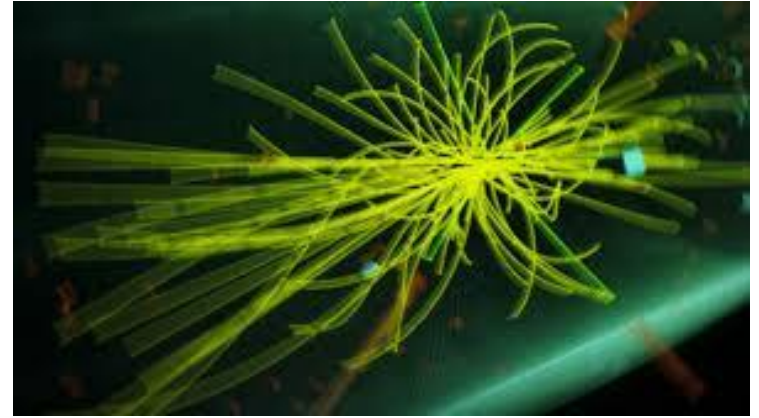


Gestione informatica dei dati

Anno accademico 2018/2019



Dalla matrice alla base di dati

La matrice dei dati

I dati possono essere rilevati attraverso un questionario, ad esempio chiedendo al cliente i dati rilevanti (nome, cognome, indirizzo, ecc.).

Il questionario è certamente uno strumento efficiente per la raccolta dei dati, scopo per il quale è stato inventato, ma costituisce un supporto inadeguato per la conservazione e, soprattutto, per l'analisi dei dati stessi.

La tecnologia permette di costruire e operare su matrici di dati.

La matrice di dati è una tabella con specifiche relazioni tra righe e colonne.



Gestione informatica dei dati

La matrice dei dati

La matrice di dati, detta anche «casi per variabili», consiste in un insieme rettangolare di numeri, dove in riga abbiamo i casi e in colonna le variabili.

In ogni cella derivante dall'incrocio fra una riga e una colonna abbiamo un dato, e cioè il valore assunto da una particolare variabile su un particolare caso.

		Codifica (tracciato record e codice)				
Informazioni		Matrice dei dati				
		VARIABILI				
		X_1	X_2	X_3	...	X_p
CASI	1	X_{11}	X_{12}	X_{13}	...	X_{1p}
	2	X_{21}	X_{22}	X_{23}	...	X_{2p}
	3	X_{31}	X_{32}	X_{33}	...	X_{3p}

	n	X_{n1}	X_{n2}	X_{n3}	...	X_{np}

La matrice dei dati

Due sono le condizioni necessarie perché le informazioni afferenti ad un certo insieme di casi possano essere organizzate nella forma di matrice-dati:

- l'unità d'analisi deve essere sempre la stessa: per esempio deve trattarsi di informazioni raccolte tutte su clienti, oppure tutte su ambiti territoriali o, ancora, su prodotti; non si può fare una matrice-dati dove alcune righe contengono clienti, altre prodotti, ecc.;
- su tutti i casi studiati devono essere state rilevate le stesse informazioni; nella matrice-dati le righe hanno la stessa lunghezza e contengono le stesse variabili; non è possibile costruire una matrice dati se su un certo numero di casi sono state raccolte certe informazioni e su altri casi ne sono state raccolte delle altre.



La matrice dei dati

L'operazione di traduzione del materiale empirico grezzo (il pacco di questionari, la pila di documenti) in matrice-dati viene chiamata codifica, e avviene con l'ausilio di due strumenti, il tracciato-record ed il codice:


- il tracciato-record indica la posizione di ogni variabile nella riga della matrice; il termine «tracciato-record» deriva dalla lingua inglese, dove per «record» si intende la riga della matrice;
- il codice (codebook) assegna ad ogni modalità della variabile un valore numerico o una etichetta.



