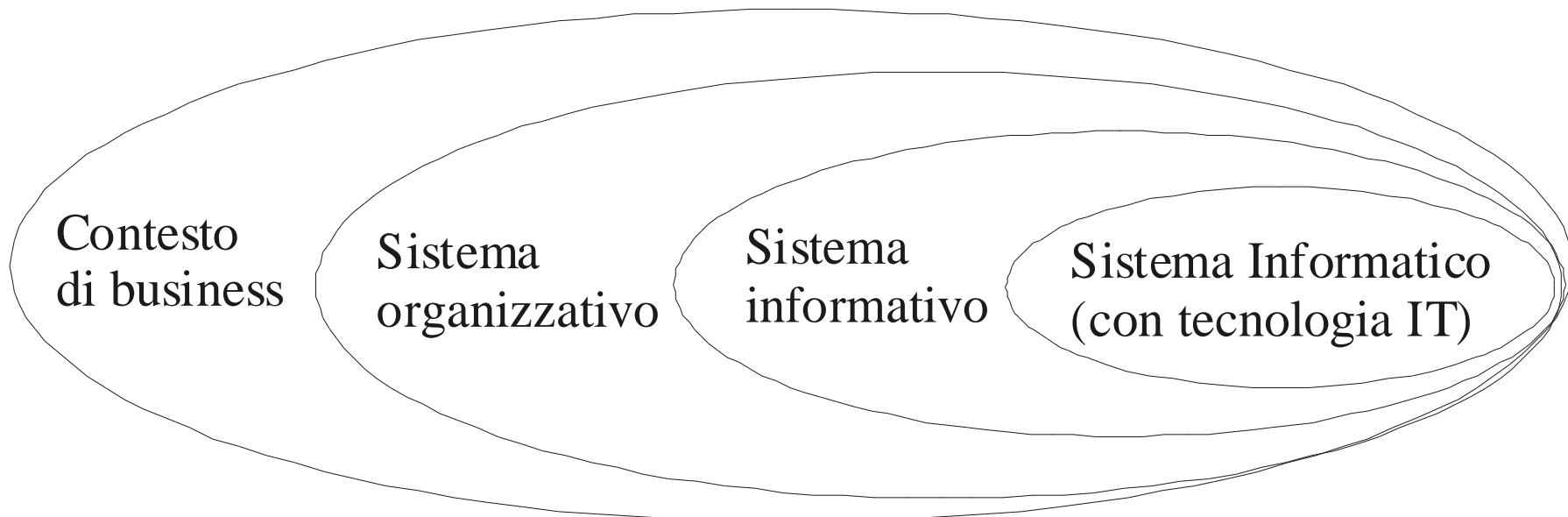


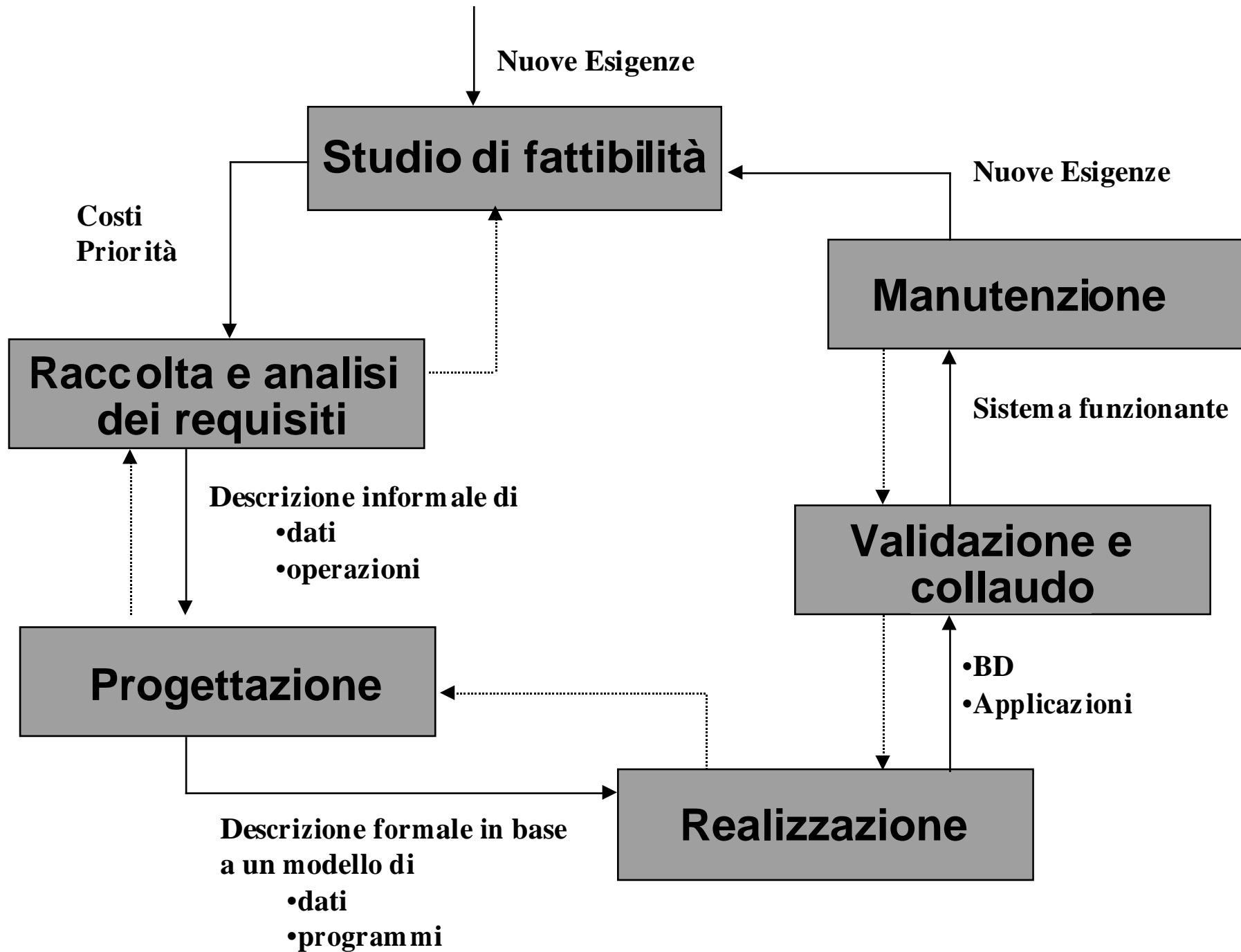
PARTE III

Progettazione di basi di dati



Progettazione di basi di dati

- È una delle attività del processo di sviluppo dei sistemi informativi
- va quindi inquadrata in un contesto più generale:
- il ciclo di vita dei sistemi informativi:
 - Insieme e sequenzializzazione delle attività svolte da analisti, progettisti, utenti, nello sviluppo e nell'uso dei sistemi informativi
 - attività iterativa, quindi *ciclo*



- Per garantire prodotti di buona qualità è opportuno seguire una
 - *metodologia di progetto*, con:
 - articolazione delle attività in fasi
 - criteri di scelta
 - modelli di rappresentazione
 - generalità e facilità d'uso



Raccolta e analisi dei requisiti

- **Raccolta requisiti: scopo**
 - Qual'è il problema?
 - Caratteristiche del sistema
 - statiche (dati)
 - dinamiche (operazioni)
- **Analisi dei requisiti: scopo**
 - Chiarire
 - Organizzare
 - Ristrutturare

Raccolta e analisi dei requisiti

- Attività difficile e non standardizzabile
- Attività iterativa
 - l'analisi inizia con i primi requisiti raccolti e spesso indirizza verso altre acquisizioni

Raccolta dei Requisiti

- Possibili fonti:
 - *utenti*, attraverso:
 - interviste
 - documentazione apposita
 - *documentazione esistente*:
 - normative (leggi, regolamenti di settore)
 - regolamenti interni, procedure aziendali
 - sistemi esistenti
 - *modulistica*
- *Risultato finale: relazione scritta*

Acquisizione per interviste

- **utenti diversi possono fornire informazioni diverse**
- **utenti a livello più alto hanno spesso una visione più ampia ma meno dettagliata**
- **le interviste portano spesso ad una acquisizione dei requisiti “per raffinamenti successivi”**

Interazione con gli utenti

- **Spunti:**
 - effettuare spesso verifiche di comprensione e coerenza
 - verificare anche per mezzo di esempi (generali e relativi a casi limite)
 - richiedere definizioni e classificazioni
 - far evidenziare gli aspetti essenziali rispetto a quelli marginali

Analisi dei Requisiti

- **Scopo: Formalizzare/Strutturare il testo**
- **Regole generali:**
 - **scelta e standardizzazione termini**
 - **scelta e standardizzazione frasi**
 - **raggruppamento requisiti per descrizioni omogenee per**
 - **dati**
 - **operazioni**

Requisiti: organizzazione di termini e concetti

- **Scelta e standardizzazione dei termini**
 - scegliere il corretto livello di astrazione
 - costruire un glossario dei termini
 - individuare omonimi e sinonimi e unificare i termini
 - rendere esplicito il riferimento fra termini

Requisiti, un esempio

Base di dati bibliografica

**Si vogliono organizzare i dati di interesse
per automatizzare la gestione dei
riferimenti bibliografici**

Base di dati bibliografica

Si vogliono organizzare i dati di interesse per automatizzare la gestione dei riferimenti bibliografici, con tutte le informazioni da riportarsi in una bibliografia.

Per ogni pubblicazione deve esistere un codice identificante costituito da sette caratteri, indicanti le iniziali degli autori, l'anno di pubblicazione e un carattere aggiuntivo per la discriminazione delle collisioni.

Base di dati bibliografica

Si vogliono organizzare i dati di interesse per automatizzare la gestione dei riferimenti bibliografici, con tutte le informazioni da riportarsi in una bibliografia. Le pubblicazioni sono di due tipi, monografie (per le quali interessano editore, data e luogo di pubblicazione) e articoli su rivista (con nome della rivista, volume, numero, pagine e anno di pubblicazione); per entrambi i tipi si debbono ovviamente riportare i nomi degli autori. Per ogni pubblicazione deve esistere un codice identificante

Un esempio più articolato

Società di formazione (1)

Si vuole realizzare una base di dati per una società che eroga corsi, di cui vogliamo rappresentare i dati dei partecipanti ai corsi e dei docenti. Per gli studenti (circa 5000), identificati da un codice, si vuole memorizzare il codice fiscale, il cognome, l'età, il sesso, il luogo di nascita, il nome dei loro attuali datori di lavoro, i posti dove hanno lavorato in precedenza insieme al periodo, l'indirizzo e il numero di telefono, i corsi che hanno frequentato (i corsi sono in tutto circa 200) e il giudizio finale.

Società di formazione (2)

Rappresentiamo anche i seminari che stanno attualmente frequentando e, per ogni giorno, i luoghi e le ore dove sono tenute le lezioni. I corsi hanno un codice, un titolo e possono avere varie edizioni con date di inizio e fine e numero di partecipanti. Se gli studenti sono liberi professionisti, vogliamo conoscere l'area di interesse e, se lo possiedono, il titolo. Per quelli che lavorano alle dipendenze di altri, vogliamo conoscere invece il loro livello e la posizione ricoperta.

Società di formazione (3)

Per gli insegnanti (circa 300), rappresentiamo il cognome, l'età, il posto dove sono nati, il nome del corso che insegnano, quelli che hanno insegnato nel passato e quelli che possono insegnare. Rappresentiamo anche tutti i loro recapiti telefonici. I docenti possono essere dipendenti interni della società o collaboratori esterni.

Glossario dei termini

Termine	Descrizione	Sinonimi	Collegamenti
Partecipante	Persona che partecipa ai corsi	Studente	Corso, Società
Docente	Docente dei corsi. Può essere esterno	Insegnante	Corso
Corso	Corso organizzato dalla società. Può avere più edizioni.	Seminario	Docente
Società	Ente presso cui i partecipanti lavorano o hanno lavorato	Posti	Partecipante

Requisiti: organizzazione di termini e concetti

- **Scelta e standardizzazione delle frasi**
 - riorganizzare le frasi per concetti
 - standardizzare la struttura delle frasi
 - suddividere le frasi articolate
 - separare le frasi sui dati da quelle sulle funzioni

Frasi di carattere generale

**Si vuole realizzare una base di dati per
una società che eroga corsi, di cui
vogliamo rappresentare i dati dei
partecipanti ai corsi e dei docenti.**

Frasi relative ai partecipanti

Per i partecipanti (circa 5000), identificati da un codice, rappresentiamo il codice fiscale, il cognome, l'età, il sesso, la città di nascita, i nomi dei loro attuali datori di lavoro e di quelli precedenti (insieme alle date di inizio e fine rapporto), le edizioni dei corsi che stanno attualmente frequentando e quelli che hanno frequentato nel passato, con la relativa votazione finale in decimi.

