONE

Scegli un sottoschema di aggregazione

1. scegli un nuovo sottoschema di aggregazione
2. unifica i due sottoschemi in un unico schema

**Considera lo schema prodotto al punto 2 e ripeti i passi
1—2 finché tutte le aggregazioni sono espresse nello
schema globale**

**Definisci la struttura delle generalizzazioni delle classi di dati
nello schema globale**

VERIFICHE FINALI

**Effettua eventuali ristrutturazioni dello schema dei
aggregati per migliorare la leggibilità e/o la minimalità**

Figura 7.1 a — Sequenza della metodologia complessiva per la progettazione dei dati aggregati e disaggregati

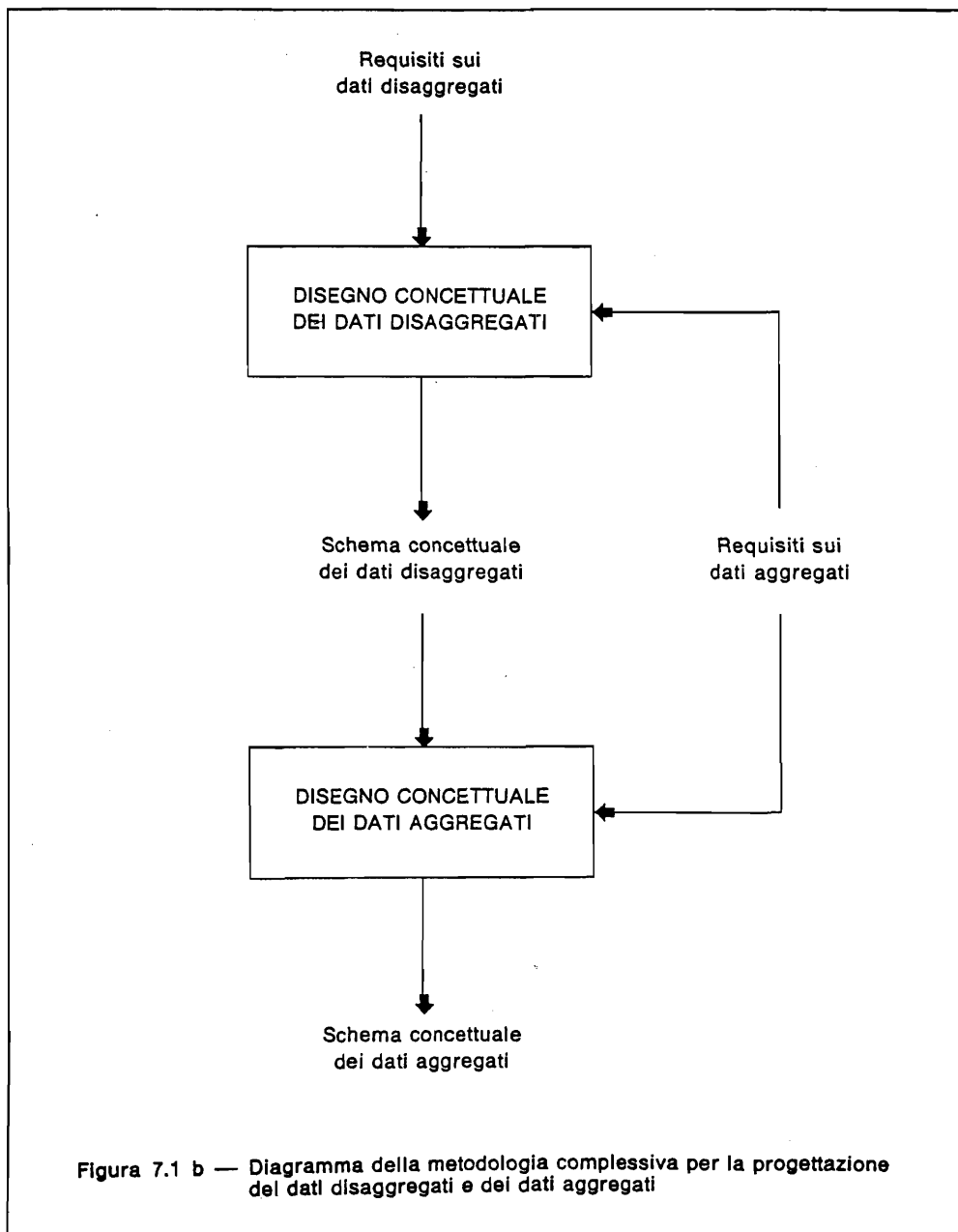
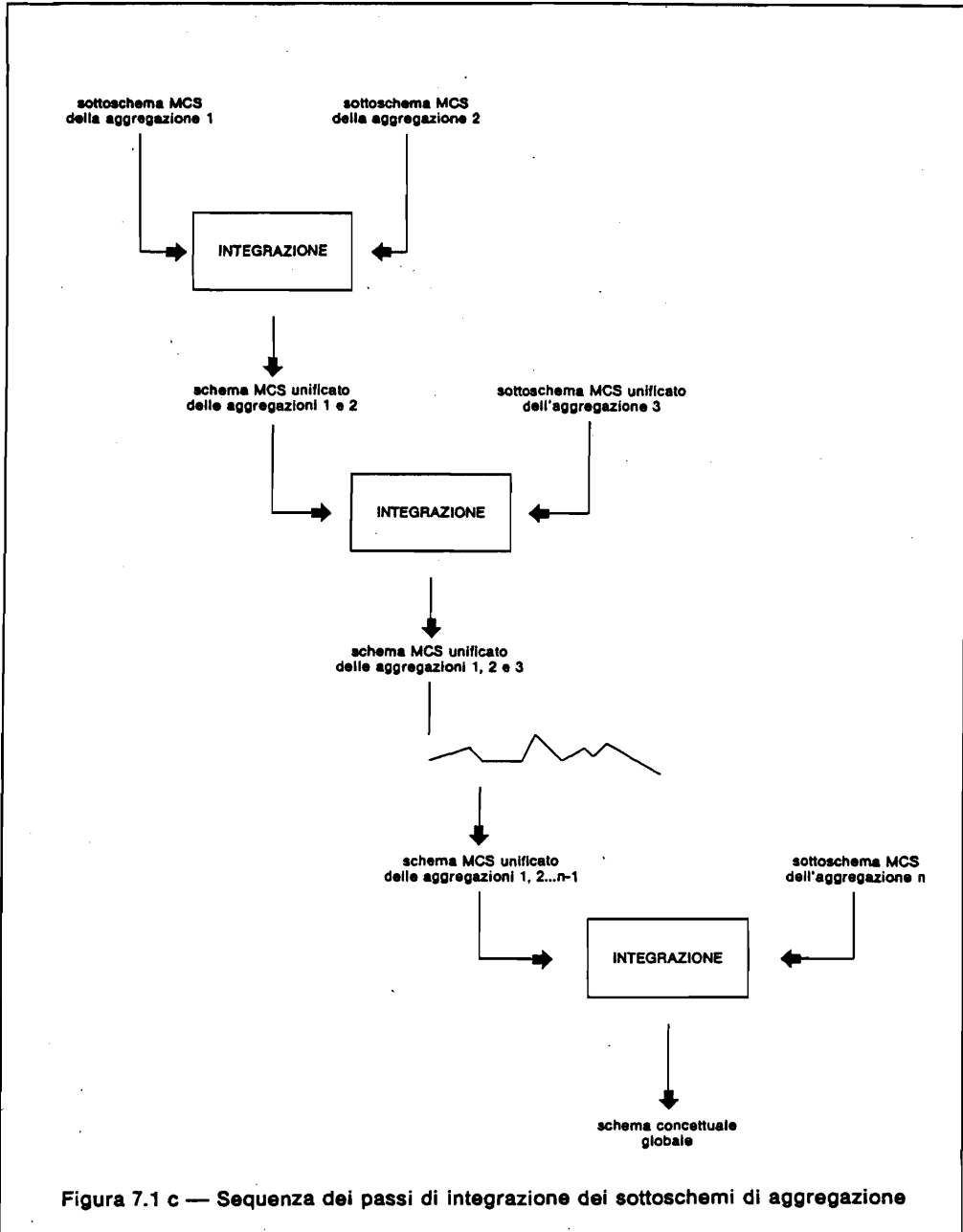


Figura 7.1 b — Diagramma della metodologia complessiva per la progettazione dei dati disaggregati e dei dati aggregati



1. occupati con eta' inferiore a 30 anni per sesso, branca di attivita' e regione geografica;
2. occupati nell'industria per titolo di studio, sesso e classe di eta' (valori assoluti e percentuali rispetto al titolo di studio);
3. media delle ore lavorate dagli occupati per branca di attivita' e ripartizione geografica;
4. persone in cerca di lavoro per titolo di studio ed eta'.

7.2. Disegno concettuale dei dati disaggregati.

La costruzione dello schema concettuale globale dei dati disaggregati segue le modalita' gia' descritte nel cap.5. Nel nostro esempio il risultato e' quindi lo schema E/R riportato nella fig. 5.6.

7.3. Disegno concettuale dei dati aggregati.

Come gia' detto, di ogni aggregazione richiesta va anzitutto verificata la compatibilita'. Il quadro di riferimento per una tale verifica e' costituito dallo schema E/R risultato della fase precedente (vedi fig. 5.6).

Prendiamo in considerazione la prima aggregazione (numero di occupati con eta' inferiore a 30 anni per sesso, branca di attivita' e regione). Si rileva che:

1. esiste una sinonimia tra i termini "occupati" dell'aggregazione e "persona che lavora" nello schema E/R (si tratta di uno stesso concetto che ha nomi differenti): si procede assegnando il nome OCCUPATO all'entita' PERSONA CHE LAVORA;
2. lo schema E/R non e' completo rispetto all'aggregazione, in quanto i concetti "sesso", "regione" ed "eta'" non vi sono contenuti, ne' sono indirettamente ricavabili da esso.

In considerazione del secondo punto, occorre prendere la decisione se tralasciare l'aggregazione, oppure procedere ad un arricchimento dello schema E/R. Nel nostro caso scegliamo la seconda alternativa, e aggiun-

giamo gli attributi ETA', SESSO e REGIONE all'entita' PERSONA (l'entita' OCCUPATO, che si trova in una gerarchia di generalizzazione con PERSONA, eredita tali attributi): a questo punto l'aggregazione e' compatibile coi dati disaggregati disponibili (vedi fig. 7.2). Non ci si puo' pero' limitare ad arricchire lo schema E/R: occorre anche modificare il questionario inserendovi i quesiti su eta', sesso e regione di residenza dell'intervistato.

Il disegno del sottoschema concettuale di aggregazione e' scomponibile logicamente in tre fasi:

1. scelta della classe di oggetti coinvolta nell'aggregazione: tale classe corrisponde ad una entita' o ad una relazione presente nello schema E/R dei dati disaggregati, con una eventuale restrizione espressa nell'etichetta;
2. specificazione degli attributi di categoria, delle loro aggregazioni e raggruppamenti;
3. scelta delle elaborazioni statistiche; le elaborazioni possono essere organizzate in gerarchie che esplicitano come determinate classi di dati vengono ottenute a partire da altre classi di dati.

A questo punto produciamo il sottoschema MCS dell'aggregazione (riportato nella fig. 7.3). Si noti che nella classe di oggetti OCCUPATI e' stata definita la restrizione "CON ETA' < 30", il che indica che non tutte le istanze dell'entita' OCCUPATO nello schema E/R verranno prese in considerazione ai fini dell'aggregazione, ma solo quelle il cui valore dell'attributo ETA' sara' inferiore a 30.