Gestione informatica dei dati

La matrice

Esempio di
tracciato
record

INDAGINE: Iscrizioni e cancellazioni all'anagrafe per trasferimento di residenza Anno 2013 					
Tracciato record valido dal 1995					
num. ordine	Posizione	Lunghezza	Acronimovariabile	TipoVariabile	Aggregazione
1	1	4	anno_di_rilevazione	-	
2	5	2	mese_di_rilevazione	Categorica	
3	7	5	filler		
4	12	5	filler		
5	17	1	conteggio	Categorica	
6	18	1	tipo_di_provvedimento	Categorica	
7	19	3	iscrizione_provincia	Categorica	
8	22	3	iscrizione_comune	Categorica	iscrizione_provincia
9	25	3	cancellazione_provincia	Categorica	

La matrice dei dati

Esempio di codebook

ELENCO MODALITA' della variabile: stato_civile (32,0) Anno: 2013

Modalita'	Descrizione	Codice aggregazione	Note
1	celibe/nubile		
2	coniugato/coniugata		
3	vedovo/vedova		
4	divorziato/divorziata		

Gestione informatica dei dati

Esempio di un questionario con diverse modalità di codifica delle variabili.

Tratto da Corbetta P., La ricerca sociale, vol. IV, il Mulino, 2003

N.ordine generale ☐ ☐ ☐ ☐ 1-4

1. Secondo lei oggi in Italia rispetto a 5 anni fa, si vive meglio, nello stesso modo o peggio?

- Meglio ☐ 1 ☐ 5
- Stesso modo ☐ 2
- Peggio ☐ 3
- Non risponde ☐ 9

2. Si parla molto di ciò che si dovrebbe fare in Italia nei prossimi 10 o 15 anni. In questo cartellino sono indicati vari obiettivi. (MOSTRARE CARTELLINO 1). Secondo lei quale di questi è il più importante, cioè quale metterebbe al primo posto? E quale al secondo?

- | | primo | secondo | |
|--|----------------------------|----------------------------|------------------------------|
| • Mantenere l'ordine nel paese | <input type="checkbox"/> 1 | <input type="checkbox"/> 1 | <input type="checkbox"/> 6-7 |
| • Aumentare per i cittadini la possibilità di partecipare al governo | <input type="checkbox"/> 2 | <input type="checkbox"/> 2 | |
| • Controllare l'aumento dei prezzi | <input type="checkbox"/> 3 | <input type="checkbox"/> 3 | |
| • Garantire la libertà d'opinione | <input type="checkbox"/> 4 | <input type="checkbox"/> 4 | |
| • Non risponde | <input type="checkbox"/> 9 | <input type="checkbox"/> 9 | |

3. Vorrei conoscere il suo punto di vista su alcuni problemi sociali. Mi dica per ciascuno di questi se lei lo considera un problema molto, abbastanza, poco o per niente importante?

- | | abba. | | | | nin. | | | | |
|---------------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-----------------------------|--|--|
| | mol. | pac. | | | | | | | |
| • L'inflazione, l'aumento dei prezzi | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 8 | | |
| • La disoccupazione | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 9 | | |
| • L'inefficienza dei servizi pubblici | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 10 | | |
| • La criminalità organizzata | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 11 | | |
| • La droga | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 12 | | |
| • L'inquinamento | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 13 | | |
| • L'immigrazione straniera | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 14 | | |
| • La corruzione politica | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 | <input type="checkbox"/> 9 | <input type="checkbox"/> 1 | <input type="checkbox"/> 15 | | |

4. Ora le leggerò un elenco di problemi urbani. Vorrei sapere di quale di questi, qui a ... (NOME DEL COMUNE), lei è maggiormente insoddisfatto (MASSIMO DUE RISPOSTE)