Frasi relative ai datori di lavoro

Relativamente ai datori di lavoro presenti e passati dei partecipanti, rappresentiamo il nome, l'indirizzo e il numero di telefono.

Frasi relative ai corsi

Per i corsi (circa 200), rappresentiamo il titolo e il codice, le varie edizioni con date di inizio e fine e, per ogni edizione, rappresentiamo il numero di partecipanti e il giorno della settimana, le aule e le ore dove sono tenute le lezioni.

Frasi relative a tipi specifici di partecipanti

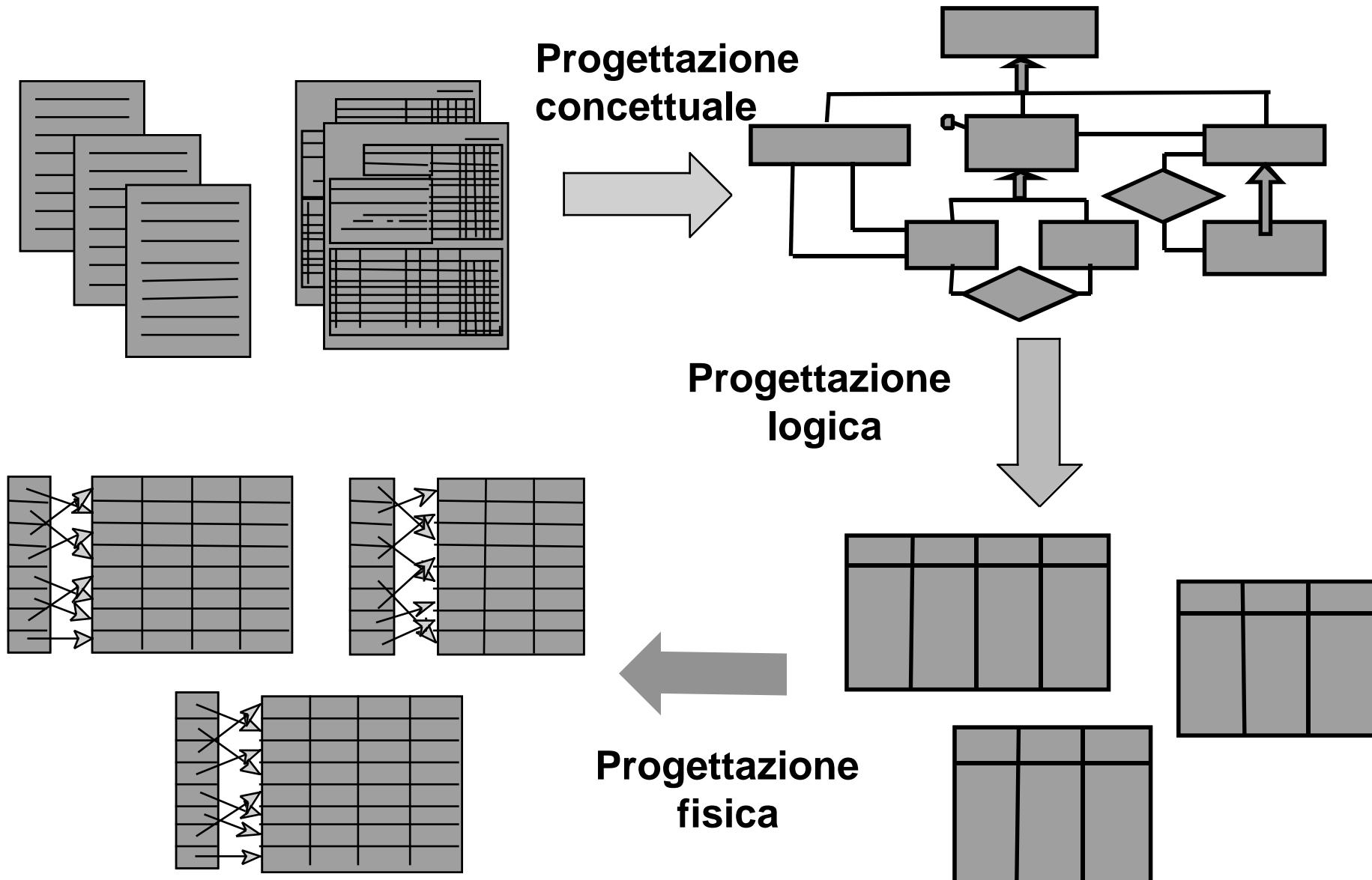
Per i partecipanti che sono liberi professionisti, rappresentiamo l'area di interesse e, se lo possiedono, il titolo professionale. Per i partecipanti che sono dipendenti, rappresentiamo invece il loro livello e la posizione ricoperta.

Frasi relative ai docenti

Per i docenti (circa 300), rappresentiamo il cognome, l'età, la città di nascita, tutti i numeri di telefono, il titolo del corso che insegnano, di quelli che hanno insegnato nel passato e di quelli che possono insegnare. I docenti possono essere dipendenti interni della società di formazione o collaboratori esterni.



Progettazione

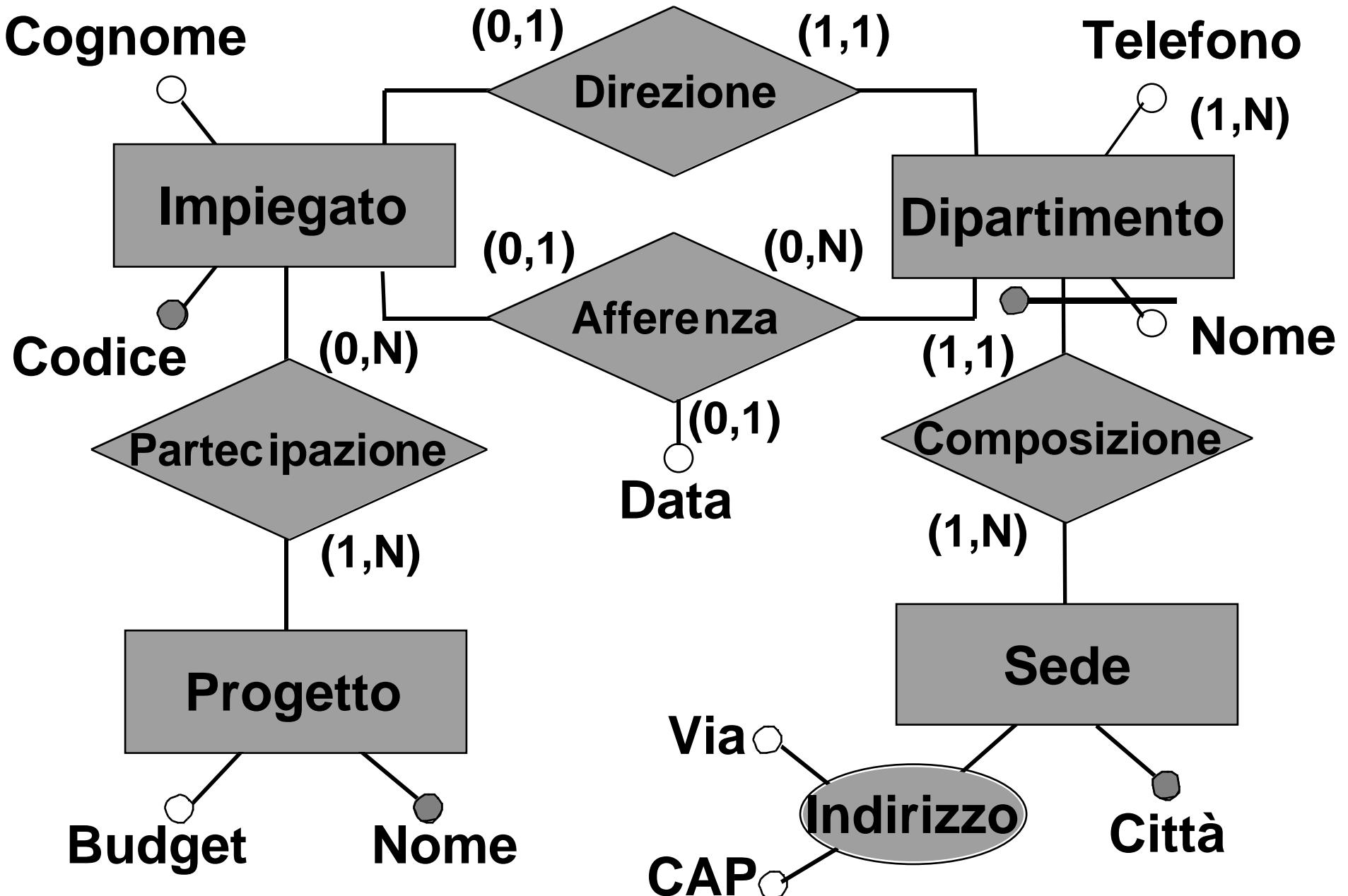


Progettazione concettuale

- **Formalizzazione dei requisiti in un modello concettuale (E-R)**
- **Vantaggi**
 - **descrizione astratta (che cosa, prima che come)**
 - **classi di dati di interesse e le loro correlazioni**
 - **rappresentazioni grafiche utili per documentazione e comunicazione**

Documentazione associata agli schemi concettuali

- **dizionario dei dati**
 - entità
 - relationship
- **vincoli non esprimibili**



Dizionario dei dati (entità)

Entità	Descrizione	Attributi	Identificatore
Impiegato	Dipendente dell'azienda	Codice, Cognome, Stipendio	Codice
Progetto	Progetti aziendali	Nome, Budget	Nome
Dipartimento	Struttura aziendale	Nome, Telefono	Nome, Sede
Sede	Sede dell'azienda	Città, Indirizzo	Città

Dizionario dei dati (relationship)

Relazioni	Descrizione	Componenti	Attributi
Direzione	Direzione di un dipartimento	Impiegato, Dipartimento	
Afferenza	Afferenza a un dipartimento	Impiegato, Dipartimento	Data
Partecipazione	Partecipazione a un progetto	Impiegato, Progetto	
Composizione	Composizione dell'azienda	Dipartimento, Sede	

Vincoli non esprimibili

Vincoli di integrità sui dati

- (1) Il direttore di un dipartimento deve afferire a tale dipartimento
- (2) Un impiegato non deve avere uno stipendio maggiore del direttore del dipartimento al quale afferisce
- (3) Un dipartimento con sede a Roma deve essere diretto da un impiegato con più di dieci anni di anzianità
- (4) Un impiegato che non afferisce a nessun dipartimento non deve partecipare a nessun progetto

Quale costrutto E-R va utilizzato per rappresentare un concetto presente nelle specifiche?