Consideriamo ora la seconda aggregazione (occupati nell'industria per sesso, titolo di studio ed eta', valori assoluti e percentuali rispetto al titolo di studio): anch'essa, da una immediata analisi dello schema E/R, risulta incompatibile. Anche in questo caso decidiamo di procedere aggiungendo l'attributo TITOLO DI STUDIO all'entita' PERSONA, e provvedendo ad inserire i corrispondenti quesiti nella bozza di questionario.

Il relativo sottoschema MCS e' riportato nella fig. 7.4.

Passiamo alla terza aggregazione richiesta (media delle ore lavorate per branca di attivita' e ripartizione geografica). Anch'essa e' incompatibile con i dati disaggregati a disposizione; puo' diventare ammissibile aggiungendo

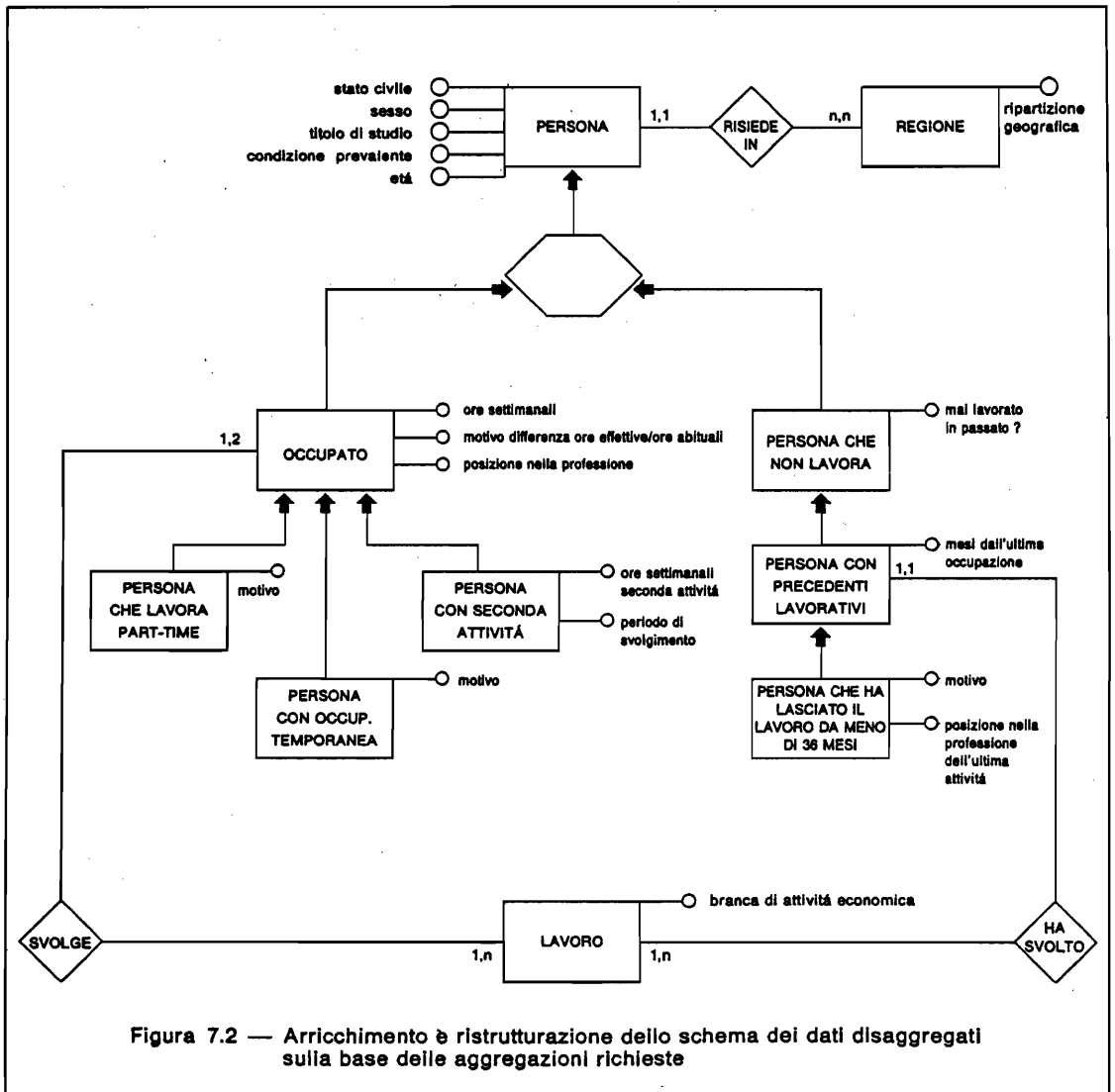


Figura 7.2 — Arricchimento e ristrutturazione dello schema dei dati disaggregati sulla base delle aggregazioni richieste

l'attributo RIPARTIZIONE GEOGRAFICA all'entità PERSONA (provvedendo cioè ad arricchire lo schema E/R). Ciò comporterebbe l'inserimento di un ulteriore quesito (relativo per l'appunto alla ripartizione) sul questionario: possiamo evitare questo ulteriore appesantimento del modello di rilevazione agendo come segue.

Procediamo dapprima ad una ristrutturazione dello schema E/R: REGIONE, da attributo di PERSONA, diviene un'entità distinta, collegata a PERSONA mediante la relazione RISIEME IN. Arricchiamo quindi lo schema aggiungendo l'attributo RIPARTIZIONE GEOGRAFICA all'entità REGIONE: ciò non comporta modifiche del questionario, in quanto il valore della ripartizione per ogni regione è noto a priori, e non deve quindi essere rilevato presso l'intervistato.

Resa ammissibile l'aggregazione, ne produciamo il relativo sottoschema MCS (vedi fig. 7.5).

Consideriamo, infine, l'ultima aggregazione (persone in cerca di lavoro per età e posizione nella professione): esaminando lo schema E/R dei dati disaggregati, modificato in seguito agli arricchimenti ed alle ristrutturazioni già effettuate (vedi fig. 7.2), si conclude che tale aggregazione è compatibile, in quanto gli oggetti "persone in cerca di lavoro" sono ricavabili a partire dall'entità PERSONA, considerando il valore dell'attributo CONDIZIONE PREVALENTE (se questo assume i valori 2 o 3 significa che la persona è in cerca di nuova occupazione o di prima occupazione).

Produciamo quindi il relativo sottoschema MCS di aggregazione, riportato nella fig. 7.6.

7.4. Integrazione dei sottoschemi

Lo scopo dell'integrazione è quello di unire i vari schemi di aggregazione sviluppati indipendentemente, ottenendo una rappresentazione integrata dell'insieme delle aggregazioni poste come requisiti. Per ottenere tale integrazione, il progettista deve scoprire tutte le parti comuni tra i vari schemi parziali e unificare la rappresentazione dei concetti che compaiono in più schemi.

La procedura per l'integrazione dei sottoschemi MCS di aggregazione è identica a quella mostrata per l'integrazione degli schemi

E/R di area nel cap.4. La sequenza dei passi per l'integrazione è riportata nella fig. 7.1c: viene scelta una prima coppia di sottoschemi di aggregazione e ne viene eseguita l'integrazione. Il nuovo schema prodotto (unificazione dei due sottoschemi originari) viene quindi a sua volta integrato con un altro tra gli schemi ottenuti nella fase precedente. Il procedimento viene ripetuto finché non sono stati considerati tutti i sottoschemi, fino a che, cioè, tutte le aggregazioni sono state espresse nello schema globale finale.

L'attività di integrazione di due sottoschemi di aggregazione può essere divisa in due momenti:

1. analisi e soluzione dei conflitti;
2. unificazione degli schemi;

I **conflitti** di rappresentazione dello stesso concetto in schemi diversi sono dovuti a due ragioni principali:

- sinonimia tra concetti: due concetti sono sinonimi quando rappresentano lo stesso aspetto della realtà con nomi diversi. Il problema delle sinonimie è in parte risolto automaticamente dalla disponibilità dello schema integrato dei dati disaggregati, nel quale le eventuali diverse denominazioni sono già state unificate.

- esistenza di concetti equivalenti, espressi però con diverse strutture di rappresentazione del modello MCS; tipico il caso di alcuni attributi che possono essere rappresentati come attributi semplici in uno schema (ad esempio: DATA), come attributi aggregati di attributi semplici in un altro (ad esempio: DATA come aggregato di GIORNO, MESE e ANNO).

Una volta risolti gli eventuali conflitti, possiamo procedere all'**unificazione** degli schemi parziali.

Il risultato finale dell'integrazione dei sottoschemi relativi alle aggregazioni considerate è lo schema MCS riportato nella fig. 7.7.

È possibile arricchire lo schema globale prendendo in considerazione eventuali "proprietà interschema".

Le **proprietà interschema** sono proprietà che non possono essere rilevate durante il disegno dei vari schemi parziali, ma solo dall'analisi di quello globale. Tipici esempi di proprietà interschema sono:

- **raggruppamenti** di attributi di categoria;
- **generalizzazioni** tra classi di dati (nuovi nodi G).

Si noti ad esempio che lo schema della fig. 7.8 e' un arricchimento di quella precedente in quanto in esso sono stati definiti un nuovo nodo G (quello relativo ai DATI SULLE FORZE LAVORO, generalizzazione dei dati sugli occupati e di quelli sulle persone in cerca di lavoro), e la relazione di raggruppamento (graficamente denotata da un arco contrassegnato da un asterisco) tra gli attributi REGIONE e RIPARTIZIONE GEOGRAFICA.

E' conveniente svolgere l'attivita' di definizione delle generalizzazioni delle classi di dati considerando l'albero delle aree definito nel corso

della progettazione dei dati disaggregati (vedi cap.4), che puo' aiutare molto nel definire le relazioni tra le classi dei dati aggregati, anche se non vi e' generalmente una corrispondenza piena tra tale albero e la gerarchia delle generalizzazioni dei dati.

7.5. Verifiche finali

Alla fine dell'attivita' di progettazione, anche nel caso dei dati aggregati, puo' essere necessaria un'analisi finale per controllare la qualita' complessiva dello schema globale; le qualita' che lo schema deve avere sono identiche a quelle considerate nel caso degli schemi E/R nel cap.4, e cioe' la correttezza, la completezza, la leggibilita' e la minimalita'.