- | | la risp. | la risp. | |
|-------------------------------|----------------------------|----------------------------|--------------------------------|
| • Trasporti pubblici | <input type="checkbox"/> 1 | <input type="checkbox"/> 1 | <input type="checkbox"/> 16-17 |
| • Orari degli uffici pubblici | <input type="checkbox"/> 2 | <input type="checkbox"/> 2 | |
| • Traffico | <input type="checkbox"/> 3 | <input type="checkbox"/> 3 | |
| • Nettezza urbana | <input type="checkbox"/> 4 | <input type="checkbox"/> 4 | |
| • Illuminazione pubblica | <input type="checkbox"/> 5 | <input type="checkbox"/> 5 | |
| • Approvvigionamento acqua | <input type="checkbox"/> 6 | <input type="checkbox"/> 6 | |
| • Non risponde | <input type="checkbox"/> 9 | <input type="checkbox"/> 9 | |

5. Ora le leggerò alcune azioni che la gente talvolta fa per protestare o per influire sul governo. Lei mi dovrebbe dire se le è mai capitato di compiere qualcuna di queste azioni.

- | | si | no-NR | |
|---------------------------------------|----------------------------|----------------------------|-----------------------------|
| • Partecipare a scioperi spontanei | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 18 |
| • Bloccare il traffico | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 19 |
| • Fare l'autoriduzione dell'affitto | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 20 |
| • Fare l'autoriduzione delle bollette | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 21 |
| • Occupare case sfitte | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 22 |
| • Occupare fabbriche | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 23 |
| • Scrivere slogan sui muri | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 24 |

6. Nel nostro studio abbiamo incontrato molte persone che, pur andando a votare, dichiarano di non interessarsi affatto di problemi politici. Secondo lei per quali ragioni in genere la gente non si interessa di politica?

☐ ☐ 25-26

Dal questionario alla matrice dei dati

1. Secondo lei oggi in Italia rispetto a 5 anni fa, si vive meglio, nello stesso modo o peggio?

- Meglio 1
- Stesso modo 2
- Peggio 3
- Non risponde 9

La prima domanda è una domanda standard che non pone problemi e dà luogo ad una variabile. Notiamo solo che fra le risposte codificate, oltre a quelle proposte dalla domanda chiusa, abbiamo aggiunto anche il valore 9 per la modalità di risposta «non risponde». Questa eventualità va sempre prevista in tutte le domande.

Dal questionario alla matrice dei dati

2. Si parla molto di ciò che si dovrebbe fare in Italia nei prossimi 10 o 15 anni. In questo cartellino sono indicati vari obiettivi. (MOSTRARE CARTELLINO 1). Secondo lei quale di questi è il più importante, cioè quale metterebbe al primo posto? E quale al secondo?

	primo	secondo
• Mantenere l'ordine nel paese	1	1
• Aumentare per i cittadini la possibilità di partecipare al governo	2	2
• Controllare l'aumento dei prezzi	3	3
• Garantire la libertà d'opinione	4	4
• Non risponde	9	9

La domanda 2 **produce due variabili**: quella relativa agli obiettivi indicati per primi, e quella relativa agli obiettivi indicati per secondi.

Gestione informatica dei dati

Dal questionario alla matrice dei dati

3. Vorrei conoscere il suo punto di vista su alcuni problemi sociali. Mi dica per ciascuno di questi se lei lo considera un problema molto, abbastanza, poco o per niente importante?

	abb.		nie.			
	mol.	poc.		NR		
• L'inflazione, l'aumento dei prezzi	1	2	3	4	9	8
• La disoccupazione	1	2	3	4	9	9
• L'inefficienza dei servizi pubblici	1	2	3	4	9	10
• La criminalità organizzata	1	2	3	4	9	11
• La droga	1	2	3	4	9	12
• L'inquinamento	1	2	3	4	9	13
• L'immigrazione straniera	1	2	3	4	9	14
• La corruzione politica	1	2	3	4	9	15

La terza domanda è di fatto una **batteria di domande**, nel senso che si tratta di più domande aventi lo stesso formato di risposta (quattro alternative). Per comodità grafica esse sono raggruppate apparentemente sotto un'unica domanda, ma di fatto si tratta di 8 domande differenti, che danno luogo a 8 variabili ognuna con la sua sequenza di risposte.