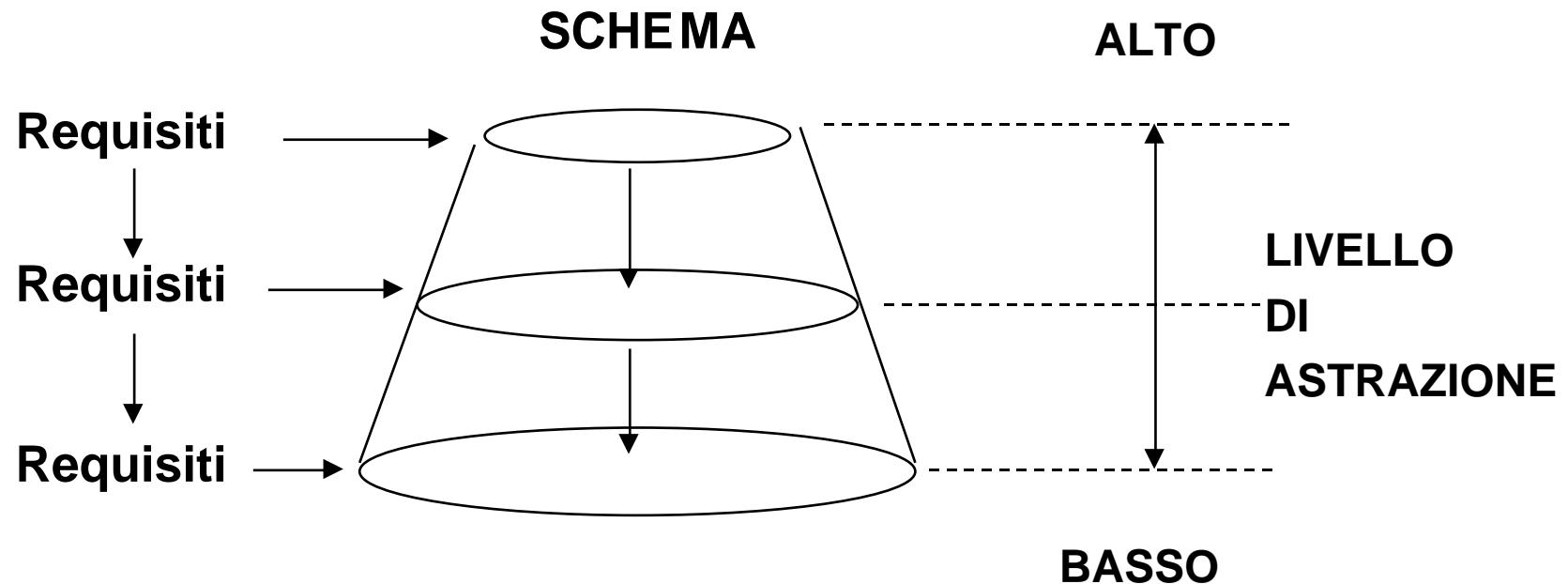
Bisogna basarsi sulle definizioni dei costrutti del modello E-R

- se ha proprietà significative e descrive oggetti con esistenza autonoma
 - **entità**
- se è semplice e non ha proprietà
 - **attributo**
- se correla due o più concetti
 - **relazione**
- se è caso particolare di un altro
 - **generalizzazione**

Strategie di progetto

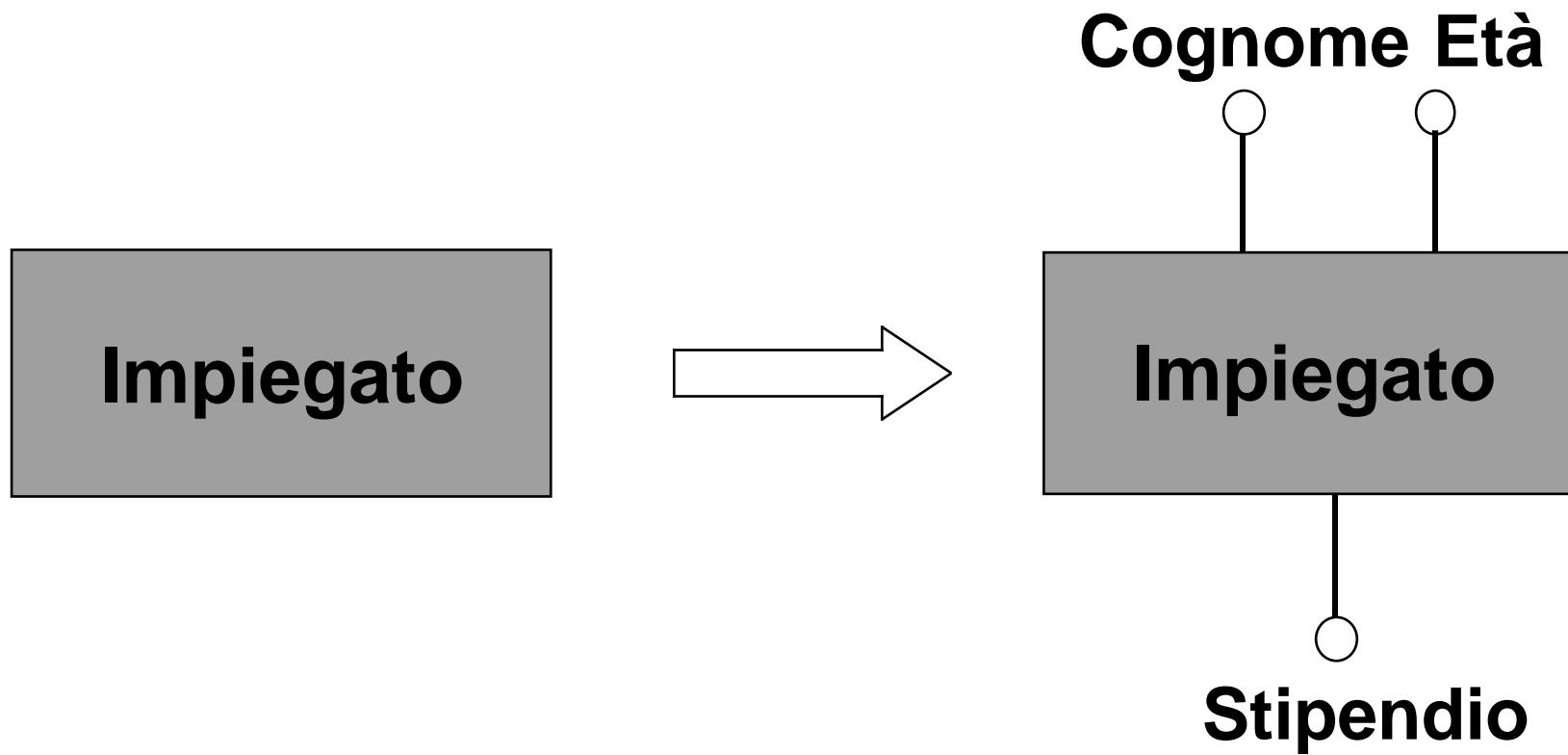
- **top-down**
- **bottom-up**
- **inside-out**

Strategia top-down

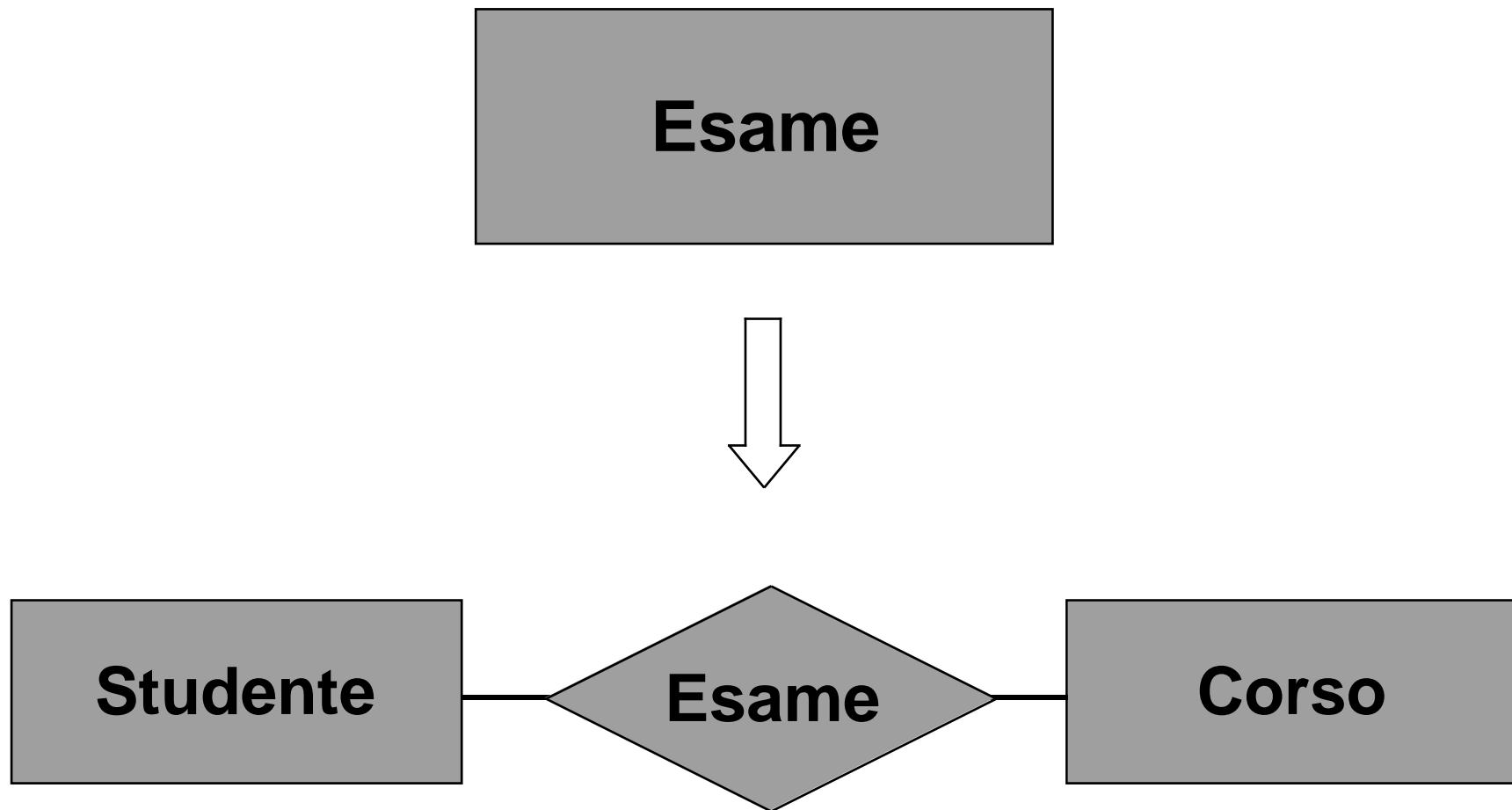


- **Vantaggi**
 - posso aggiungere dettagli pian piano
(controllo della complessità)
- **Svantaggi**
 - occorre una visione globale del sistema