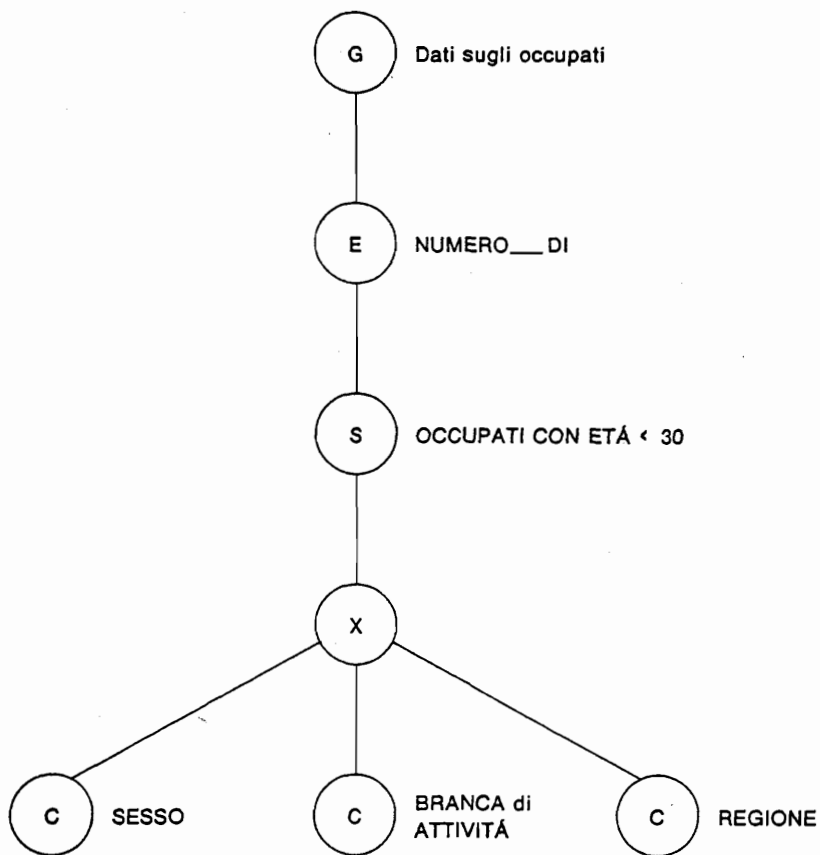


Figura 7.3 — Numero di occupati di età inferiore ai 30 anni per sesso, branca di attività e regione

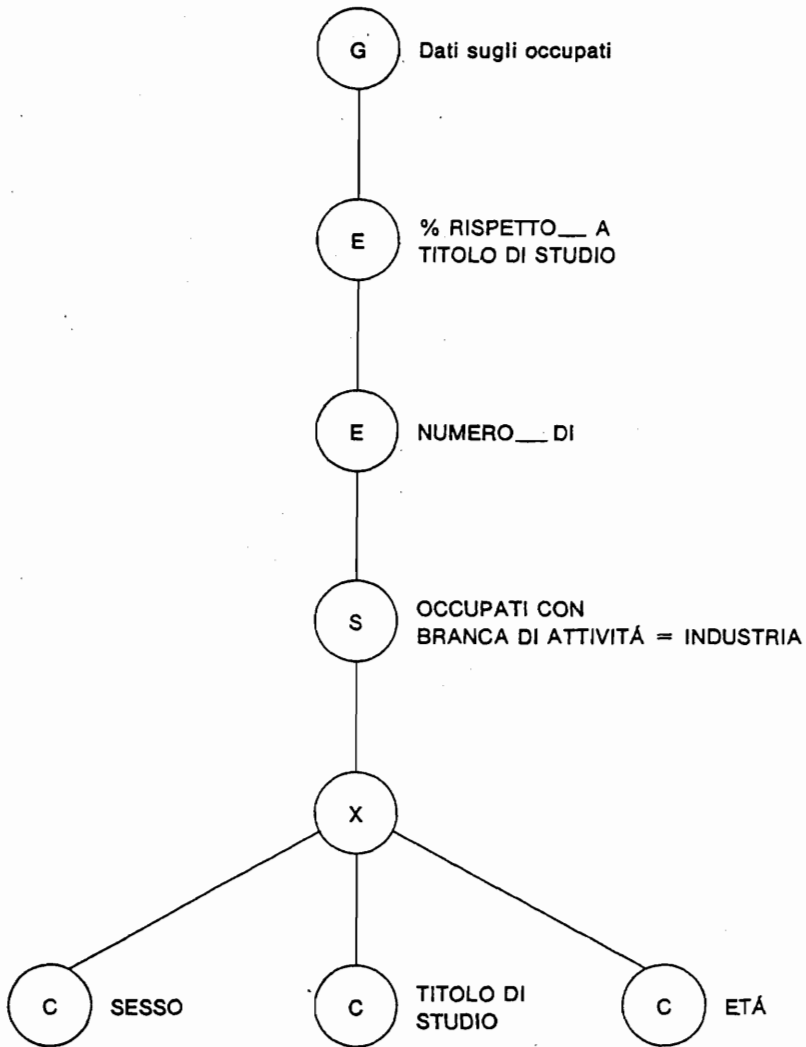
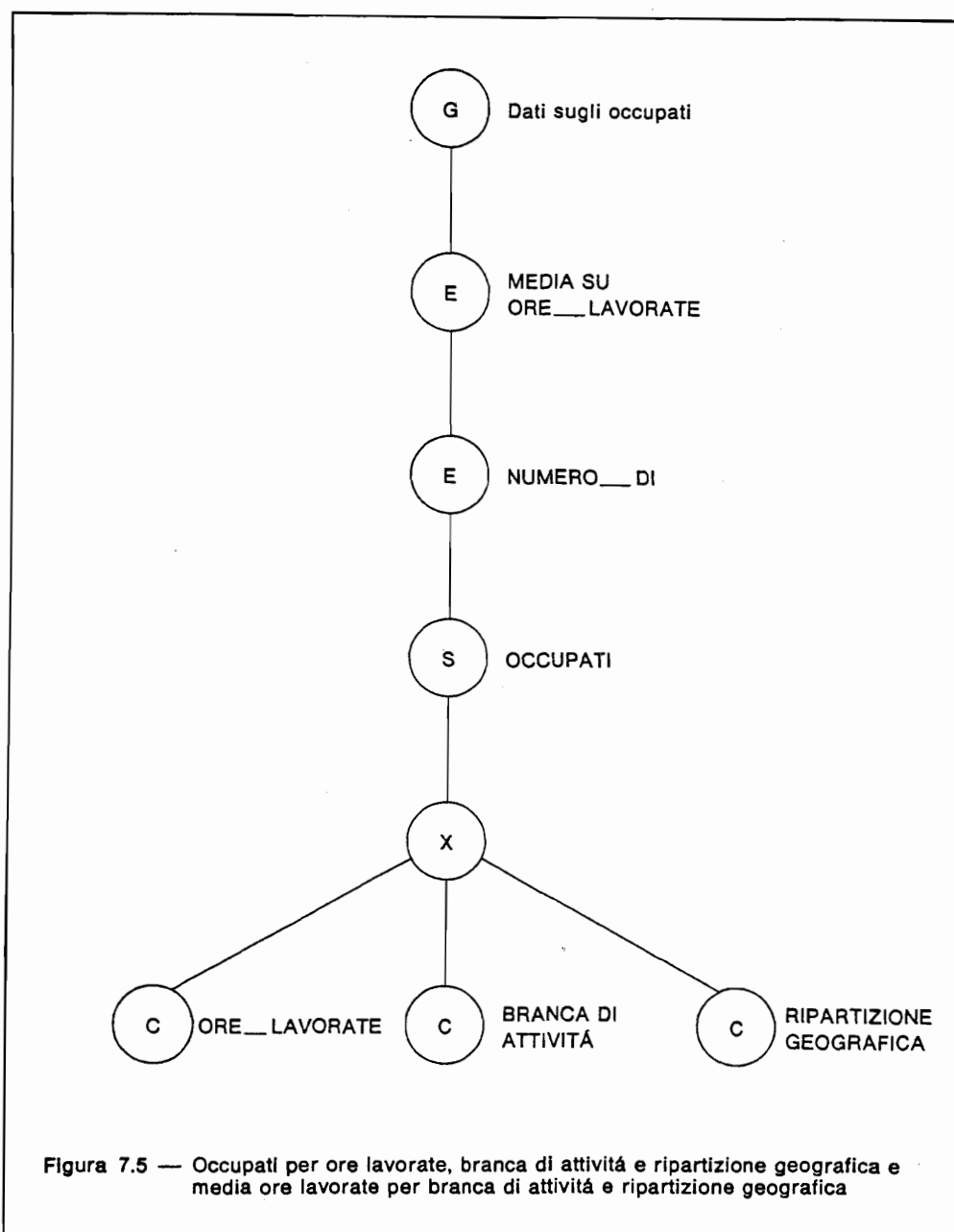
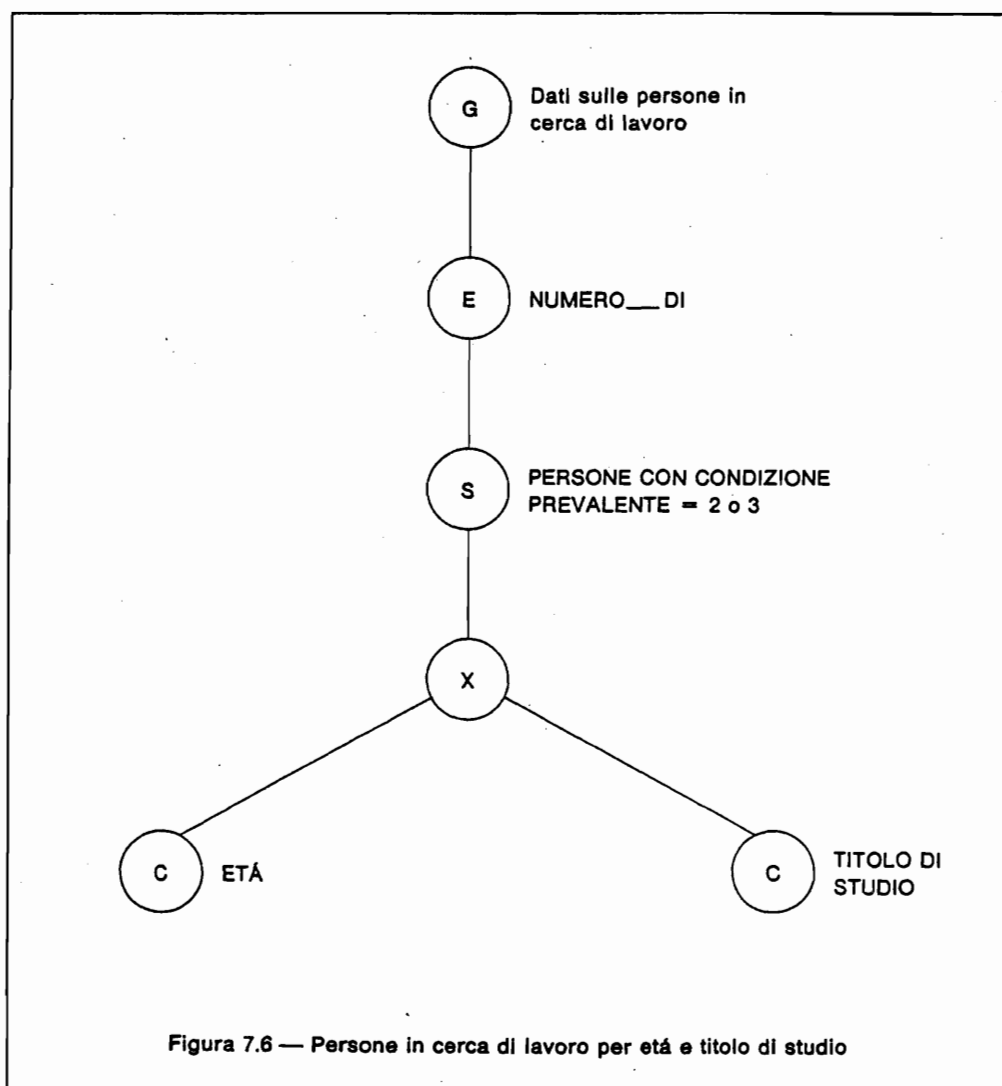
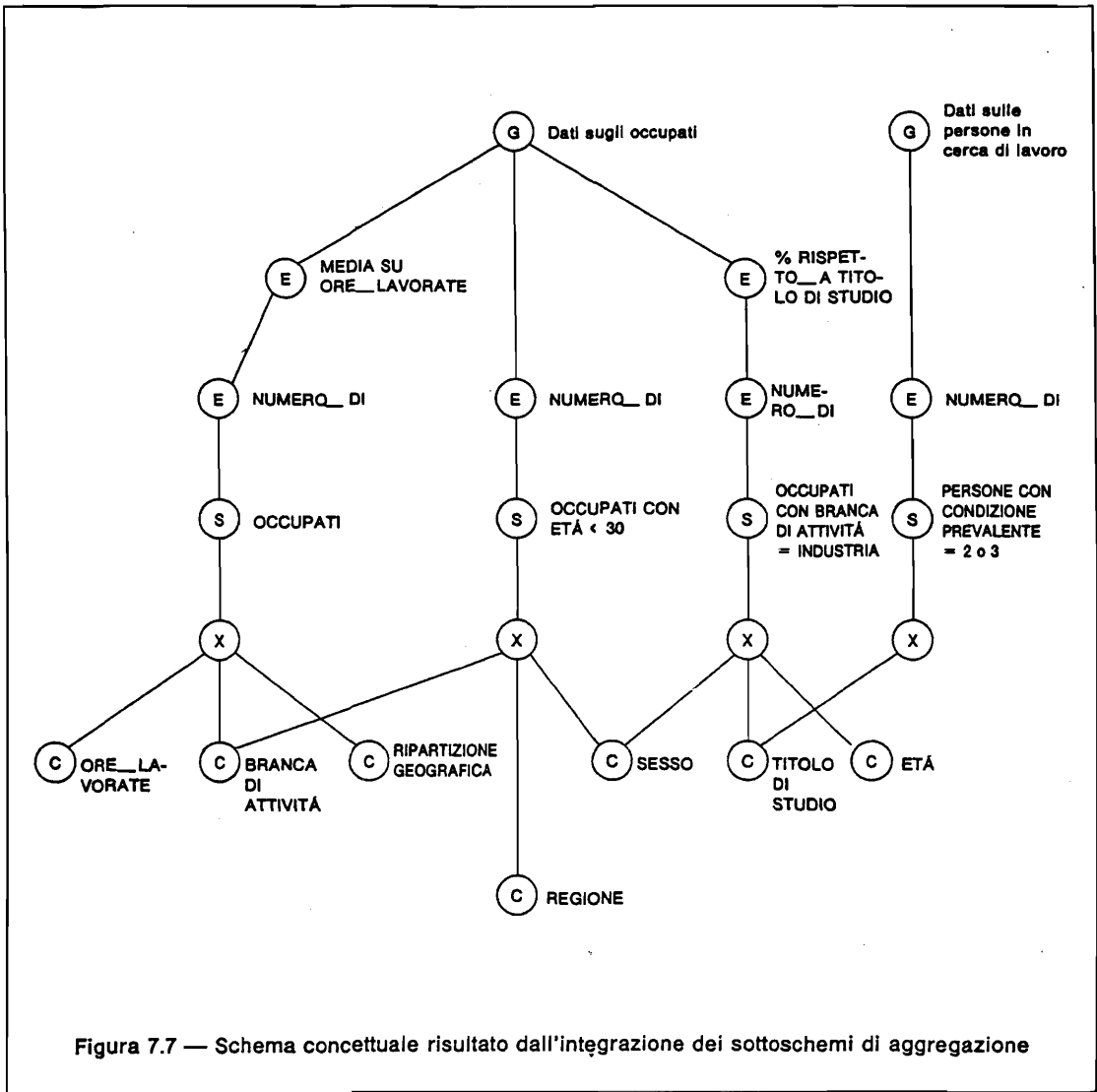