Dal questionario alla matrice dei dati

4. Ora le leggerò un elenco di problemi urbani. Vorrei sapere di quale di questi, qui a ... (NOME DEL COMUNE), lei è maggiormente insoddisfatto (MASSIMO DUE RISPOSTE)

	1a risp.	2a risp.	16-17
• Trasporti pubblici	1	1	
• Orari degli uffici pubblici	2	2	
• Traffico	3	3	
• Nettezza urbana	4	4	
• Illuminazione pubblica	5	5	
• Approvvigionamento acqua	6	6	
• Non risponde	9	9	

La domanda 4 è a **risposta multipla**; cioè la stessa domanda ammette più risposte. Normalmente la domanda ammette una sola risposta. Se la domanda ammette più risposte, la domanda genera più variabili (tante variabili quanto è il numero delle risposte che il soggetto può dare). La domanda 4 può avere al massimo due risposte.

Gestione informatica dei dati

Dal questionario alla matrice dei dati

Anche la domanda 5 è a **risposta multipla**: il soggetto può dare tante risposte quante sono le alternative presenti. Poiché le alternative proposte sono 7, occorre prevedere 7 colonne, in ciascuna delle quali si registra se il soggetto ha risposto sì o no a quella specifica alternativa.

5. Ora le leggerò alcune azioni che la gente talvolta fa per protestare o per influire sul governo. Lei mi dovrebbe dire se le è mai capitato di compiere qualcuna di queste azioni.

	sì	no-NA	
• Partecipare a scioperi spontanei	1	2	<input type="checkbox"/> 18
• Bloccare il traffico	1	2	<input type="checkbox"/> 19
• Fare l'autoriduzione dell'affitto	1	2	<input type="checkbox"/> 20
• Fare l'autoriduzione delle bollette	1	2	<input type="checkbox"/> 21
• Occupare case sfitte	1	2	<input type="checkbox"/> 22
• Occupare fabbriche	1	2	<input type="checkbox"/> 23
• Scrivere slogan sui muri	1	2	<input type="checkbox"/> 24

Dal questionario alla matrice dei dati

Infine, la domanda 6 è una **domanda aperta**, cioè non viene offerto all'intervistato un ventaglio di possibili risposte, ma egli risponde come meglio gli aggrada. In questo caso la codifica viene effettuata a posteriori. In sede di intervista l'intervistatore trascrive sul questionario la risposta liberamente data.

6. Nel nostro studio abbiamo incontrato molte persone che, pur andando a votare, dichiarano di non interessarsi affatto di problemi politici. Secondo lei per quali ragioni in genere la gente non si interessa di politica?

25-26'

Gestione informatica dei dati

La matrice dei dati

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	1	7	4	8	3	7	8	9	2	0	3	8	1	8	1
2	2	0	2	4	5	6	3	3	2	7	6	8	5	8	0
3	3	0	2	1	4	1	1	3	5	5	7	2	3	0	7
4	4	3	5	2	6	6	4	8	0	3	4	5	1	6	4
5	5	9	2	5	9	1	4	4	0	0	3	4	8	9	8
6	6	7	2	2	6	0	7	1	8	6	6	3	5	6	6
7	7	7	1	6	4	0	6	3	6	0	6	7	9	5	6
8	8	1	5	4	9	9	9	0	4	2	2	7	3	0	6
9	9	5	4	3	2	7	5	5	8	3	4	6	8	7	4
10	10	6	1	3	1	2	6	7	7	1	3	9	4	6	0
11															

La trasformazione delle risposte dei questionari in un file informatico produce una matrice rettangolare $n \times m$ (n = casi, m =variabili)



Gestione informatica dei dati

La matrice dei dati

Una tabella è una raccolta di informazioni omogenee, cioè relative a un insieme di dati riguardanti lo stesso argomento. Ad esempio, una tabella Giocatori conterrà tutti i dati utili a descrivere gli atleti in campo (cognome, nome, ecc.).

Una tabella è organizzata in righe e colonne.