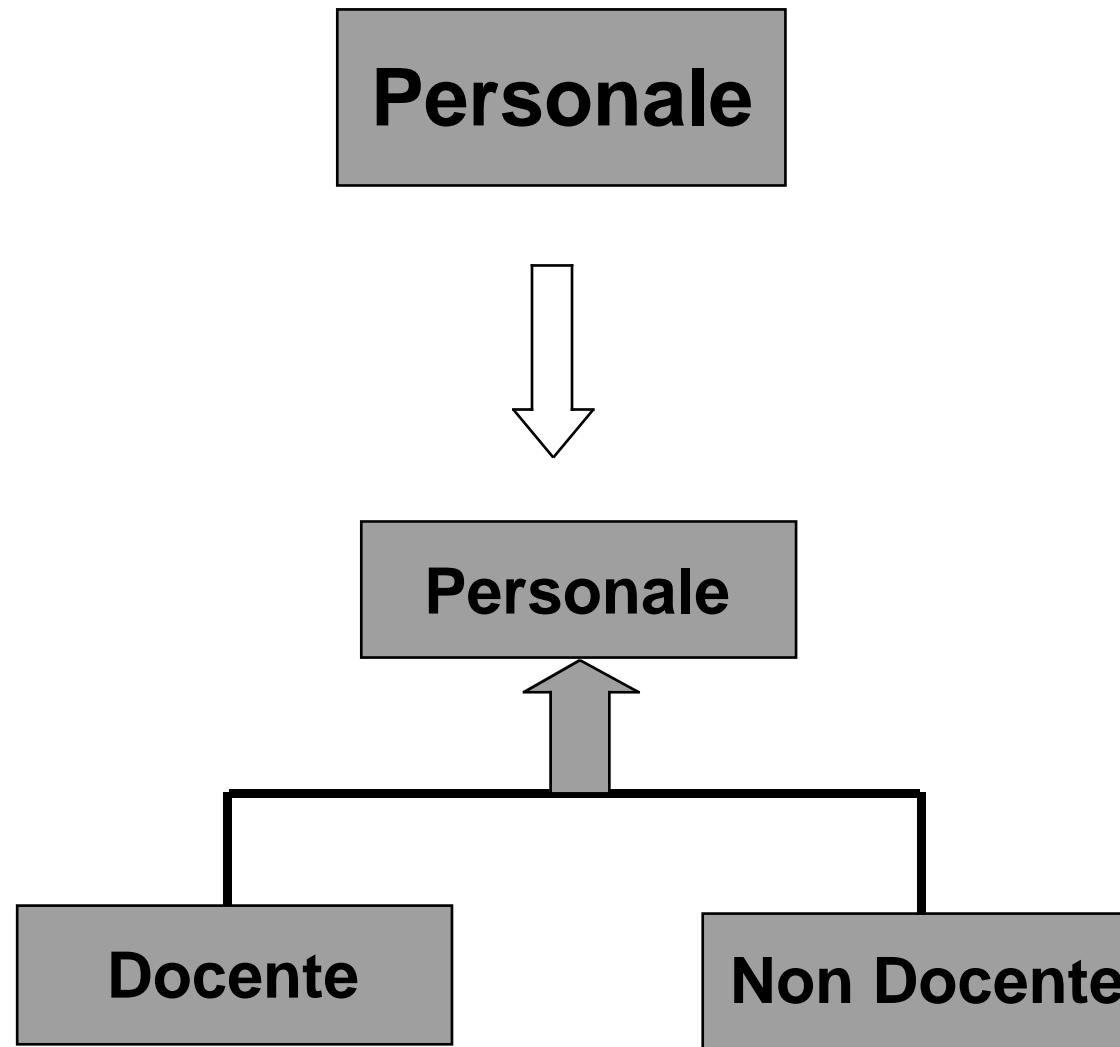
Primitive di raffinamento top-down



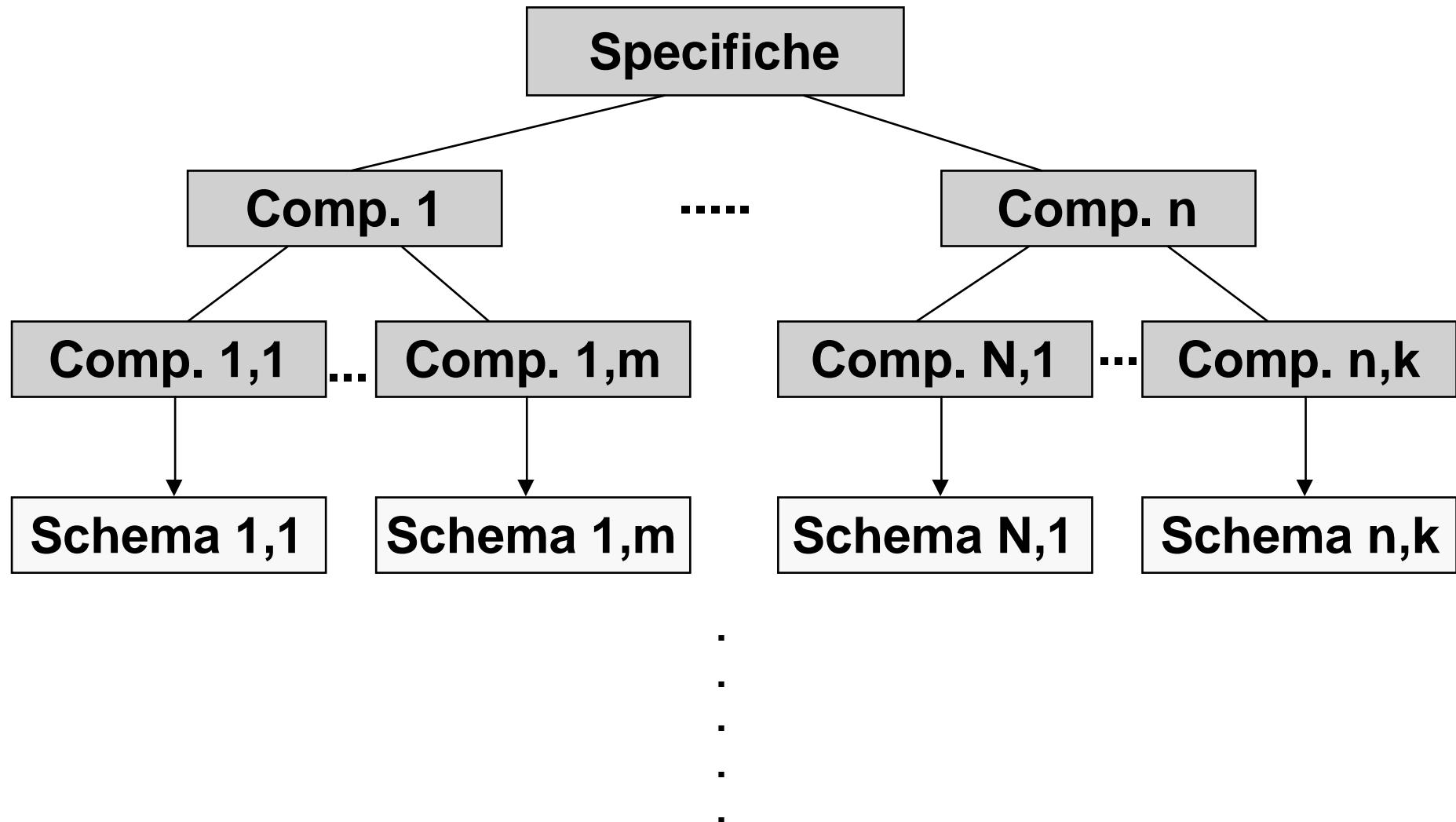
Primitive di raffinamento top-down



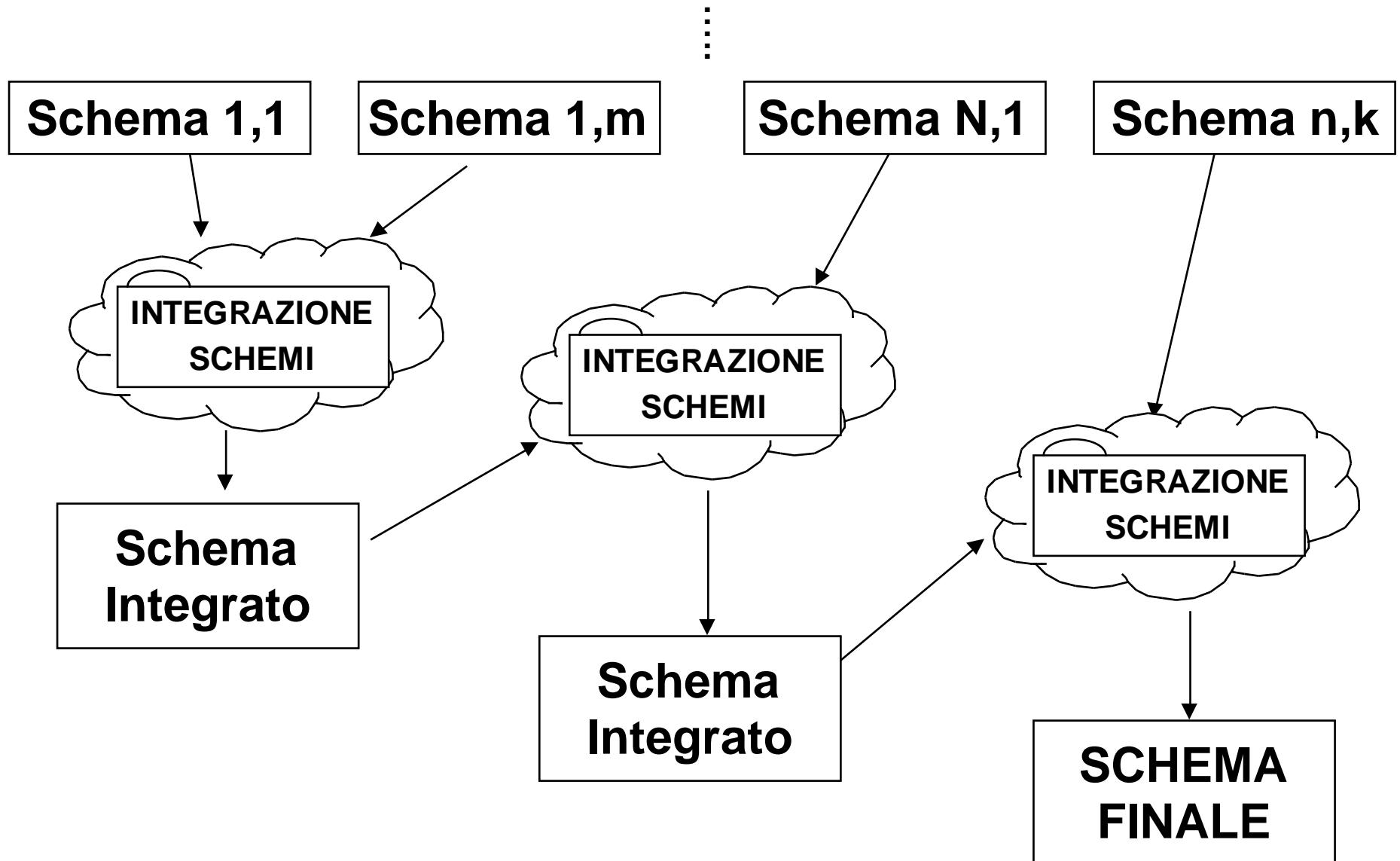
Primitive di raffinamento top-down



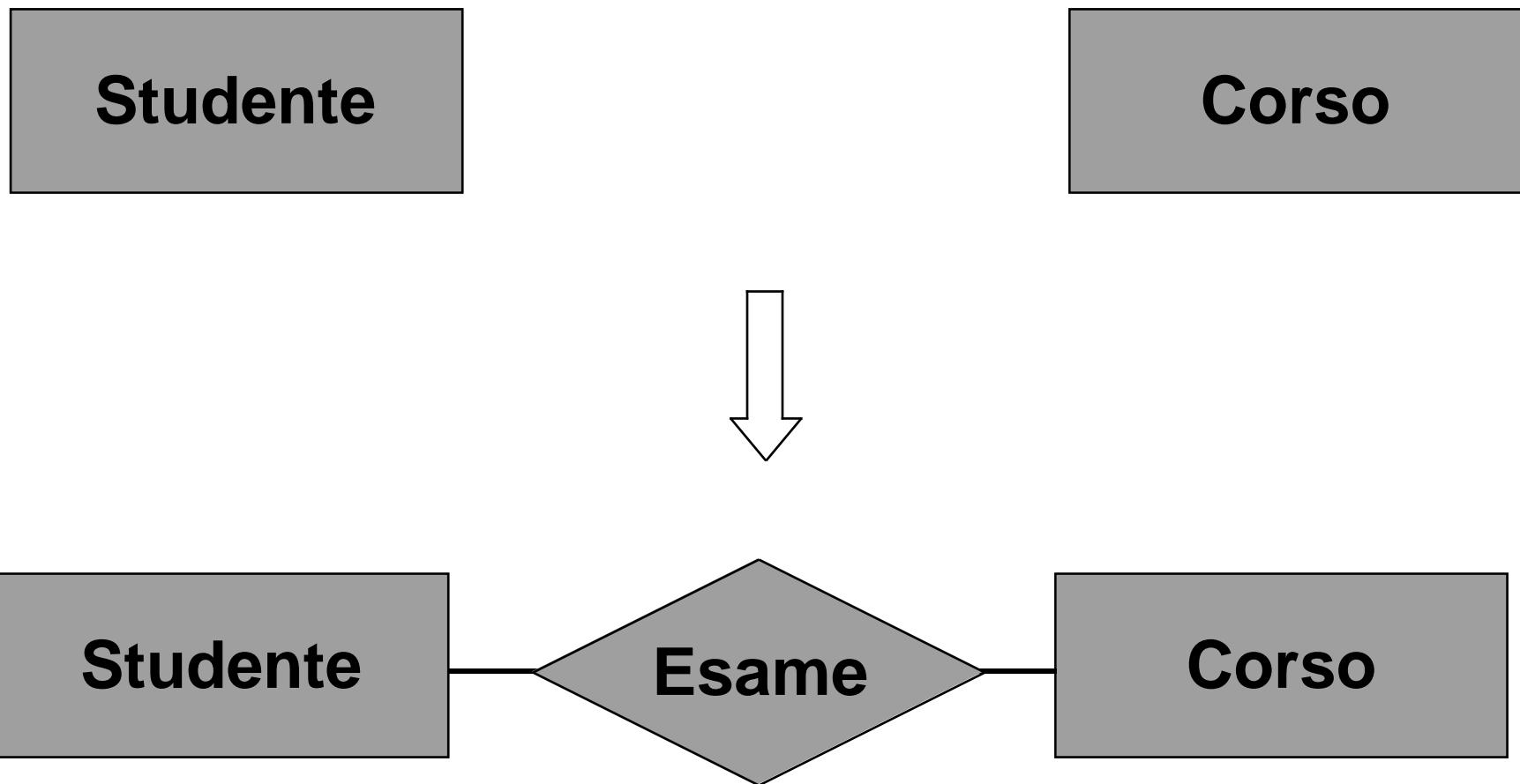
Strategia bottom-up



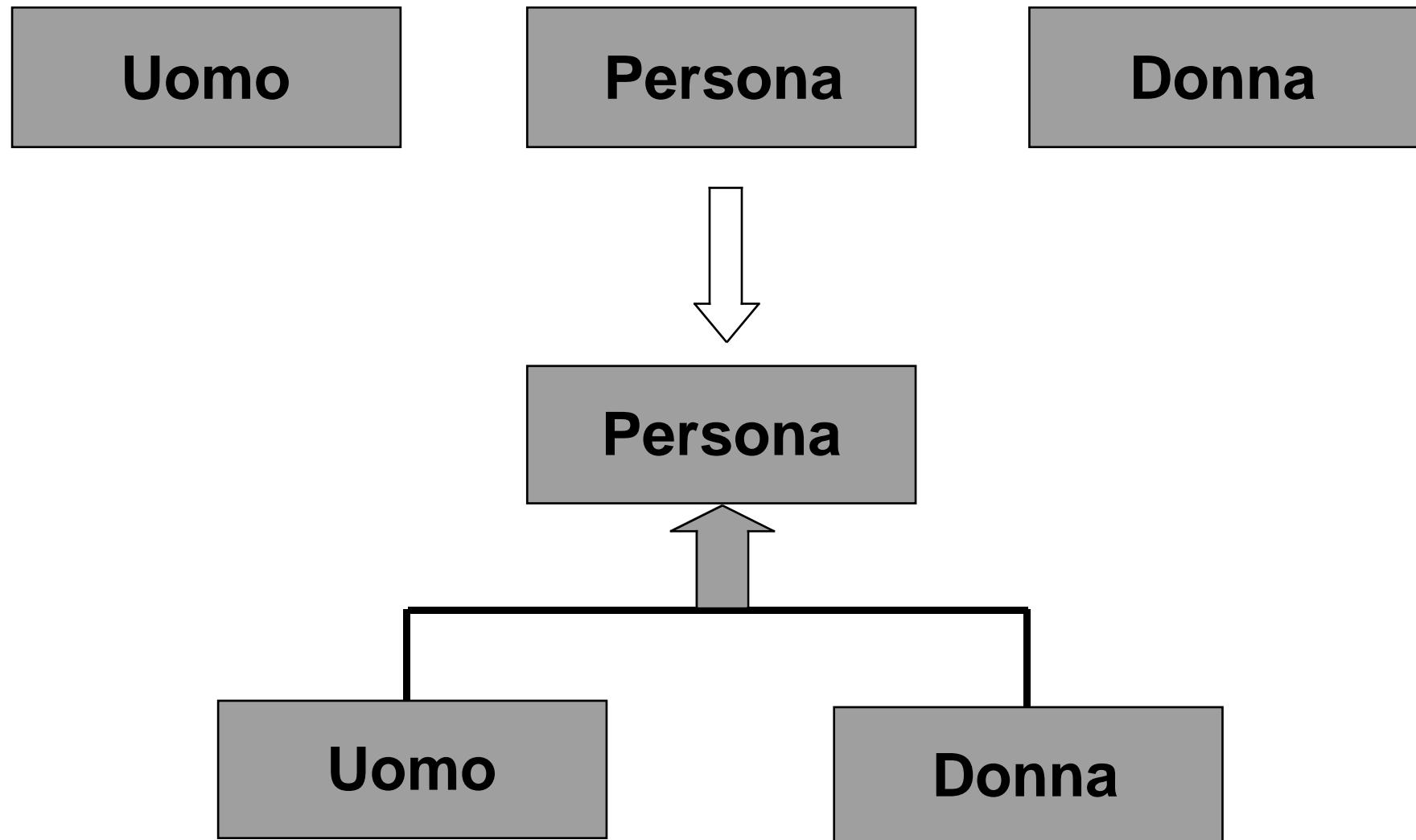
Strategia bottom-up (Cont.)



Primitive di raffinamento Bottom-up



Primitive di raffinamento Bottom-up



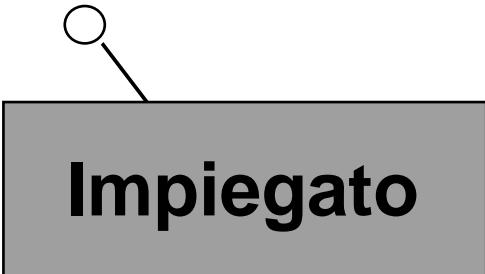
Strategia bottom-up

- **Vantaggi**
 - posso operare con conoscenza incompleta
 - ogni sottospecifica un progettista