Figura 7.4 — Occupati nell'industria per sesso, titolo di studio ed età (valori assoluti e percentuali rispetto al titolo di studio)







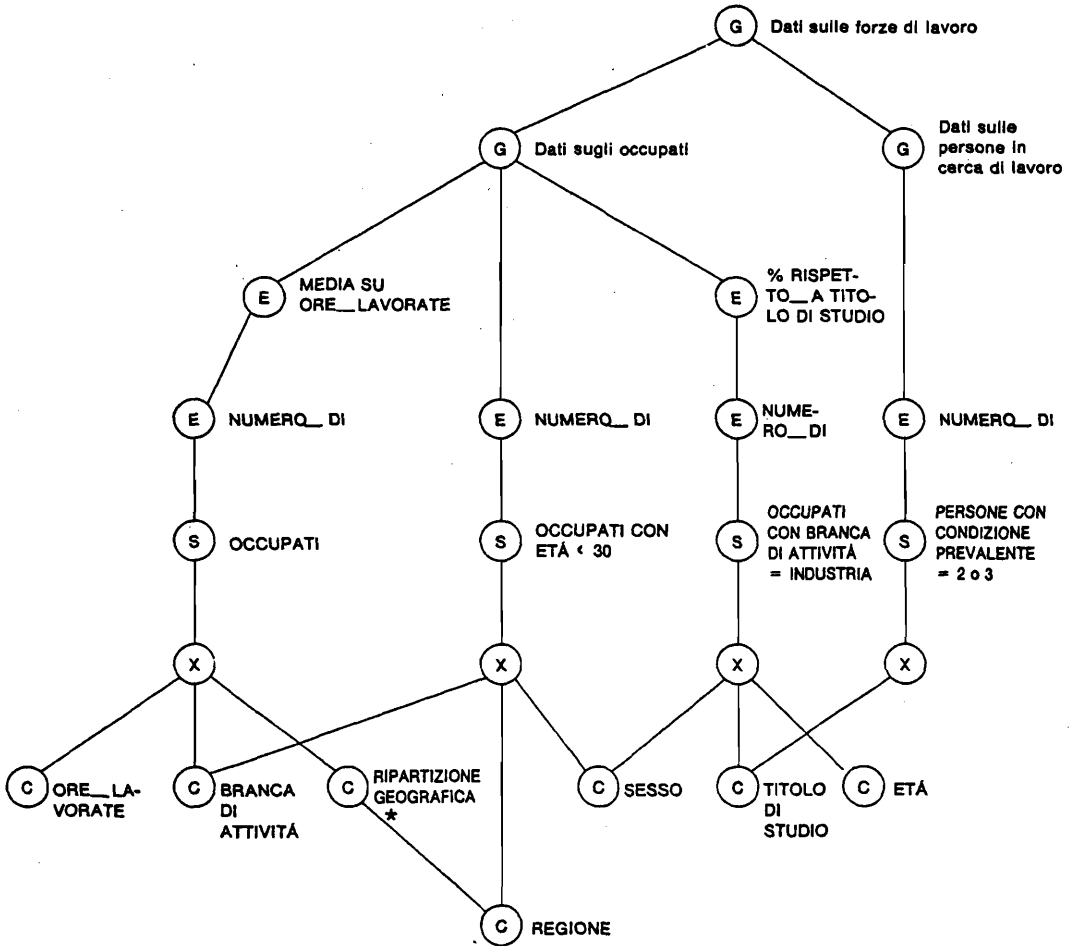


Figura 7.8 — Schema concettuale risultato dall'integrazione dei sottoschemi di aggregazione in seguito all'analisi delle proprietà interschema ed alla definizione della struttura delle generalizzazioni

APPENDICE A: LA DESCRIZIONE LINGUISTICA DI SCHEMI E/R

Illustriamo qui le modalita' per descrivere le strutture di rappresentazione del MODELLO E/R in termini linguistici, permettendo cosi' di arricchire la documentazione prodotta dalla progettazione concettuale. Gli esempi che verranno fatti si riferiscono alle classi ed ai vincoli illustrati nel paragrafo precedente.

Nel caso di uno Schema in cui la definizione delle entita' e' rappresentata da espressioni sintattiche di una certa complessita', un primo passo di una certa utilita' e' quello di associare ad ogni denominazione di ENTITA' un riferimento sintetico, nella forma di un acronimo, per identificare univocamente ed in maniera veloce le diverse entita' dello schema concettuale:

< acronimo entita' > < denominazione entita' >

L'insieme delle coppie (acronimo, denominazione) formera' una **Tabella degli Acronimi**, riferimento sia per il progettista che per chiunque si ponga di fronte allo Schema Concettuale. La sua definizione permette di avere la lista completa di tutte le entita' dello Schema. Per una documentazione omogenea e' consigliabile usare lo stesso acronimo contemporaneamente nella versione grafica che in quella sintattica. L'esempio della tabella A.1 mostra l'operazione di costruzione della tabella nel caso delle entita' illustrate in precedenza. Vale la pena notare che nel nostro caso le dizioni alfabetiche non presentano una particolare complessita' o lunghezza quindi il caso riportato ha solo un significato esemplificativo, ricordando che se puo' non essere necessario l'uso di un acronimo e' comunque sempre utile disporre di un elenco di tutte le entita' coinvolte nello Schema Concettuale.

TABELLA DEGLI ACRONIMI

< DIP > < dipendente >

< DIR > < dirigente >

< IMP > < impiegato >

< OPR > < operato >

< REP > < reparto >

Tabella A.1 - Tabella degli acronimi

Per gli ATTRIBUTI costruiamo un elenco composto da:

< codice > < denominazione attributo >

dove il codice e' un progressivo, unico per ogni attributo dell'intero Schema, utile per individuare lo stesso attributo associato a classi differenti (Tabella A.2).

LISTA DEGLI ATTRIBUTI

< 001 > < matricola >
 < 002 > < nome >
 < 003 > < cognome >
 < 004 > < sesso >
 < 005 > < nome reparto >
 < ... > < >

Tabella A.2 - Lista degli attributi

Nel caso si voglia rappresentare contemporaneamente le entita' con i propri attributi si potra' usare il seguente formalismo:

< **acronimo entita'** > < **denominazione entita'** >

ENTAGG

(lista attributi)

la notazione ENTAGG (ENTITA' AGGREGATO DI) sottolinea che l'entita' in oggetto e' aggregazione dell'insieme degli attributi elencati (Tabella A.3).