Ogni riga rappresenta un record, ovvero un insieme di informazioni.

Ogni colonna rappresenta un campo, ovvero un dato specifico di informazione.

	A	B	C	D
1	Buffon	Gianluigi	1	
2	Zambrotta	Gianluca	19	
3	Cannavaro	Fabio	5	
4	Materazzi	Marco	23	
5	Grosso	Fabio	3	
6	Gattuso	Gennaro	8	
7	Pirlo	Andrea	21	
8	Camoranesi	Mauro German	16	
9	Perrotta	Simone	20	
10	Totti	Francesco	10	
11	Toni	Luca	9	
12				
13				

Gestione informatica dei dati

Sistema di gestione di basi di dati DataBase Management System (DBMS)



Accezione generica, metodologica:

- Insieme organizzato di dati utilizzati per il supporto allo svolgimento delle attività di un ente (azienda, ufficio, persona)

Accezione specifica, metodologica e tecnologica:

- insieme di dati gestito da un DBMS

Gestione informatica dei dati

Sistema di gestione di basi di dati DataBase Management System (DBMS)

È importante sottolineare la differenza tra un database e un programma per la gestione di un database.

Il database contiene i dati.

Il programma per la gestione del database (DBMS: Database Management System) serve a organizzare i dati e consente all'utente di effettuare le operazioni su di essi, come inserire i dati, modificarli, cancellarli, ricercarli, visualizzarli, stamparli.



Gestione informatica dei dati

Sistema di gestione di basi di dati DataBase Management System (DBMS)

Il DBMS gestisce collezioni di dati:

- grandi
- persistenti
- condivise

Il DBMS garantisce:

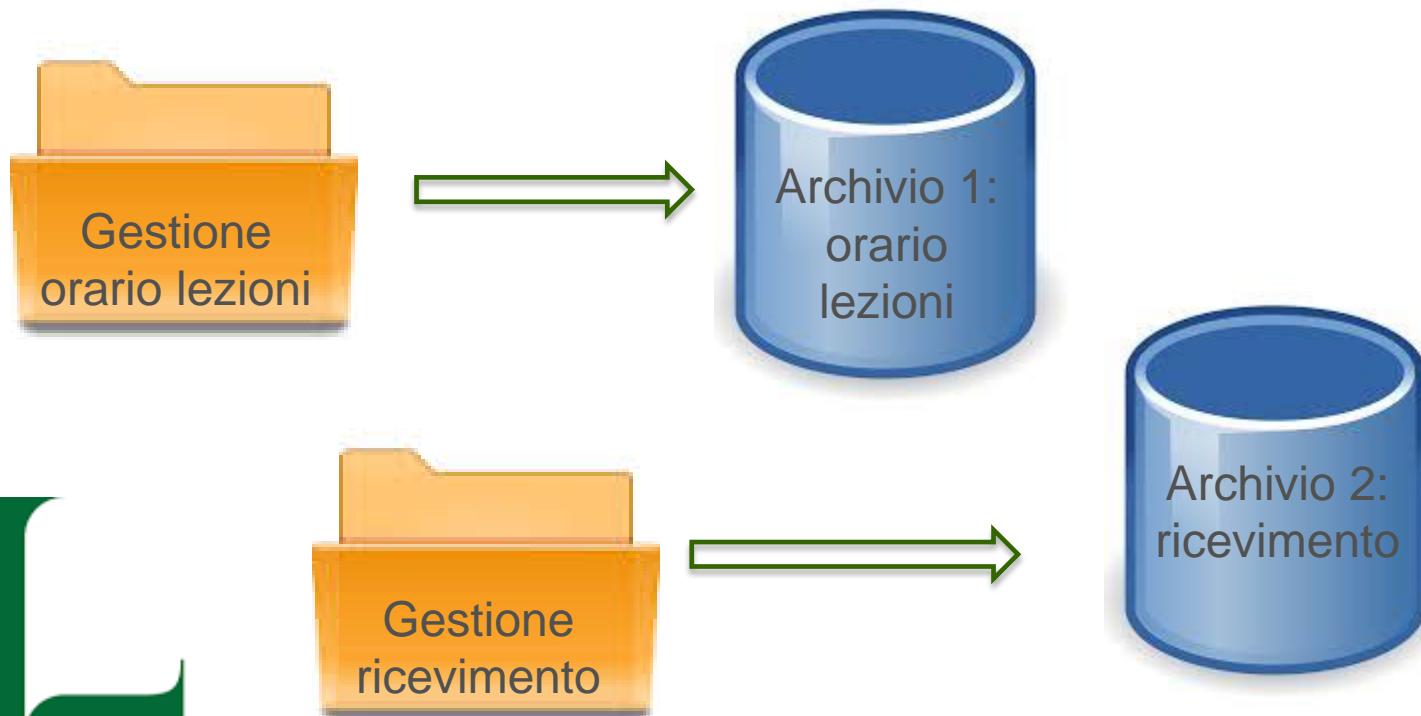
- privacy
- affidabilità
- efficienza
- efficacia