- **Svantaggi**
 - integrazione di schemi

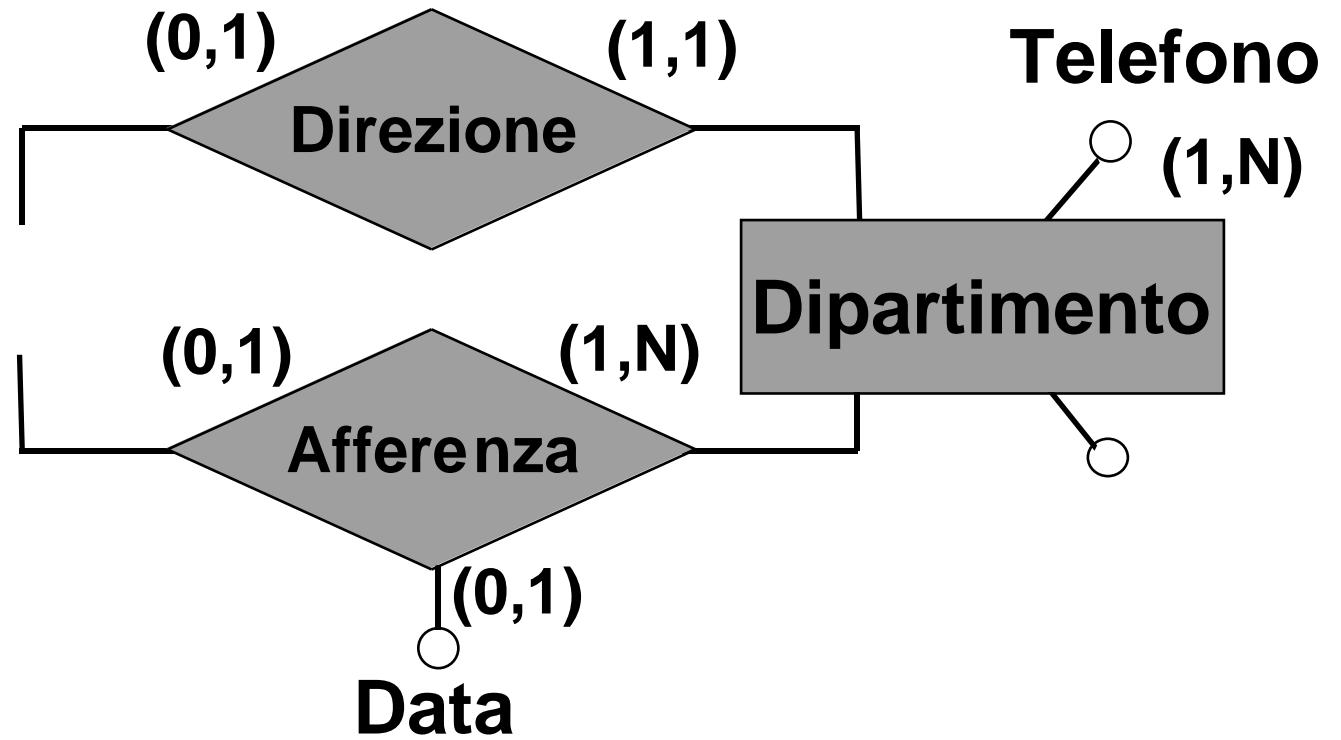
Strategia inside-out

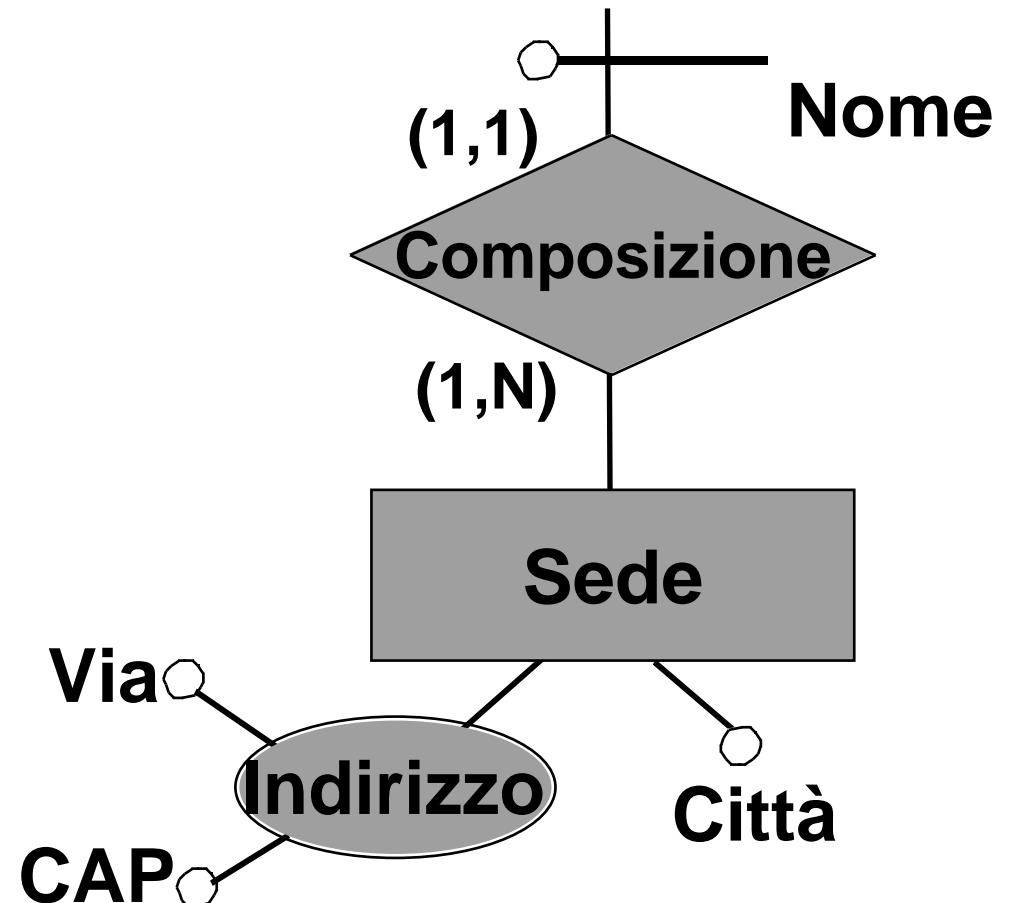
- **Si parte da un insieme di concetti tipicamente omogenei per area funzionale**
- **Si estende la costruzione verso le aree collegate**

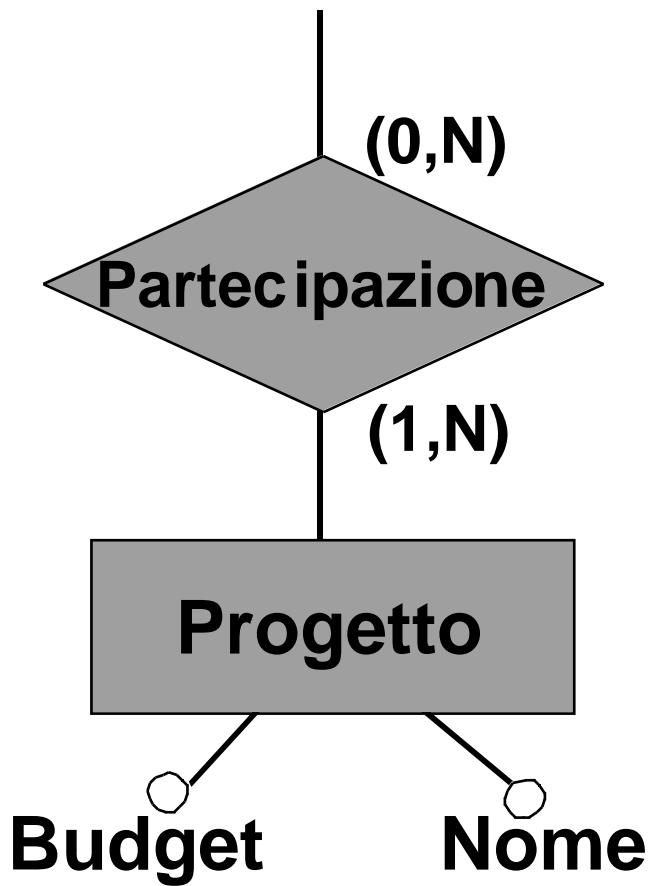
Cognome



Codice







In pratica

- si procede di solito con una strategia ibrida (*mista*):
 - si individuano i concetti principali e si realizza uno *schema scheletro*
 - sulla base di questo si può decomporre
 - poi si raffina, si espande, si integra