LISTA ENTITA' E ATTRIBUTI

< DIP > < dipendente >
ENTAGG
 < 001 > < matricola >
 < 002 > < nome >
 < 003 > < cognome >
 < 004 > < sesso >
 < ... > < >

Tabella A.3 - Lista entita' e attributi

L'espressione di ASTRAZIONI DI GENERALIZZAZIONE e' ottenuta attraverso la seguente sintassi:

< **acr. entita' padre** > **SOTDI** < **elenco acr. entita' figlio** >

dove per SOTDI si intende SOTTOINSIEME DI e sta ad individuare una RELAZIONE DI SOTTOINSIEME tra l'entita' padre e distintamente ciascuna entita' figlio; per esempio:

< **DIP** > **SOTDI** < **IMP** >

mentre in:

< **acr. entita' padre** > **GENDI** < **elenco acr. entita' figlio** >

GENDI sta per GENERALIZZAZIONE DI ed esprime la rappresentazione di una GERARCHIA DI GENERALIZZAZIONE in cui ogni entita' figlio e' una partizione dell'entita' padre; in tal modo ogni istanza della classe di ordine superiore e' anche istanza di almeno una delle classi di ordine inferiore. Nel nostro esempio:

**<DIP> GENDI <DIR>
 <IMP>
 <OPR>**

Per esprimere la classe RELAZIONE avremo:

**<nome relazione> REL <acr. entita' 1> <cmin,cmax>
 <acr. entita' 2> <cmin,cmax>**

il nome della relazione consiste generalmente in una forma verbale, in quanto essa consente di rappresentare il concetto espresso dalla relazione nei termini dei fatti che intercorrono tra le entita' coinvolte. REL rappresenta la dizione RELAZIONE, mentre cmin e cmax rappresentano rispettivamente la cardinalita' minima e massima cioe' il numero minimo e massimo di istanze coinvolte nella relazione. Per gli attributi della relazione valgono gli stessi formalismi usati per le entita':

**<LAVORA IN> REL <DIP> <1,n>
 <REP> <1,n>**

Per quanto riguarda la rappresentazione degli ATTRIBUTI AGGREGATI abbiamo:

<denom. attr. aggregato> ATTAG <elenco attributi aggregati>

con ATTAG intendiamo ATTRIBUTO AGGREGATO DI; ad esempio:

<periodo di appartenenza> ATTAG <data inizio> <data fine>

Per l'espressione del dominio ammissibile per i singoli attributi utilizziamo la seguente sintassi:

<codice attributo> <denominazione> INSDI <dominio>

dove INSDI sta per INSIEME DI. Per la descrizione dei domini, possono essere utilizzati due diversi standard: il primo, piu' generico, individua solo il tipo di dati ammessi:

- interi (INTEGER)
- logici (vero o falso) (BOOLEAN)
- stringhe di caratteri (STRING)

il secondo, di maggior dettaglio, riporta l'elenco completo dei valori ammessi:

<004> < sesso> INSDI <Maschio, Femmina>

APPENDICE B: LE OPERAZIONI SUI DATI DISAGGREGATI

Ricordiamo che obiettivo primario della progettazione concettuale e' quello di costruire uno schema composto da un insieme di strutture di rappresentazione e, se questo e' relativamente semplice per i dati ed i vincoli di integrita', risulta invece piu' complesso e quindi meno preciso per esprimere l'evoluzione di queste strutture in termini di operazioni. La loro esplicitazione anche se meno espressiva di quella dei dati risulta comunque utile innanzitutto per verificare, alla luce dei dati coinvolti dalle operazioni, la completezza dello schema concettuale dei dati. In secondo luogo puo' essere utilizzato come un primo abbozzo di analisi per la realizzazione, nel seguito, delle procedure applicative. I concetti qui illustrati fanno riferimento al MODELLO E/R, o meglio a cio' che in precedenza e' stato definito come arricchimento del modello originario.

Specifichiamo che per operazione intendiamo la rappresentazione di procedure che operano sugli oggetti che compongono la realta' di interesse; in termini di Schema Concettuale cio' significa la rappresentazione di operazioni di manipolazione sulle istanze delle classi definite. Nell'ambito del MODELLO E/R l'operazione e' detta **TRANSAZIONE**. La descrizione di una transazione e' possibile grazie all'uso di "operazioni elementari", in pratica la possibilita' di aggiungere o eliminare istanze di entita' o relazioni:

INSERISCI-ent (x,X) inserisce l'istanza x nella entita' X

INSERISCI-rel (x,X) inserisce l'istanza x nella relazione X

CANCELLA-ent (x,X) cancella l'istanza x nella entita' X

CANCELLA-rel (x,X) cancella l'istanza x nella relazione X

Nell'uso di queste operazioni si pone la restrizione che vengano rispettati i vincoli impliciti del modello. Vediamo due semplici esempi in cui descriviamo, in termini generali, prima l'inserimento poi la cancellazione di una istanza da una entita'; negli esempi faremo uso di un formalismo riconducibile ad un generico linguaggio di programmazione:

Operazione elementare **INSERISCI-ent (x,X)**:

Per ogni entita' Y di ordine superiore all'entita' X

ESEGUI

Se x non e' presente nelle istanze di Y

Allora

INSERISCI-ent (x,X)

INSERISCI-ent (x,Y)

oppure

INSERISCI-ent (x,X)

FINE

dove si fa salva la regola per cui le istanze di ogni entita' figlio, in una relazione di sottoinsieme tra classi, appartengono anche a quella della entita' padre.

Vediamo ora il caso della cancellazione:

Operazione elementare CANCELLA-ent (x,X)

Se x appartiene alle istanze di X

Allora SEQUENZA

ESEGUI

CANCELLA-ent (x,X)

Per ogni entita' Y sottoinsieme di X

CANCELLA-ent (x,Y)

Per ogni istanza r della relazione R

dove e' presente l'istanza x

CANCELLA-rel (r,R)

FINE SEQUENZA

dove, oltre al vincolo che discende dall'esistenza di una relazione di sottoinsieme tra classi, viene rispettato il vincolo relativo ad eventuali relazioni che legano l'istanza cancellata con istanze di un'altra entita'.

Distinguiamo le operazioni secondo il tipo:

1. la "sequenza", cioe' l'esecuzione una dopo l'altra di tutte le istruzioni specificate;
2. la "alternativa" per cui le istruzioni che seguono una determinata condizione saranno eseguite solo al suo verificarsi, in caso contrario non sara' eseguita nessuna istruzione o verranno svolte delle istruzioni alternative;
3. il "ciclo" che implica la ripetizione di una o piu' istruzioni fino al verificarsi di una certa condizione.

Vale la pena notare che nell'esempio precedente, CANCELLA-ent (x,X), sono presenti contemporaneamente le tre operazioni definite:

- la "alternativa" nei termini SE...ALLORA

- la "sequenza" con il comando SEQUENZA che indica di svolgere le istruzioni che seguono

il "ciclo" con la notazione PER OGNI.

In generale il formalismo per rappresentare una TRANSAZIONE potra' essere:

TRANSAZIONE < nome > < parametri >
Tipo di operazione
< lista istruzioni >
Fine operazione

dove il nome identifica l'operazione svolta, mentre con < parametri > si identificano eventuali parametri utili per lo svolgimento della transazione. Vediamo un esempio in relazione alla figura 6.2: sia da specificare l'operazione di rilevazione di una nuova forza lavoro; la transazione sara' del tipo:

TRANSAZIONE AGGIUNGI-PERS (x:PERSONA)

SEQUENZA

INSERISCI-ent (x,PERSONA)

Se attributo CONDIZIONE = 1

allora INSERISCI-ent (x, OCCUPATO)

INSERISCI-rel (x, SVOLGE)

oppure

INSERISCI-ent (x, PERSONA CHE NON LAVORA)

FINE SEQUENZA

dove l'inserimento di una nuova forza lavoro (istanza, come specificato nei parametri, dell'entità PERSONA) deve prevedere un suo inserimento anche nell'entità OCCUPATO (e quindi nella relazione SVOLGE con l'entità LAVORO), oppure nell'entità PERSONA CHE NON LAVORA.

APPENDICE C: LA DESCRIZIONE LINGUISTICA DI SCHEMI MCS

Così come esiste la possibilità di descrivere schemi E/R di dati disaggregati mediante un linguaggio formale (vedi Appendice A), in modo analogo è possibile definire una sintassi per l'espressione formale di uno schema MCS.

La forma generale di tale sintassi è riportata nella tabella C.1.

Valgono le seguenti regole:

- nella lista degli attributi di categoria, vanno indicati quelli di tipo multiplo secondo il seguente formalismo:

< attributo > M

- nella lista delle relazioni di raggruppamento, ogni relazione deve seguire la forma:

< attributo > RAGGRUPPATO IN < attributo >

con l'indicazione di una eventuale regola di raggruppamento;

- nella lista degli attributi aggregati, per ognuno di essi va seguita la forma:

< attributo > AGGREGATO DI < lista attributi >

- ogni classificazione statistica deve essere indicata nel modo seguente:

< numero classificazione > : < lista attributi collegati >

- ogni classe di dati primitiva segue la forma:

< numero classe di dati > : < nome classe di dati >

PRODOTTA DA < classe di oggetti >

| < PARAMETRO > < lista attributi > |

USANDO OPERATORE < nome operatore >

| USANDO CLASSIFICAZIONE < numero classificazione > |

- ogni classe di dati derivata segue la forma:

< numero classe di dati > : < nome classe di dati >

PRODOTTA DA < lista numeri classi di dati >

| < PARAMETRO > < lista attributi > |

USANDO OPERATORE < nome operatore >

| USANDO CLASSIFICAZIONE < numero classificazione > |

- infine, nella lista delle generalizzazioni, per ognuna di esse occorre indicare:

< nome generalizzazione > GENERALIZZAZIONE DI

< lista generalizzazioni o numeri di classi di dati >

Attraverso un linguaggio formale quale quello introdotto, è possibile stabilire un rapporto biunivoco tra schemi grafici e descrizioni linguistiche. Queste ultime sono facilmente elaborabili mediante calcolatore, il che permetterà, disponendo di appositi tools, il trattamento automatico delle corrispondenze tra i diversi schemi.

Avendo come riferimento la figura 5.16, la corrispondente descrizione linguistica è riportata nella tabella C.2.

SCHEMA

< nome schema >

CLASSI DI OGGETTI

< lista classi di oggetti >;

ATTRIBUTI DI CATEGORIA

< lista attributi di categoria >;

RELAZIONI DI RAGGRUPPAMENTO DI ATTRIBUTI< lista relazioni di raggruppamento
tra attributi di categoria con regola di raggruppamento >;**ATTRIBUTI AGGREGATI**

< lista attributi aggregati con attributi dipendenti >;

CLASSIFICAZIONI STATISTICHE

< lista classificazioni statistiche >;

CLASSI DI DATI**PRIMITIVE**< lista classi di dati prodotti da elaborazioni
su classi di oggetti >;**DERIVATE**< lista classi di dati prodotti da elaborazioni
su classi di dati >;**GENERALIZZAZIONI**

< lista delle generalizzazioni >;

FINE-SCHEMA < nome-schema >

Tabella C.1 - Sintassi per l'espressione di schemi MCS

SCHEMA persone abitazioni**CLASSI DI OGGETTI**

persone, abitazioni

ATTRIBUTI DI CATEGORIA

sesso, comune, provincia, numero stanze, classi di eta'

RELAZIONI DI RAGGRUPPAMENTOcomune **RAGGRUPPATO IN** provincia**CLASSIFICAZIONI STATISTICHE**

1: sesso, comune, classi di eta'

2: numero stanze, comune

CLASSI DI DATI**PRIMITIVE**1: Numero di persone **PRODOTTA DA** persone
USANDO OPERATORE numero-di
USANDO CLASSIFICAZIONE 12: Numero di abitazioni **PRODOTTA DA** abitazioni
USANDO OPERATORE numero-di
USANDO CLASSIFICAZIONE 2**DERIVATE**3: Indice di affollamento **PRODOTTA DA** 1,2
USANDO OPERATORE rapporto **PER** comune**GENERALIZZAZIONI****DATI SU** persone **GENERALIZZAZIONE DI** 1**DATI SU** abitazioni **GENERALIZZAZIONE DI** 2**DATI SU** persone e abitazioni **GENERALIZZAZIONE DI****DATI SU** persone, **DATI SU** abitazioni**FINE-SCHEMA** persone abitazioni