Gestione informatica dei dati

Dagli archivi alle basi di dati

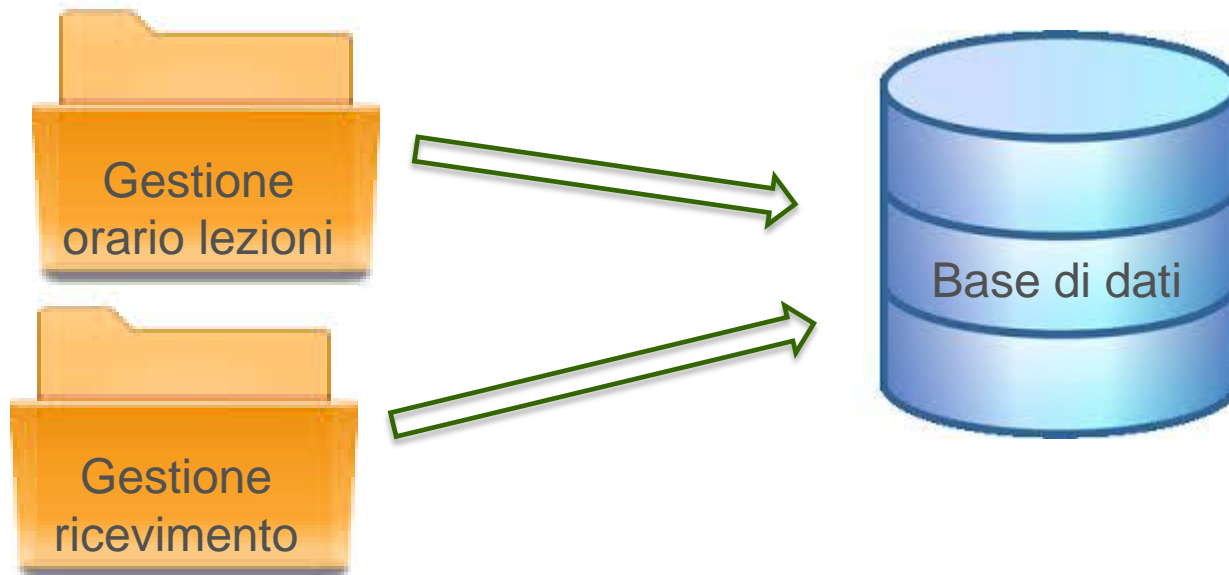
- Ridondanza:
 - informazioni ripetute
- Rischio di incoerenza:
 - le versioni possono non coincidere



Gestione informatica dei dati

Dagli archivi alle basi di dati

- Una base di dati è una risorsa integrata, condivisa fra applicazioni
- Conseguenze
 - Attività diverse su dati condivisi:
 - meccanismi di autorizzazione
 - Accessi di più utenti ai dati condivisi:
 - controllo della concorrenza



Dagli archivi alle basi di dati

- I modelli sono insiemi di costrutti utilizzati per organizzare i dati di interesse e descriverne la dinamica
- Componente fondamentale dei modelli sono i meccanismi di strutturazione (o costruttori di tipo)
- Ogni modello di dati prevede alcuni costruttori (esempio: il costruttore «relazione» permette di definire insiemi di record omogenei)




Gestione informatica dei dati

Dagli archivi alle basi di dati

- Per costruire una base di dati necessitano uno schema e una istanza.