Definizione dello schema scheletro

- **Si individuano i concetti più importanti, ad esempio perché più citati o perché indicati esplicitamente come cruciali e li si organizza in un semplice schema concettuale**

Analisi di Qualità di uno Schema Concettuale

- **Correttezza**
 - sintattica
 - semantica
- **Completezza rispetto ai requisiti**
- **Leggibilità**
- **Minimalità**

Progettazione Logica

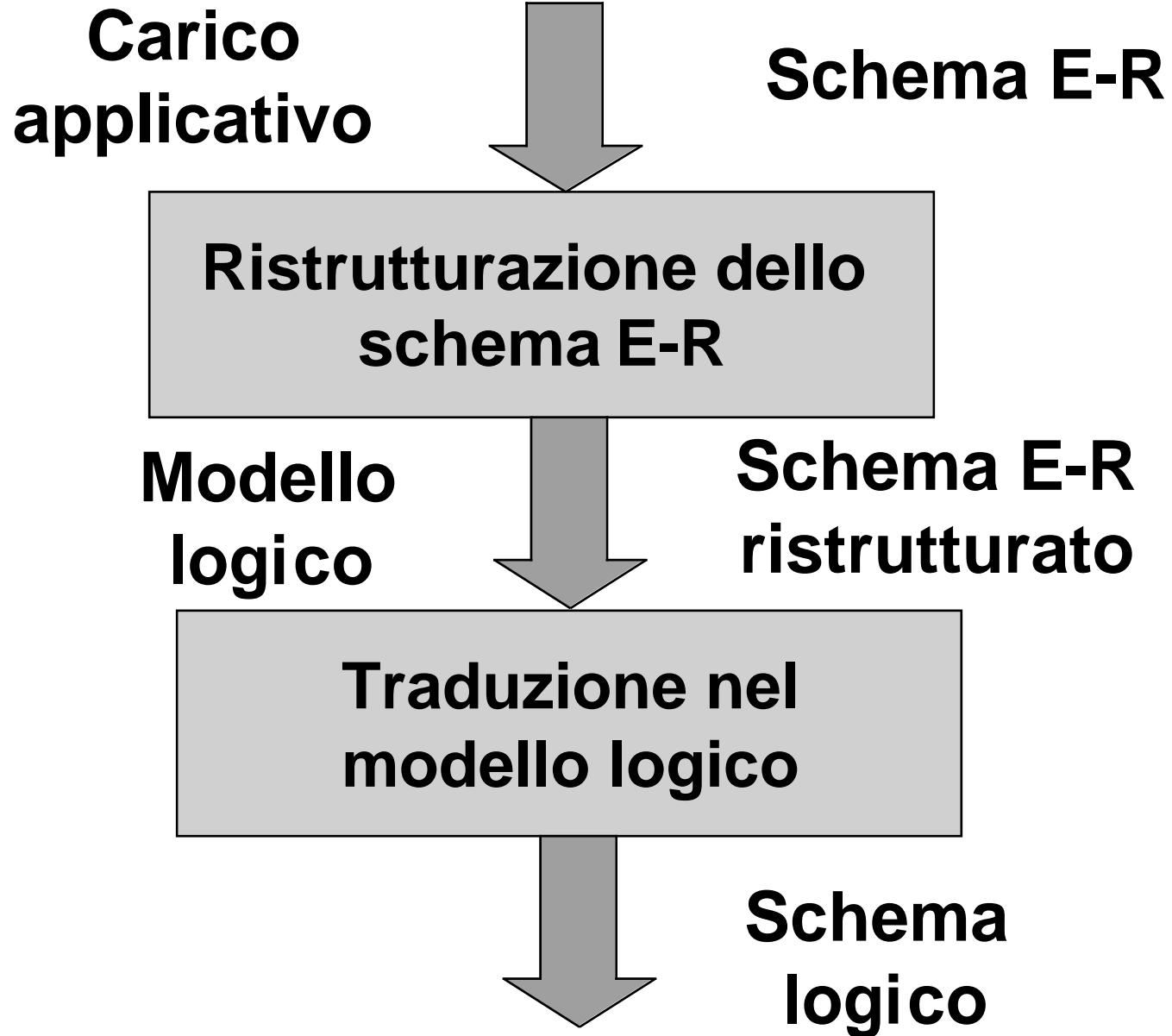
"tradurre" lo schema concettuale in uno schema logico che rappresenti gli stessi dati in maniera corretta ed efficiente

Dati di ingresso e uscita

- **Ingresso:**
 - **schema concettuale**
 - **informazioni sul carico applicativo**
 - **modello logico**
- **Uscita:**
 - **schema logico**
 - **documentazione associata**

Non si tratta di una pura e semplice traduzione

- alcuni aspetti non sono direttamente rappresentabili**
- è necessario considerare le prestazioni**



Ristrutturazione schema E-R

- **Motivazioni:**
 - semplificare la traduzione
 - "ottimizzare" le prestazioni
- **Osservazione:**
 - uno schema E-R ristrutturato non è (più) uno schema concettuale nel senso stretto del termine

**Per ottimizzare il risultato abbiamo
bisogno di analizzare le prestazioni a
questo livello**

- Ma:
 - le prestazioni non sono valutabili con precisione su uno schema concettuale!

Consideriamo “indicatori” dei parametri che regolano le prestazioni

- ***spazio***: numero di occorrenze previste
- ***tempo***: numero di occorrenze (di entità e relationship) visitate durante un’operazione

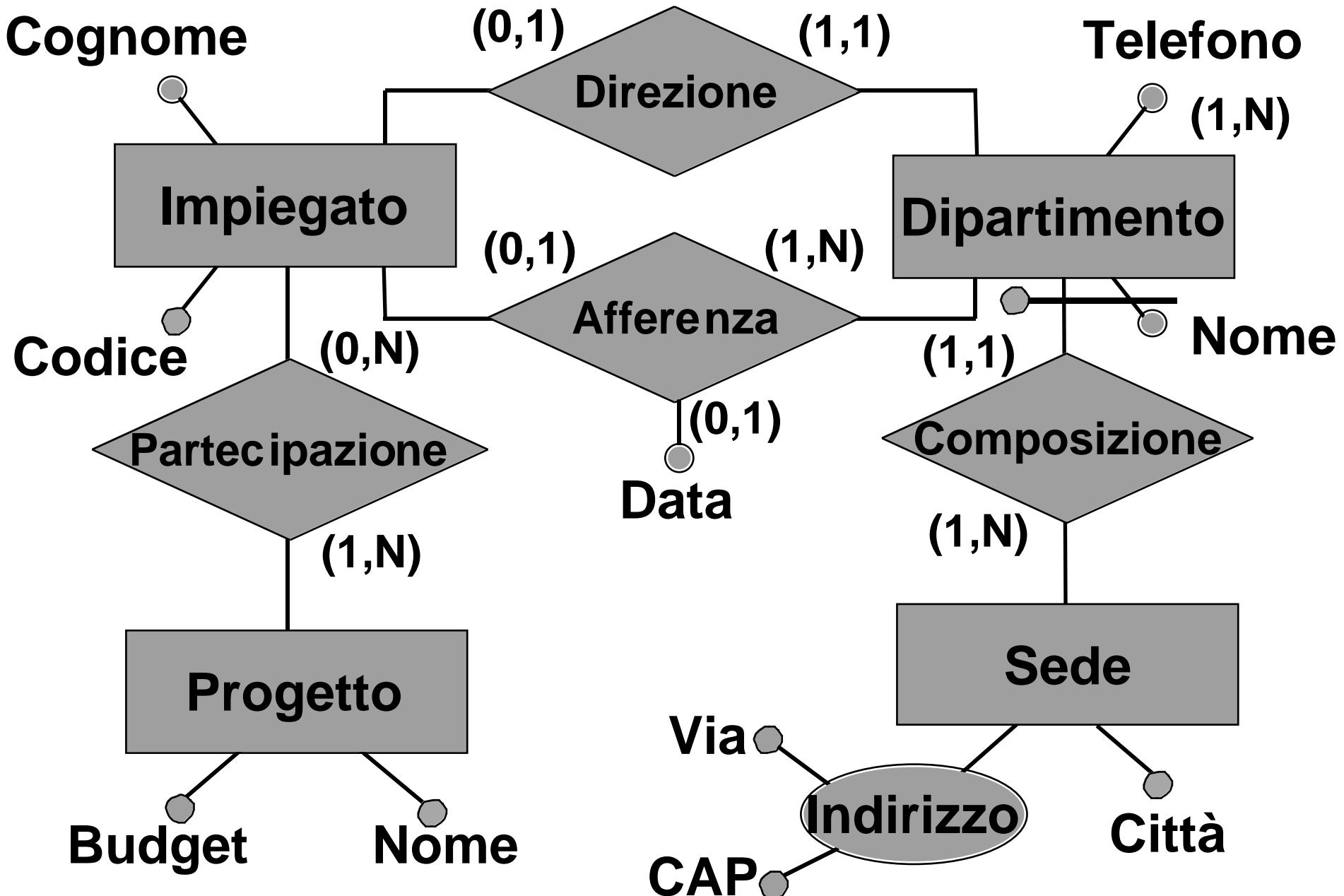


Tavola dei volumi

Concetto	Tipo	Volume
Sede	E	10
Dipartimento	E	80
Impiegato	E	2000
Progetto	E	500
Composizione	R	80
Afferenza	R	1900
Direzione	R	80
Partecipazione	R	6000

Esempio di valutazione di costo

- Operazione:
 - trova tutti i dati di un impiegato, del dipartimento nel quale lavora e dei progetti ai quali partecipa
- Si costruisce una tavola degli accessi basata su uno schema di navigazione

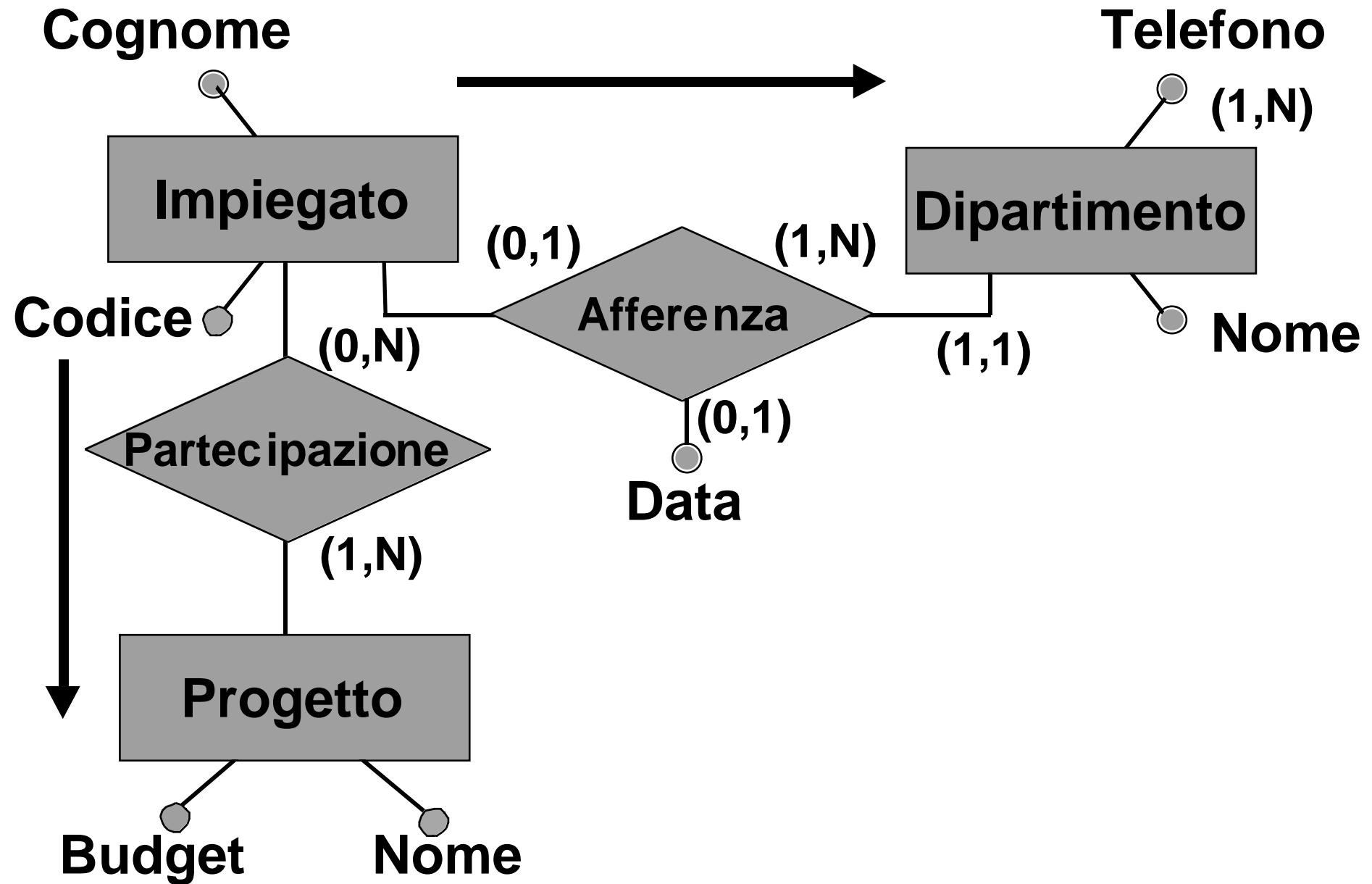


Tavola degli accessi

Concetto	Costrutto	Accessi	Tipo
Impiegato	Entità	1	L
Afferenza	Relazione	1	L
Dipartimento	Entità	1	L
Partecipazione	Relazione	3	L
Progetto	Entità	3	L