Tabella C.2 - Descrizione linguistica di uno schema MCS

GLOSSARIO

aggregato di attributi di categoria - nel modello MCS, un attributo di categoria composto da due o più attributi di categoria semplici

aggregazione (statistica) - espressione che individua un dato statistico a partire dai dati elementari e dalle variabili coinvolte, e le eventuali elaborazioni da effettuare sui dati stessi

albero delle generalizzazioni - nel modello MCS, una gerarchia di nodi G che individua le generalizzazioni tra le classi di dati dello schema

astrazione - procedimento mentale attraverso il quale è possibile giungere alla definizione di un oggetto di ordine superiore a partire da un insieme di oggetti considerati omogenei rispetto ad un insieme di proprietà

astrazione di classificazione - definizione di una classe a partire da oggetti aventi proprietà in comune

astrazione di aggregazione - definizione di una classe a partire da altre classi viste come componenti della prima

astrazione di generalizzazione - definizione di una classe a partire da altre classi viste come sottoinsiemi della prima

attributo - nel modello E/R, proprietà elementare di un'entità

attributo identificatore - nel modello E/R, attributo o insieme di attributi mediante il quale è possibile identificare univocamente le singole istanze di un'entità

attributo multiplo - nel modello E/R, attributo ad ogni istanza del quale è possibile associare più valori

attributo aggregato - nel modello E/R, attributo composto da due o più attributi semplici

attributo di categoria - nel modello MCS, la variabile utilizzata per classificare statisticamente classi di oggetti e classi di dati

cardinalità minima e massima - nel modello E/R, il numero minimo e massimo di volte che ogni istanza di una entità è coinvolta in una relazione

classe di dati - nel modello MCS, l'insieme di dati ottenuto mediante applicazione di funzioni di elaborazione a classi di oggetti o ad altre classi di dati; nel primo caso avremo classi di dati primitive, nel secondo classi di dati derivate

classe di oggetti - in uno schema concettuale di dati, concetto che rappresenta un insieme di oggetti aventi comuni proprietà

classificazione statistica - nel modello MCS, la relazione intercorrente tra una classe di oggetti o una classe di dati, e l'insieme di attributi di categoria utilizzati per classificarli

dati aggregati - nel processo produttivo statistico, dati ottenuti da elaborazioni di dati disaggregati

dati disaggregati - dati in ingresso al processo produttivo statistico, possono o meno coincidere coi dati elementari

dati elementari - dati raccolti presso le unità di rilevazione dell'indagine

dominio - riferito ad attributi, attributi di categoria, variabili e parametri, indica l'insieme di valori che essi possono assumere

elaborazione statistica - nel modello MCS, una funzione che, applicata a classi di oggetti o a classi di dati, produce una nuova classe di dati; è caratterizzata dall'utilizzo di operatori monadici o poliadici

entità - nel modello E/R indica una classe di oggetti

etichetta - nel modello MCS, una definizione associabile ai nodi dello schema dei dati per poterne descrivere le caratteristiche

funzione di aggregazione - nel processo produttivo statistico, funzione statistico-matematica applicata ai dati disaggregati

generalizzazione - nel modello MCS, l'astrazione di generalizzazione tra classi di dati

gerarchia di generalizzazione - nel modello E/R, relazione tra piu' entita' di ordine inferiore ed una di ordine superiore, tale che ogni istanza di quest'ultima e' anche istanza di una ed una sola entita' tra quelle di ordine inferiore

grouping - vedi 'raggruppamento di attributi'

Istanza - elemento di una classe di oggetti

metadati - l'insieme di informazioni che permettono di comprendere il significato dei dati cui si riferiscono

metodo bottom-up - metodo che permette la definizione di oggetti di ordine superiore a partire da quelli di ordine inferiore

metodo top-down - metodo che permette la definizione di oggetti di ordine inferiore a partire da quelli di ordine superiore

metodologia di progettazione - insieme di fasi e di strumenti operativi che permettono l'esecuzione di attivita' di progettazione in modo efficiente ed efficace

Metodologia Integrata per la rappresentazione dei DATI Statistici (MIDAS) - metodologia di progettazione concettuale sia dei dati disaggregati che di quelli aggregati

modello concettuale dei dati - modello che permette di fornire rappresentazioni dei dati indipendenti dalle caratteristiche dall'ambiente tecnologico

Modello Concettuale Statistico (MCS) - modello per la progettazione concettuale dei dati aggregati

modello Entita'/Relazione (E/R) - modello concettuale per la rappresentazione dei dati disaggregati

operatore - nell'etichetta di un'elaborazione statistica di uno schema MCS rappresenta una particolare funzione matematico-statistica; l'operatore puo' essere monadico se applicato

ad un'unica classe di oggetti o classe di dati, poliadico altrimenti

operazione - manipolazione eseguita su istanze di una o piu' classi di oggetti

processo produttivo statistico - processo di produzione di informazioni statistiche caratterizzato da attivita' di trasformazione dei dati in ingresso (dati disaggregati) per ottenere dati in uscita (dati aggregati)

progettazione concettuale dei dati - fase della costruzione dei sistemi informativi avente l'obiettivo di fornire una rappresentazione formale dei dati, non ambigua ed indipendente dall'ambiente tecnologico di sviluppo

progettazione logica dei dati - fase della costruzione dei sistemi informativi avente l'obiettivo di tradurre gli schemi concettuali dei dati in schemi compatibili con l'ambiente tecnologico di sviluppo

progettazione fisica dei dati - fase della costruzione dei sistemi informativi avente l'obiettivo, sulla base dello schema logico dei dati, del software e dell'hardware disponibili, di scegliere i parametri fisici di memorizzazione, il tipo e l'organizzazione dei supporti da utilizzare, i metodi di accesso, ecc. in modo da garantire l'efficienza delle applicazioni

proprieta' interschema - proprieta' dello schema concettuale integrato, non presente nei diversi schemi parziali e evidenziata durante l'attivita' di integrazione degli schemi stessi

raggruppamento di attributi - nel modello MCS, una relazione gerarchica che lega due attributi di categoria quando e' verificata la seguente proprieta': i valori dell'attributo di livello superiore determinano una partizione in sottoinsiemi del dominio dell'attributo di livello inferiore, e tali insiemi sono disgiunti, non hanno cioe' valori comuni (esempio: REGIONE-PROVINCIA)

relazione - nel modello E/R, legame logico tra due o piu' entita'

relazione di sottoinsieme - nel modello E/R, si ha una relazione di sottoinsieme tra due entita' se tutte le istanze della seconda sono anche istanze della prima, ma non vale il viceversa

requisiti - in un sistema informativo, l'insieme dei dati di interesse (requisiti statici), delle operazioni sui dati, degli eventi che regolano

l'ordine e la causalità delle operazioni (requisiti dinamici), così come sono descritti dagli utenti

ridondanza - in uno schema concettuale, indica la presenza di concetti la cui eliminazione non ne riduce il contenuto informativo

schema concettuale - rappresentazione formale, integrata e indipendente dalla tecnologia, dei dati e delle operazioni di interesse di una applicazione; output della fase di progettazione concettuale

sistema informatico - insieme degli strumenti informatici per la gestione automatizzata di un sistema informativo

sistema informativo - insieme strutturato di risorse umane, di strumenti di memorizzazione, scambio, elaborazione e diffusione di informazioni, di procedure manuali ed automatizzate

per il trattamento delle informazioni, e di regole organizzative per la loro gestione

struttura di rappresentazione - struttura simbolica di cui un modello fa uso per descrivere una data realtà

transazione - nel modello E/R, descrizione formale di un'operazione

vincoli di integrità - in un modello concettuale dei dati, le proprietà che le istanze appartenenti ad una data classe debbono soddisfare per rappresentare correttamente la realtà di interesse

vincoli espliciti - in uno schema concettuale dei dati, vincoli dichiarati esplicitamente mediante un formalismo ad hoc

vincoli impliciti - in uno schema concettuale dei dati, vincoli implicitamente descritti dalle strutture di rappresentazione utilizzate

PRINCIPALI PUBBLICAZIONI ISTAT

PUBBLICAZIONI MENSILI E TRIMESTRALI (1)

BOLLETTINO MENSILE DI STATISTICA

La più completa ed autorevole raccolta di dati congiunturali concernenti l'evoluzione dei fenomeni demografici, sociali, economici e finanziari.

Abbonamento annuo L. 101.000 (Estero L. 124.000) Ogni fascicolo L. 13.000

INDICATORI MENSILI

Forniscono dati riassuntivi e tempestivi sull'andamento mensile dei principali fenomeni interessanti la vita nazionale.

Abbonamento annuo L. 23.000 (Estero L. 30.000) Ogni fascicolo L. 3.000

NOTIZIARIO ISTAT

Fornisce i primi risultati delle rilevazioni ed elaborazioni statistiche riguardanti l'attività produttiva, i prezzi, il commercio interno, gli scambi internazionali come pure lo stato ed il movimento della popolazione e le sue caratteristiche sociali e sanitarie.

I dati, esposti in grafici e tabelle, sono accompagnati da commenti, illustrazioni e note interpretative.

Serie 1 - Statistiche demografiche e sociali; abbonamento annuo L. 17.000 (Estero L. 25.000) una copia L. 1.200

Serie 2 - Statistiche dell'attività produttiva; abbonamento annuo L. 54.000 (Estero L. 74.000) una copia L. 1.200

Serie 3 - Statistiche del lavoro, delle retribuzioni e dei prezzi; abbonamento annuo L. 19.000 (Estero L. 25.000) una copia L. 1.200

Serie 4 - Argomenti vari; abbonamento annuo L. 10.000 (Estero L. 15.000) una copia L. 1.200. Tutte le serie: abbonamento annuo L. 93.000 (Estero L. 130.000)

INDICATORI TRIMESTRALI

Conti economici trimestrali

Abbonamento annuo L. 9.000 (Estero L. 11.000) Ogni fascicolo L. 3.000

STATISTICA DEL COMMERCIO CON L'ESTERO

Documentazione statistica ufficiale, a periodicità trimestrale, sul commercio dell'Italia con l'estero; fornisce, per tutte le merci comprese nella classificazione merceologica della tariffa dei dazi doganali, l'andamento delle importazioni e delle esportazioni da e per i principali Paesi.

Abbonamento annuo L. 87.000 (Estero L. 105.000) Ogni fascicolo L. 27.000

Abbonamento annuo cumulativo a tutti i periodici, compresa la «Statistica del commercio con l'estero»: L. 276.000 (Estero L. 358.000); esclusa la «Statistica del commercio con l'estero»: L. 189.000 (Estero L. 253.000)

Gli abbonamenti decorrono dal 1° gennaio anche se sottoscritti nel corso dell'anno. In tal caso l'abbonato riceverà i numeri dell'annata già pubblicati. L'abbonato ai periodici ISTAT ha diritto a ricevere gratuitamente i fascicoli non pervenutigli soltanto se ne segnalerà il mancato arrivo entro 10 giorni dal ricevimento del fascicolo successivo. Decorso tale termine, si spediscono solo contro rimessa dell'importo. Le variazioni di indirizzo devono essere segnalate dall'abbonato per iscritto. Nel sottoscrivere l'abbonamento cumulativo, gli interessati possono chiedere che l'ISTAT provveda, senza ulteriori richieste, all'invio di tutte le pubblicazioni non periodiche non appena liberate dalle stampe, contro assegno o con emissione di fattura, con lo sconto del 30%. Le singole pubblicazioni possono essere richieste direttamente all'Istituto centrale di statistica (Via Cesare Balbo, 16 - 00100 Roma) versando il relativo importo, maggiorato del 10% per spese di spedizione, sul c/c postale n. 619007.