Lo schema, sostanzialmente invariante nel tempo, ne descrive la struttura (possiamo chiamarlo l'aspetto intensionale).



Nome campo	Tipo dati
CODICE-txt	Testo
CODICE-num	Numerico
REG	Numerico
N_PROV	Numerico
N_COM	Numerico
SIGLA	Testo
COMUNE	Testo

Gestione informatica dei dati

Dagli archivi alle basi di dati

L'istanza, ovvero i valori attuali della tabella, che possono variare (possiamo chiamarlo l'aspetto estensionale del data base)

0_Classificazioni						
CODICE-txt	CODICE-nun	REG	N_PROV	N_COM	SIGLA	COMUNE
081001	81001	19	81	1	TP	ALCAMO
081002	81002	19	81	2	TP	BUSETO PALIZZOLO
081003	81003	19	81	3	TP	CALATAFIMI
081004	81004	19	81	4	TP	CAMPOBELLO DI MAZARA
081005	81005	19	81	5	TP	CASTELLAMMARE DEL GOLFO
081006	81006	19	81	6	TP	CASTELVETRANO
081007	81007	19	81	7	TP	CUSTOMACI
081008	81008	19	81	8	TP	ERICE
081009	81009	19	81	9	TP	FAVIGNANA
081010	81010	19	81	10	TP	GIRFILI INA

Architettura di un DBMS

Schema logico: descrizione dell'intera base di dati nel modello logico “principale” del DBMS