Attività della ristrutturazione

- *Analisi delle ridondanze*
- Eliminazione delle generalizzazioni
- Partizionamento/accorpamento di entità e relationship
- Scelta degli identificatori primari

Analisi delle ridondanze

- Una ridondanza in uno schema E-R è una informazione significativa ma derivabile da altre
- in questa fase si decide se eliminare le ridondanze eventualmente presenti o mantenerle

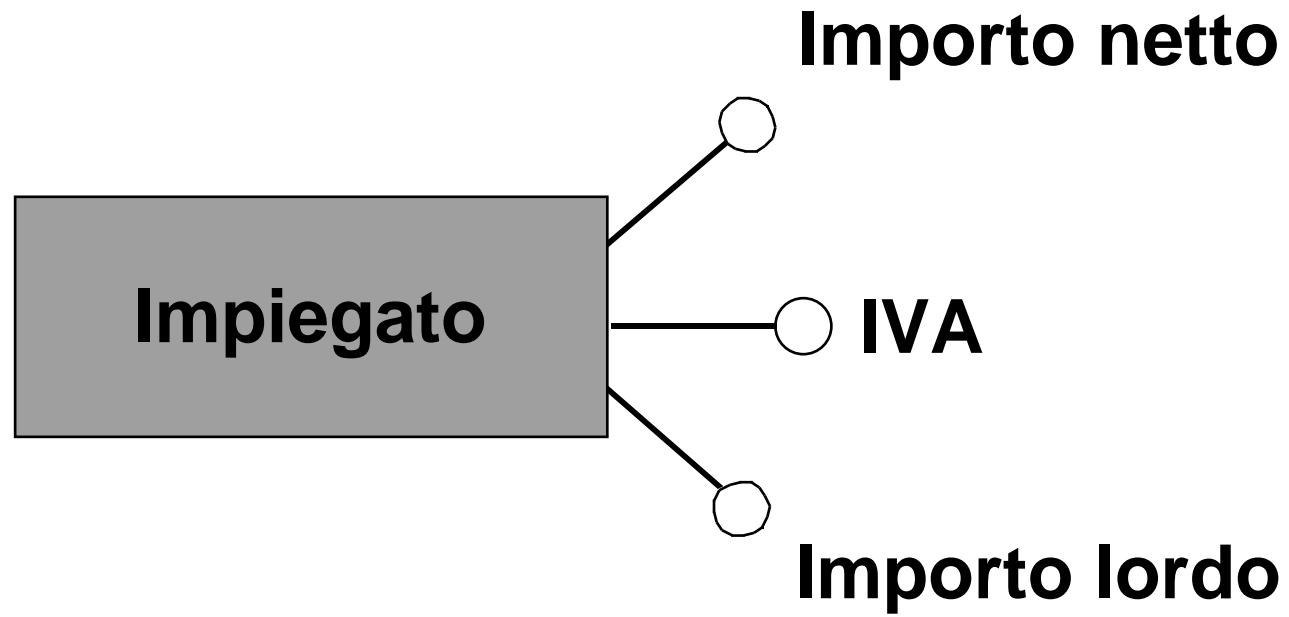
Ridondanze

- **Vantaggi**
 - semplificazione delle interrogazioni
- **Svantaggi**
 - appesantimento degli aggiornamenti
 - maggiore occupazione di spazio

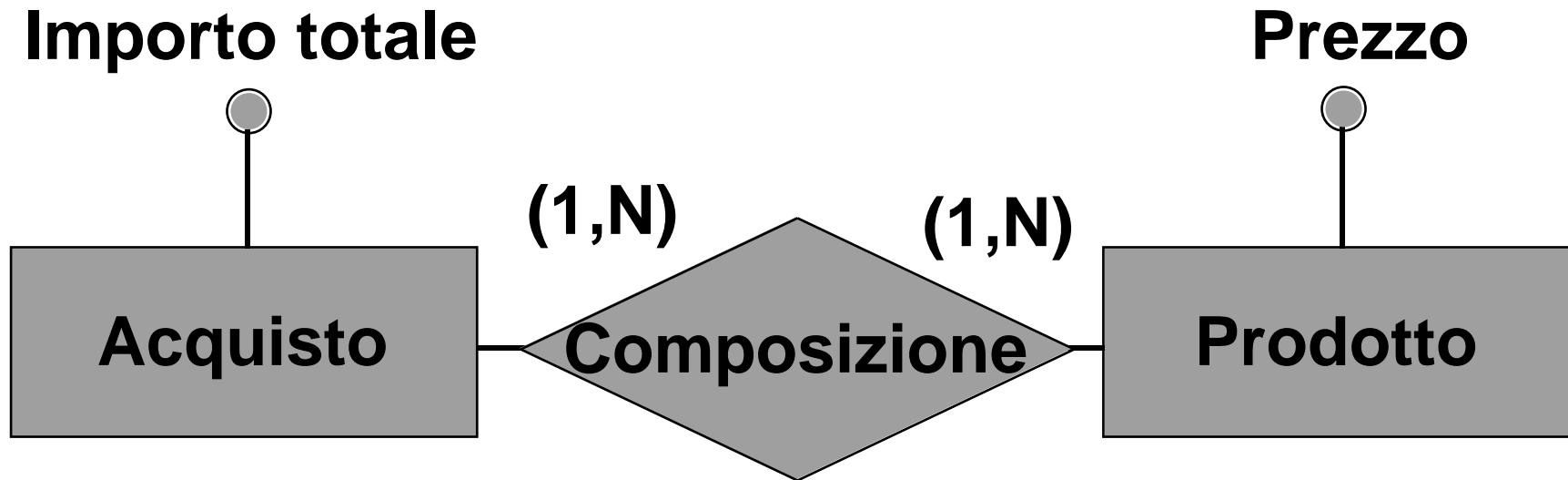
Forme di ridondanza in uno schema E-R

- **attributi derivabili:**
 - da altri attributi della stessa entità (o relazione)
 - da attributi di altre entità (o relazioni)
- **relazioni derivabili dalla composizione di altre relazioni in presenza di cicli**

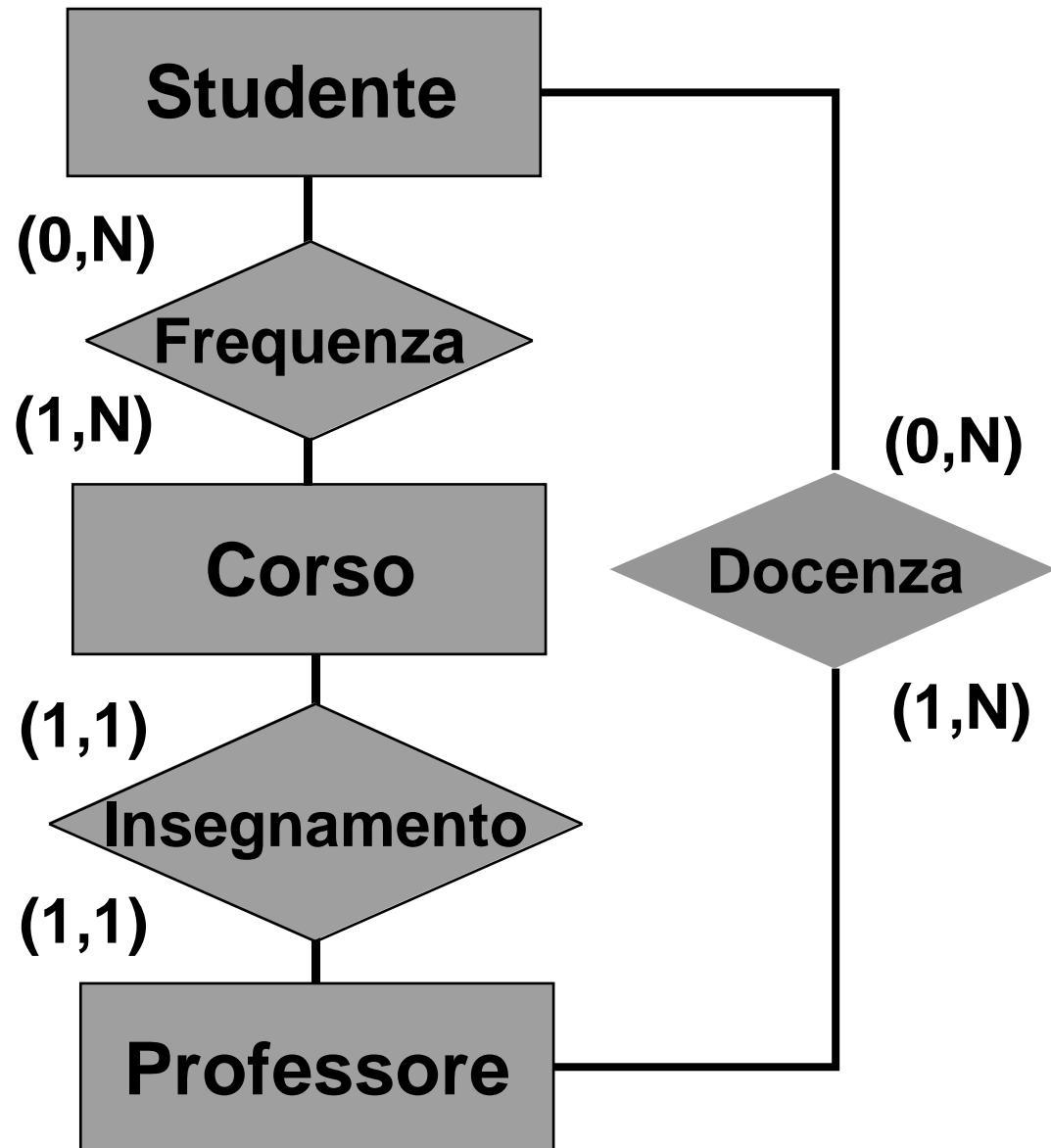
Attributo derivabile



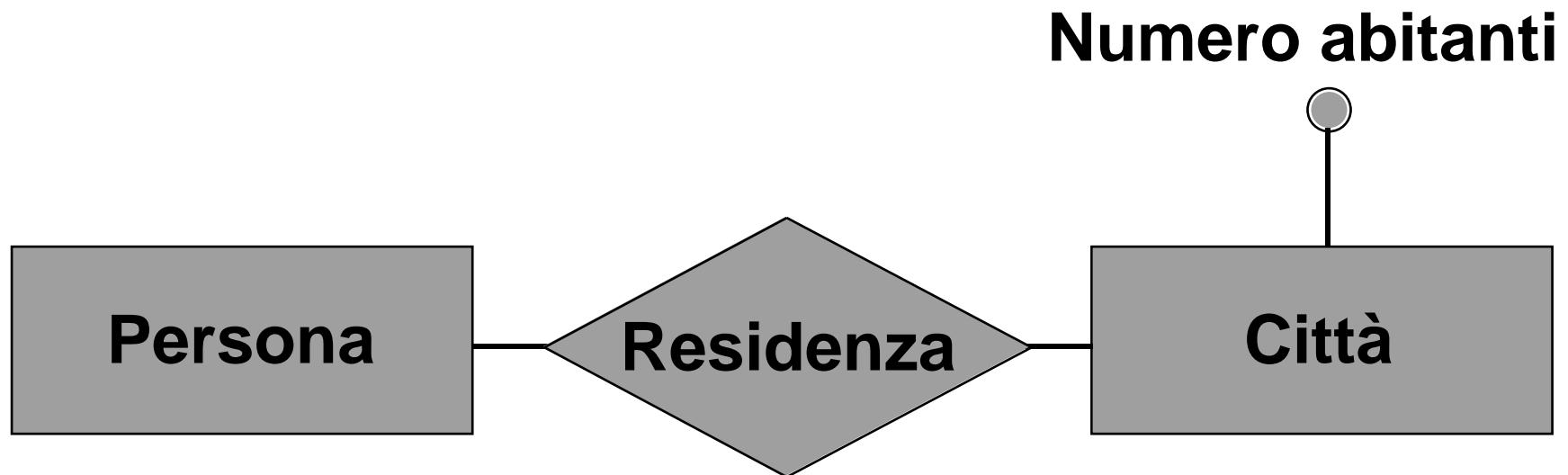
Attributo derivabile da altra entità



Ridondanza dovuta a ciclo



Analisi di una ridondanza



Concetto	Tipo	Volume
Città	E	200
Persona	E	1000000
Residenza	R	1000000

- **Operazione 1: memorizza una nuova persona con la relativa città di residenza (500 volte al giorno)**
- **Operazione 2: stampa tutti i dati di una città (incluso il numero di abitanti) (2 volte al giorno)**