(1) I prezzi dei singoli fascicoli ed i relativi canoni di abbonamento sono riferiti all'anno 1989.

PUBBLICAZIONI ANNUALI

ANNUARIO STATISTICO ITALIANO - Edizione 1988 L. 35.000

Sintetizza i risultati delle rilevazioni ed elaborazioni statistiche di maggior interesse nazionale.

COMPENDIO STATISTICO ITALIANO - Edizione 1988 L. 16.000

Sintetizza in semplici tabelle numeriche di facile lettura ed attraverso appropriate note illustrative e rappresentazioni grafiche, i dati fondamentali della vita economica, demografica e sociale e fornisce un quadro panoramico della corrispondente situazione degli altri principali Paesi del mondo.

I CONTI DEGLI ITALIANI - Vol. 22, edizione 1988 L. 12.000

Illustra in forma divulgativa i principali aspetti quantitativi dell'economia italiana.

STATISTICHE DEMOGRAFICHE

Vol. 32, ed. 1988, Tomo 2 Parte prima - Nascite e decessi - Anni 1982-83 L. 33.500

Tomo 2 - Parte seconda - Matrimoni, separazioni e scioglimenti del matrimonio - Anni 1982-83 L. 12.600

Vol. 35, ed. 1988, Tomo 1 Parte seconda - Iscrizioni e cancellazioni anagrafiche per trasferimento di residenza - Anni 1985 e 1986 - Espatriati e rimpatriati - Anno 1986 L. 15.800

Vol. 36, ed. 1988, Tomo 1 Parte prima - Movimento e calcolo della popolazione residente secondo gli atti anagrafici - Anno 1987 L. 18.900

STATISTICHE SANITARIE

Vol. 28, edizione 1988 - Dati 1982 L. 33.500

Statistiche ospedaliere, malattie infettive, interruzioni volontarie della gravidanza e cause di morte.

Vol. 29, edizione 1988 - Dati 1983

Tomo 1 - Statistiche ospedaliere, malattie infettive, interruzioni volontarie della gravidanza L. 15.800

Tomo 2 - Statistiche sulle cause di morte e di nati/mortalità L. 22.000

Vol. 30, edizione 1988 - Dati 1984

Tomo 2 - Statistiche sulle cause di morte e di nati-mortalità L. 22.000

STATISTICHE DELLA PREVIDENZA, DELLA SANITÀ E DELL'ASSISTENZA SOCIALE

Vol. 26, edizione 1988 L. 15.800

Pubblicazione nella quale vengono illustrate alcune forme di attività svolte dai vari istituti nel settore della previdenza sociale, i conti economici delle unità sanitarie locali e degli istituti ospedalieri pubblici, nonché i principali aspetti dell'assistenza sociale - Dati 1985-86

STATISTICHE DELL'ISTRUZIONE

Vol. 38 - Dati anno scolastico 1984-85

Tomo 1 - Dati analitici: nazionali, regionali e provinciali L. 21.000

Tomo 2 - Dati riassuntivi comunali L. 15.800

Vol. 39, edizione 1988 - Tomo 2 - Dati riassuntivi comunali anno scol. 1985-86 L. 15.800

Quadro statistico completo ed aggiornato della situazione scolastica del Paese, attraverso dati sui vari rami d'insegnamento esaminati sotto i più interessanti aspetti dell'ordinamento degli studi e dei risultati conseguiti dagli iscritti.

STATISTICHE CULTURALI - Vol. 28, edizione 1989 L. 14.000

Documentazione ufficiale completa sulle principali attività culturali concernenti, tra l'altro, la produzione libraria, la pubblicazione di riviste scientifiche, la stampa periodica e le biblioteche nell'anno 1986.

STATISTICHE GIUDIZIARIE - Vol. 34, edizione 1988 L. 33.500

Ampia documentazione statistica dell'attività giudiziaria nonché dei principali fenomeni in materia civile e penale nel campo della criminalità e degli Istituti di prevenzione e pena, nell'anno 1986.

STATISTICHE AGRARIE - Vol. 32, edizione 1988 L. 18.900

Riunisce i dati significativi dell'economia agricola italiana con particolare riguardo alle coltivazioni - Dati 1984

STATISTICHE FORESTALI - Vol. 39, edizione 1988 L. 12.600

Fornisce un quadro completo sulla struttura delle foreste italiane e delle relative utilizzazioni legnose, unitamente ad alcuni aspetti economici - Dati 1985

STATISTICHE METEOROLOGICHE - Vol. 24, edizione 1988 L. 15.800

Raccoglie i dati relativi alle temperature, piovosità e altri fattori climatici rilevati da una rete di stazioni ed osservatori distribuiti nel territorio nazionale nell'anno 1983.

STATISTICHE DELLA CACCIA, PESCA E COOPERAZIONE - Vol. 2, edizione 1988 L. 8.400

Raccoglie i dati sull'attività della pesca e sulla consistenza del relativo naviglio, su alcuni aspetti del settore venatorio, nonché sulla consistenza e l'attività degli impianti collettivi dei principali settori agricolo-zootecnici.

STATISTICHE DELLA ZOOTECCNIA, PESCA E CACCIA - Vol. 26, edizione 1988 - Dati 1986 L. 15.800

Raccoglie i dati sulla consistenza e produttività degli allevamenti, sull'attività della pesca e sulla consistenza del relativo naviglio, nonché su alcuni aspetti del settore venatorio dell'anno 1984.

STATISTICHE INDUSTRIALI - Vol. 27, edizione 1987 L. 15.000

Nel suo genere, unica e veramente preziosa pubblicazione in cui sono organicamente raccolte tutte le informazioni statistiche fondamentali concernenti il complesso ed importante settore dell'industria - Dati 1986

STATISTICHE DELL'ATTIVITÀ EDILIZIA E DELLE OPERE PUBBLICHE - Vol. 28, edizione 1988 L. 15.800

Statistica ufficiale dedicata esclusivamente al settore dell'attività edilizia e delle opere pubbliche effettuate dallo Stato e da Enti pubblici, nonché da privati con finanziamento parziale dello Stato - Dati 1985

STATISTICHE DEL COMMERCIO INTERNO - Anni 1985-86 Vol. 28, edizione 1988 L. 15.800

Fornisce, con riferimento all'anno 1986, i risultati delle rilevazioni correnti relativi al fenomeno della distribuzione. Vi figurano gli indici mensili delle vendite al minuto, nonché la più recente distribuzione per Comune delle licenze di esercizio

STATISTICHE DEL TURISMO - Vol. 1, edizione 1988 L. 12.600

Descrive il sistema delle informazioni statistiche sul turismo ed espone, in un quadro organico, statistiche, dati ed indicatori aventi per oggetto i suoi principali aspetti, con riferimento agli anni 1985 e 1986.

STATISTICHE DELLA NAVIGAZIONE MARITTIMA, anno 1985 Vol. 40, L. 22.000

Contiene i dati statistici sul movimento dei natanti e del relativo carico avvenuto nei porti marittimi e negli altri approdi autorizzati del territorio nazionale.

STATISTICA DEGLI INCIDENTI STRADALI - Vol. 35, edizione 1988 L. 15.800

La più completa ed aggiornata raccolta di dati su una materia di viva attualità.

STATISTICA ANNUALE DEL COMMERCIO CON L'ESTERO, anno 1986 Vol. 43

Tomo 1 - Dati generali e riassuntivi L. 38.000

Tomo 2 - Merci per Capitoli merceologici e Paesi

- Parte prima: da Cap. 1 a Cap. 24 L. 12.600

- Parte seconda: da Cap. 25 a Cap. 40 L. 15.800

- Parte terza: da Cap. 41 a Cap. 67 L. 18.900

- Parte quarta: da Cap. 68 a Cap. 83 L. 15.800

- Parte quinta: da Cap. 84 a Cap. 85 L. 22.000

- Parte sesta: da Cap. 86 a Cap. 99 L. 15.800

- Appendice: L. 8.400

Riporta, con riferimento all'anno 1986, i dati definitivi sull'andamento delle importazioni e delle esportazioni con l'analisi completa del movimento per merci e per Paesi. Nel tomo primo è riportata, tra l'altro, un'ampia documentazione sul movimento delle merci nei depositi doganali e sul commercio di transito.

STATISTICHE DEI BILANCI DELLE AMMINISTRAZIONI REGIONALI, PROVINCIALI E COMUNALI - Vol. XXVII, edizione 1982 L. 14.000

Esponde i dati relativi ai bilanci delle Amministrazioni, tenendo conto dell'aspetto contabile, funzionale ed amministrativo dei documenti contabili. Per le Amministrazioni provinciali e comunali è stata dedicata particolare attenzione ai dati riguardanti i servizi sociali, i settori d'intervento nel campo economico ed il personale.

STATISTICHE DEL LAVORO - Vol. 26, edizione 1986 L. 12.000

Organica ed aggiornata documentazione statistica su tutti i principali aspetti del mondo del lavoro, nell'anno 1984.

ANNUARIO DI CONTABILITÀ NAZIONALE - Vol. 15, edizione 1987 L. 17.000

Contiene i dati sulla struttura e sulla evoluzione delle principali grandezze del sistema economico italiano.

PUBBLICAZIONI PLURIENNALI ED OCCASIONALI

ELEZIONI DELLA CAMERA DEI DEPUTATI E DEL SENATO DELLA REPUBBLICA, 26 giugno 1983 - Dati sommari L. 4.000

ELEZIONI AMMINISTRATIVE, dal 1975 al 1980 L. 3.500

L'ATTREZZATURA ALBERGHIERA IN ITALIA al 1° gennaio 1973 L. 5.500

TAVOLE ATTUARIALI 1970-72 L. 8.000

IL VALORE DELLA LIRA DAL 1861 al 1982 L. 5.000

INDAGINE SULLA STRUTTURA DELLE AZIENDE AGRICOLE 1977 L. 6.500

2° RILEVAZIONE NAZIONALE SULLO STATO DELL'EDILIZIA SCOLASTICA - 8 novembre 1976

Vol. I - Dati provinciali: Tomo 1 - Totale L. 10.000

Tomo 2 - Comuni capoluoghi L. 10.000

Vol. II - Dati comunali: Tomo 1 - Italia settentrionale L. 13.000

Tomo 2 - Italia centrale, meridionale e insulare L. 13.000

STATISTICHE SOCIALI - Vol II, 1981 L. 7.000

INDAGINE SU ALCUNE SPECIE DI ALBERI DA FRUTTO (MELO, PERO, PESCO, ARANCIO) - 1982 L. 13.000

RILEVAZIONE STATISTICA DEGLI ACQUEDOTTI E DELL'APPROVIGIONAMENTO IDRICO IN ITALIA - Situazione al 31-12-1975 L. 8.000

STATISTICHE AMBIENTALI - Vol. I, 1984 L. 9.000

INDAGINE SULLE STRUTTURE ED I COMPORTAMENTI FAMILIARI L. 16.000

POPOLAZIONE RESIDENTE E PRESENTE DEI COMUNI - Censimenti dal 1861 al 1981 L. 14.000