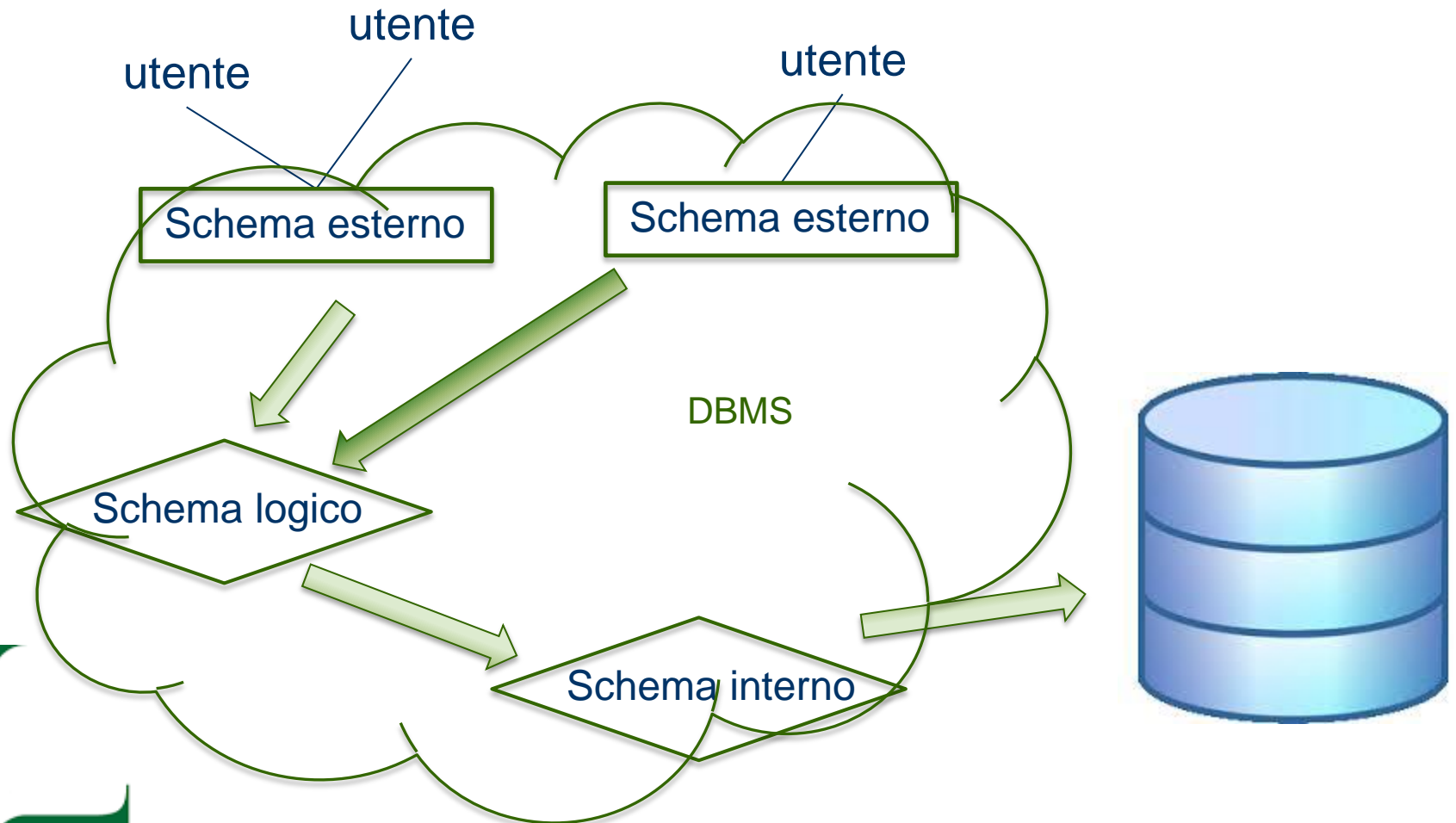
Schema interno (o fisico): rappresentazione dello schema logico per mezzo di strutture fisiche di memorizzazione

Schema esterno: descrizione di parte della base di dati in un modello logico (“viste” parziali, derivate, anche in modelli diversi)



Gestione informatica dei dati

Architettura di un DBMS



Architettura di un DBMS

Come conseguenza dell'articolazione in livelli della banca dati possiamo indicare:

- che l'accesso avviene solo tramite il livello esterno (che potrebbe anche coincidere con il livello logico) e che può essere tarato per agevolare quanto più possibile l'utente, senza dover intervenire sul livello fisico.
- Grazie ai livelli diversi che compongono l'architettura di un DBMS possiamo avere una indipendenza fisica e una indipendenza logica



Architettura di un DBMS

- il livello logico e quello esterno sono indipendenti da quello **interno** in quanto:
 - una relazione è utilizzata nello stesso modo qualunque sia la sua realizzazione fisica
 - la realizzazione fisica può cambiare senza che debbano essere modificati i programmi
- il livello esterno è indipendente da quello **logico** in quanto:
 - aggiunte o modifiche alle viste non richiedono modifiche al livello logico
 - modifiche allo schema logico che lascino inalterato lo schema esterno sono trasparenti per l'utente



Architettura di un DBMS

Linguaggi per un data base

Data definition language (DDL): linguaggi utilizzati per programmi per la definizione di schemi (logici, esterni, fisici) e altre operazioni generali. Viene utilizzato per la creazione della struttura fisica del DB e degli schemi che la connettono con la struttura logica.

Data manipulation language (DML): linguaggi utilizzati per programmi di interrogazione e l'aggiornamento di (istanze di) basi di dati.

Storage definition language (SDL): viene utilizzato per specificare lo schema interno (viene usato assieme al DDL nei DB che mantengono una separazione tra schema concettuale e schema interno).

View definition language (VDL): specifica gli schemi esterni per gli utenti e li connette allo schema concettuale.



Gestione informatica dei dati

Architettura di un DBMS