Presenza di ridondanza

Operazione 1

Concetto	Costrutto	Accessi	Tipo
Persona	Entità	1	S
Residenza	Relazione	1	S
Città	Entità	1	L
Città	Entità	1	S

Operazione 2

Concetto	Costrutto	Accessi	Tipo
Città	Entità	1	L

Assenza di ridondanza

Operazione 1

Concetto	Costrutto	Accessi	Tipo
Persona	Entità	1	S
Residenza	Relazione	1	S

Operazione 2

Concetto	Costrutto	Accessi	Tipo
Città	Entità	1	L
Residenza	Relazione	5000	L

Presenza di ridondanza

- Costi:
 - Operazione 1: 1500 accessi in scrittura e 500 accessi in lettura al giorno
 - Operazione 2: trascurabile.
- Contiamo doppi gli accessi in scrittura
 - Totale di 3500 accessi al giorno

Assenza di ridondanza

- Costi:
 - Operazione 1: 1000 accessi in scrittura
 - Operazione 2: 10000 accessi in lettura al giorno
- Contiamo doppi gli accessi in scrittura
 - Totale di 12000 accessi al giorno

Attività della ristrutturazione

- Analisi delle ridondanze
- *Eliminazione delle generalizzazioni*
- Partizionamento/accorpamento di entità e relazioni
- Scelta degli identificatori primari

Eliminazione delle gerarchie

- **il modello relazionale non può rappresentare direttamente le generalizzazioni**
- **entità e relazioni sono invece direttamente rappresentabili**
 - **si eliminano perciò le gerarchie, sostituendole con entità e relazioni**

Tre possibilità

- 1. accorpamento delle figlie della generalizzazione nel genitore**
- 2. accorpamento del genitore della generalizzazione nelle figlie**
- 3. sostituzione della generalizzazione con relazioni**

- **la scelta fra le alternative si può fare con metodo simile a quello visto per l'analisi delle ridondanze (però non basato solo sul numero degli accessi)**
- **è possibile seguire alcune semplici regole generali**

- 1. conviene se gli accessi al padre e alle figlie sono contestuali**
 - 2. conviene se gli accessi alle figlie sono distinti**
 - 3. conviene se gli accessi alle entità figlie sono separati dagli accessi al padre**
-
- **sono anche possibili soluzioni “ibride”, soprattutto in gerarchie a più livelli**

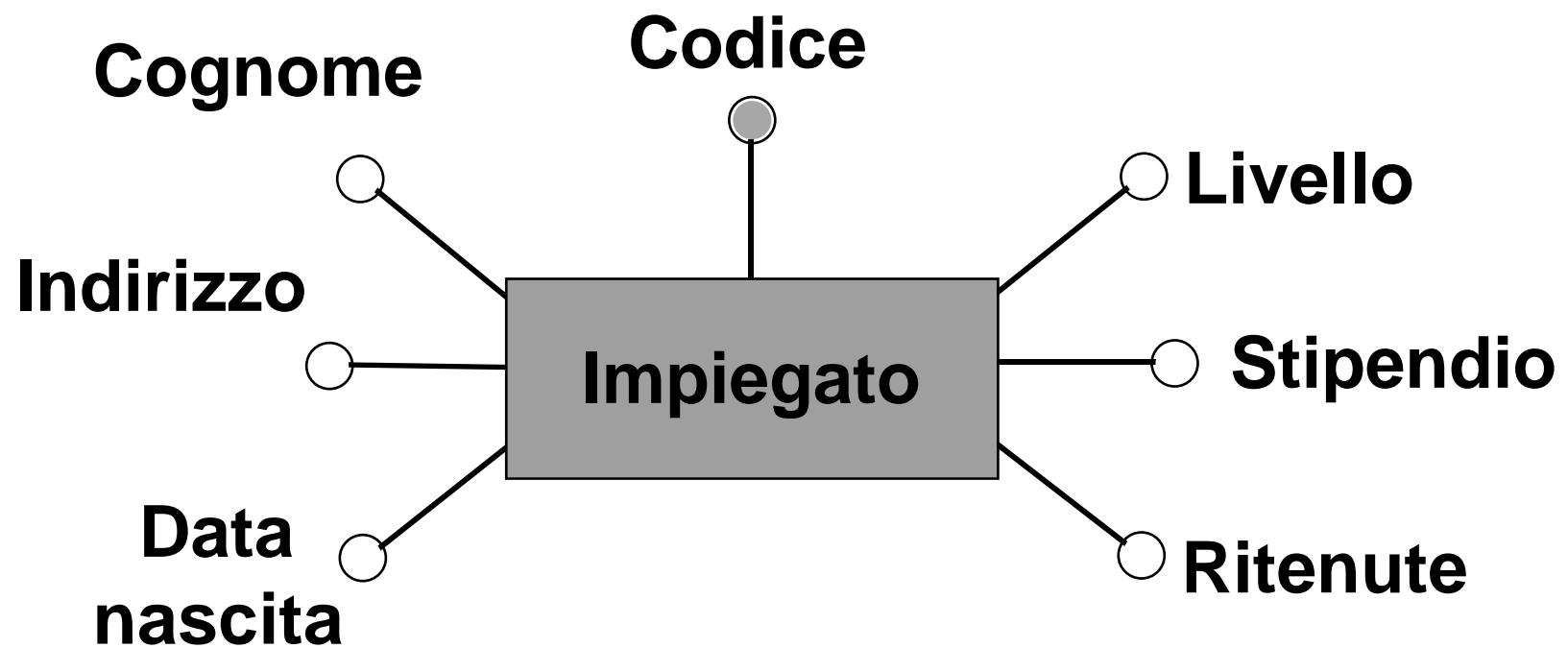
Attività della ristrutturazione

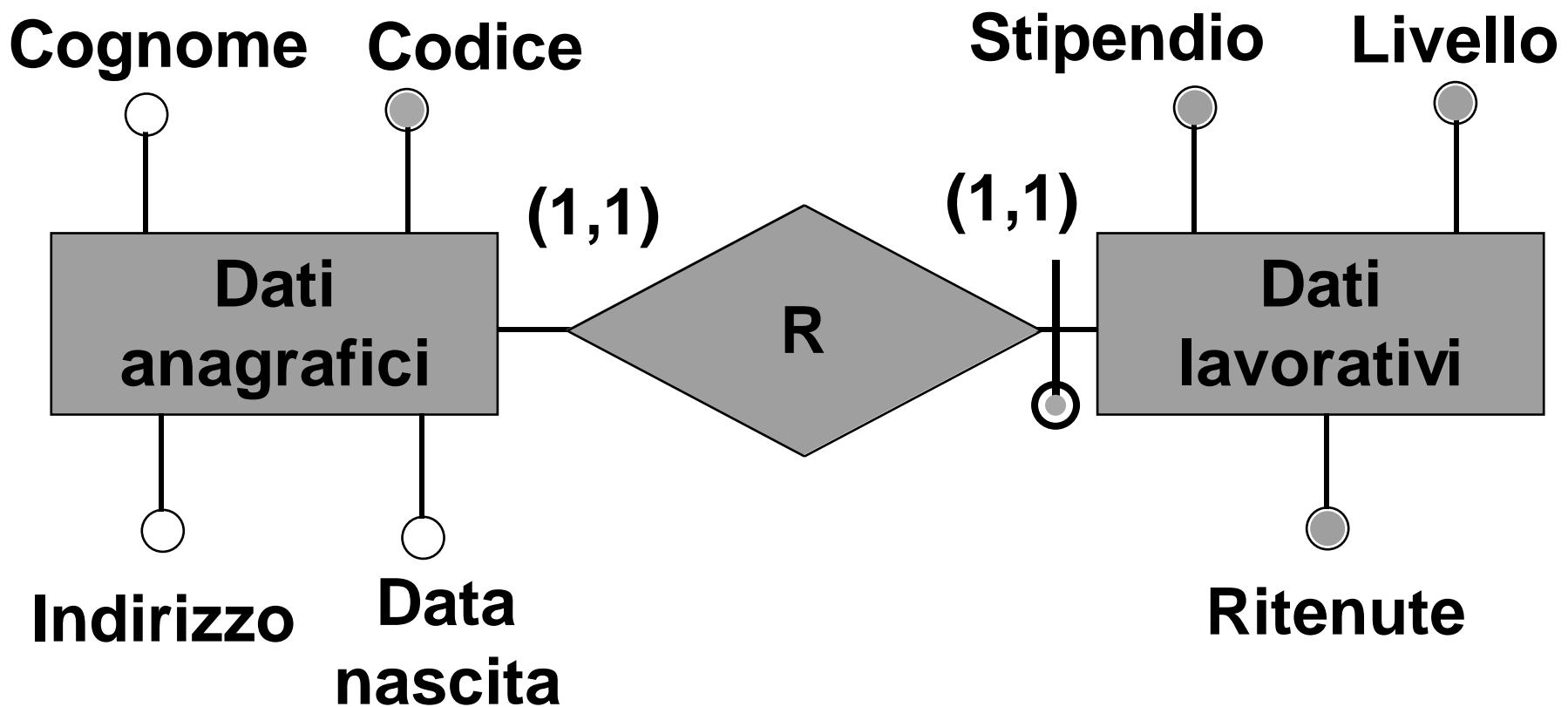
- Analisi delle ridondanze
- Eliminazione delle generalizzazioni
- *Partizionamento/accorpamento di entità e relazioni*
- Scelta degli identificatori primari

- **Ristrutturazioni effettuate per rendere più efficienti le operazioni in base a un semplice principio**
- **Gli accessi si riducono:**
 - **separando attributi di un concetto che vengono acceduti separatamente**
 - **raggruppando attributi di concetti diversi acceduti insieme**

Ristrutturazioni, casi principali

- **partizionamento verticale di entità**
- **partizionamento orizzontale di relationship**
- **eliminazione di attributi multivалore**
- **accorpamento di entità/
relationship**





Attività della ristrutturazione

- Analisi delle ridondanze
- Eliminazione delle generalizzazioni
- Partizionamento/accorpamento di entità e relazioni
- *Scelta degli identificatori primari*

Scelta degli identificatori primari

- **operazione indispensabile per la traduzione nel modello relazionale**
- **Criteri**
 - **assenza di opzionalità**
 - **semplicità**
 - **utilizzo nelle operazioni più frequenti o importanti**

**Se nessuno degli identificatori soddisfa i
requisiti visti?**

**Si introducono nuovi attributi (codici)
contenenti valori speciali generati
appositamente per questo scopo**

Traduzione verso il modello relazionale

- **le regole già viste:**
 - **le entità diventano relazioni sugli stessi attributi**
 - **le associazioni (ovvero le relazioni E-R) diventano relazioni sugli identificatori delle entità coinvolte (più gli attributi propri)**
 - **compattamento per associazioni 1:M e 1:1**

Strumenti di supporto

- **Esistono sul mercato prodotti CASE che forniscono un supporto a tutte le fasi della progettazione di basi di dati**