L'ITALIA DEI CENSIMENTI L. 10.000

STATISTICHE SULLA PUBBLICA AMMINISTRAZIONE - II ed. L. 18.000

SOMMARIO DI STATISTICHE STORICHE - 1926-1985 L. 35.000

IMMAGINI DELLA SOCIETÀ ITALIANA - 1988 L. 30.000

MORTALITÀ PER CAUSA E UNITÀ SANITARIA LOCALE - Dati 1980-82 L. 35.000

ATLANTE STATISTICO ITALIANO 1988 L. 50.000

Censimenti

12° CENSIMENTO GENERALE DELLA POPOLAZIONE - 25 ottobre 1981

DATI SULLE CARATTERISTICHE STRUTTURALI DELLA POPOLAZIONE E DELLE ABITAZIONI - Campione al 2% dei fogli di famiglia - Dati provvisori L. 5.000

Vol. I - Primi risultati provinciali e comunali sulla popolazione e sulle abitazioni (dati provvisori) L. 6.500

- Vol. II - Dati sulle caratteristiche strutturali della popolazione e delle abitazioni:
 Tomo I - Fascicoli provinciali - Prezzi vari
 Tomo 2 - Fascicoli regionali - Prezzi vari
 Tomo 3 - Fascicolo nazionale - Italia L. 25.000
- Vol. III - Popolazione delle frazioni geografiche e delle località abitate dei comuni - Fascicoli regionali e nazionale - Prezzi vari
- Vol. IV - Atti del censimento - L. 26.500

POPOLAZIONE LEGALE DEI COMUNI L. 8.000

6° CENSIMENTO GENERALE DELL'INDUSTRIA, DEL COMMERCIO, DEI SERVIZI E DELL'ARTIGIANATO - 26 ottobre 1981

- Vol. I - Primi risultati sulle imprese e sulle unità locali - Dati provvisori
 Tomo I - Dati nazionali, regionali e provinciali (esaurito)
 Tomo 2 - Dati comunali (esaurito)
- Vol. II - Dati sulle caratteristiche strutturali delle imprese e delle unità locali
 Tomo I - Fascicoli provinciali - Prezzi vari
 Tomo 2 - Fascicoli regionali - Prezzi vari
 Tomo 3 - Fascicolo nazionale - Italia L. 14.000
- Vol. III - Atti del censimento - L. 11.000
- Vol. IV - Relazione generale sul censimento - L. 26.500

3° CENSIMENTO GENERALE DELL'AGRICOLTURA - 24 ottobre 1982

CARATTERISTICHE STRUTTURALI DELLE AZIENDE AGRICOLE - L. 14.000

- Vol. I - Primi risultati provinciali e comunali - Dati provvisori L. 8.000
- Vol. II - Caratteristiche strutturali delle aziende agricole:
 Tomo I: Fascicoli provinciali - Prezzi vari
 Tomo 2: Fascicoli regionali - Prezzi vari
 Tomo 3: Fascicolo nazionale - Italia L. 11.000

Vol. III - Atti del censimento L. 33.500

TIPOLOGIA DELLE AZIENDE AGRICOLE - Campione al 10% dei questionari d'azienda L. 6.000

INDAGINE SULLE SUPERFICI A VITE

- Vol. I - Caratteristiche delle aziende con vite
 Tomo I: Dati provinciali, regionali e nazionali L. 33.500
 Tomo 2: Dati comunali L. 15.000
- Vol. II - Caratteristiche dei vitigni L. 33.500

Collana d'informazione

Anno 1988

- n. 15 - RILEVAZIONE DELLE FORZE DI LAVORO - Gennaio 1988 L. 8.400
- n. 16 - I CONSUMI DELLE FAMIGLIE NEL 1986 L. 22.000
- n. 17 - GLI IMPIEGHI DELL'ENERGIA IN ITALIA NEL 1982 - L. 15.800
- n. 18 - VALORE AGGIUNTO DELL'AGRICOLTURA, PER REGIONI - Anni 1980-87 L. 8.400
- n. 19 - CONTI ECONOMICI NAZIONALI, 1983-1987 L. 8.400
- n. 20 - STATISTICHE FORESTALI, DELLA CACCIA, PESCA E COOPERAZIONE - Anni 1986-87 L. 8.400
- n. 21 - INDAGINE STATISTICA SUI TRATTAMENTI PENSIONISTICI AL 31-12-86 L. 8.400
- n. 22 - RILEVAZIONE DELLE FORZE DI LAVORO - Aprile 1988 L. 8.400
- n. 23 - CONTI ECONOMICI NAZIONALI, OCCUPAZIONE E REDDITI DA LAVORO DIPENDENTE - Anni 1980-87 L. 12.600
- n. 24 - STATISTICHE DELL'ISTRUZIONE - DATI SOMMARI ANNO SCOLASTICO 1987-88 L. 12.600
- n. 25 - LAVORO E RETRIBUZIONI - Anno 1987 L. 8.400
- n. 26 - FATTURATO, PRODOTTO LORDO, INVESTIMENTI DELLE IMPRESE INDUSTRIALI, DEL COMMERCIO, DEI TRASPORTI E COMUNICAZIONI E DI ALCUNI TIPI DI SERVIZI - Anno 1985 L. 15.800
- n. 27 - FATTURATO, PRODOTTO LORDO, INVESTIMENTI DELLE IMPRESE INDUSTRIALI, DEL COMMERCIO, DEI TRASPORTI E COMUNICAZIONI E DI ALCUNI TIPI DI SERVIZI - Anno 1986 L. 15.800

Anno 1989

- n. 1 - STATISTICHE DELLA RICERCA SCIENTIFICA - Consuntivo 1986 - Previsione 1987 e 1988 L. 10.000
- n. 2 - RILEVAZIONE DELLE FORZE DI LAVORO - Luglio 1988 L. 10.000
- n. 3 - LA DISTRIBUZIONE QUANTITATIVA DEL REDDITO IN ITALIA NELLE INDAGINI SUI BILANCI DI FAMIGLIA - Anni 1986-87 L. 10.000
- n. 4 - ASPETTI STRUTTURALI DELLE PRINCIPALI COLTIVAZIONI LEGNOSE AGRARIE - Anno 1987 L. 10.000

- n. 5 - LA SUPERFICIE FORESTALE NELLE COMUNITÀ MONTANE AL 31 DICEMBRE 1986 L. 10.000
- n. 6 - STRUTTURA DELLE AZIENDE AGRICOLE -1985 L. 38.000
- n. 7 - VALORE AGGIUNTO DELL'AGRICOLTURA PER REGIONE - Anni 1970-1980 L. 14.000
- n. 8 - CONTI ECONOMICI DELLE AMMINISTRAZIONI PUBBLICHE E DELLA PROTEZIONE SOCIALE - 1980-87 L. 14.000
- n. 9 - BILANCI CONSUNTIVI DELLE REGIONI E DELLE PROVINCE AUTONOME - Anno 1986 (in corso di stampa)

Note e relazioni

Anno 1987

- n. 1 - TAVOLE DI MORTALITÀ DELLA POPOLAZIONE ITALIANA PER REGIONE 1979-83 L. 6.300
- n. 2 - LA DESTAGIONALIZZAZIONE DEGLI INDICI DELLA PRODUZIONE INDUSTRIALE - Anni 1977-1985 L. 6.300
- n. 3 - TAVOLE DI MORTALITÀ E TAVOLE ATTUARIALI DELLA POPOLAZIONE ITALIANA NEL 1981 L. 15.000
- n. 4 - LA CRIMINALITÀ ATTRAVERSO LE STATISTICHE - Anni 1981-1982-1983 L. 6.300

Anno 1988

- n. 1 - LA CRIMINALITÀ ATTRAVERSO LE STATISTICHE - Anni 1984-1985-1986 L. 8.400
- n. 2 - INDAGINE SUGLI SPORT E SULLE VACANZE - Le vacanze degli italiani nel 1985 L. 15.800
- n. 3 - INDAGINE SUGLI SPORT E SULLE VACANZE - Gli sport degli italiani nel 1985 L. 8.400

Metodi e norme

Serie A

- n. 18 - NUMERI INDICI DEL COSTO DI COSTRUZIONE DI UN FABBRICATO RESIDENZIALE: Base 1976 = 100 L. 1.500
- n. 20 - NUMERI INDICI DEI PREZZI: Base 1980 = 100 L. 4.500
- n. 21 - NUMERI INDICI DEI PREZZI DEI PRODOTTI VENDUTI E DEI BENI ACQUISTATI DAGLI AGRICOLTORI: Base 1980 = 100 L. 5.000
- n. 22 - NUMERI INDICI DELLA PRODUZIONE INDUSTRIALE: Base 1980 = 100 L. 9.000
- n. 23 - NUMERI INDICI DEI PREZZI AL CONSUMO: Base 1985 = 100 L. 6.300
- n. 24 - NUMERI INDICI DEL FATTURATO, DEGLI ORDINATIVI E DELLA CONSISTENZA DEGLI ORDINATIVI: BASE 1980 = 100 L. 6.300

Serie B

- n. 18 - ANAGRAFE DELLA POPOLAZIONE L. 4.500
- n. 20 - NORME TECNICHE PER LE RILEVAZIONI STATISTICHE DEI PREZZI ALL'INGROSSO DELLE MERCI L. 2.000
- n. 21 - ISTRUZIONI PER LA RILEVAZIONE STATISTICA DEL MOVIMENTO DELLA POPOLAZIONE L. 4.000
- n. 22 - ISTRUZIONI PER LA RILEVAZIONE DEI DATI DELLE STATISTICHE FORESTALI L. 6.000
- n. 23 - ISTRUZIONI PER LA RILEVAZIONE DELL'ATTIVITÀ EDILIZIA L. 8.400
- n. 24 - ISTRUZIONI PER LE RILEVAZIONI DELLE STATISTICHE GIUDIZIARIE
Tomo 1 - Procedura di rilevazione L. 15.800
Tomo 2 - Modelli di rilevazione - L. 15.800

Serie C

- n. 8 - CLASSIFICAZIONE DELLE ATTIVITÀ ECONOMICHE L. 6.500
- n. 9 - CLASSIFICAZIONE DELLE PROFESSIONI L. 6.500
- n. 10 - CLASSIFICAZIONI DELLE MALATTIE, TRAUMATISMI E CAUSE DI MORTE - Ristampa 1986
Vol. 1: Introduzione e parte sistematica L. 16.000
Vol. 2: Indici alfabetici L. 25.000

Annali di statistica

Serie IX

- Vol. 1 - ATTI DEL 2° CONVEGNO SULL'INFORMAZIONE STATISTICA IN ITALIA (Roma, 17-19 giugno 1981) L. 10.000
- Vol. 2 - PREVISIONI DELLA POPOLAZIONE RESIDENTE DAL 1986 AL 2001 PER SESSO, ETÀ E REGIONE L. 14.000
- Vol. 3 - STUDI STATISTICI SUI CONSUMI - Edizione 1983 L. 9.500
- Vol. 4 - CONTABILITÀ NAZIONALE - FONTI E METODI L. 9.000
- Vol. 5 - ATTI DEL SEMINARIO SULLA VALUTAZIONE DEI RISULTATI E DELLA METODOLOGIA DEI CENSIMENTI (Roma, 7-11 maggio 1984) L. 25.000
- Vol. 6 - ATTI DEL CONVEGNO «LA FAMIGLIA IN ITALIA» (Roma, 29-30 ottobre 1985) L. 14.000
- Vol. 7 - ATTI DEL CONVEGNO SULL'INFORMAZIONE STATISTICA E I PROCESSI DECISIONALI (Roma, 11-12 dicembre 1986) L. 15.000

