

Progetto Formativo Aziendale  
ORGANIZZAZIONE E UTILIZZO DI DATASET DI DATI ANAMNESTICI EDESITI  
DIAGNOSTICI: RACCOLTA ED ARCHIVIAZIONE DEI DATI  
Roma, 7-8 giugno 2016

## Cosa intendiamo per “dato”

Marcello Sala  
Istituto Zooprofilattico Sperimentale delle Regioni Lazio e Toscana  
Osservatorio Epidemiologico Veterinario Regione Lazio  
Roma



# Dato

- È una unità di “informazione” che diviene informazione quando è in grado di modificare la nostra percezione del mondo; la conoscenza è la trasformazione dell’informazione in valore

*(John Browning and Spencer Rice, Encyclopedia of the new economy – wired)*



# Dato

- È un elemento "grezzo"
- Non agito dal ricevente
- I dati opportunamente interpretati forniscono le informazioni



# Informazione

- Dato elaborato cognitivamente cioè..
- ....messaggio che riduce l'incertezza
- ....tiene conto delle relazioni esistenti in un contesto e deve essere utilizzabile in TEMPO REALE



# Conoscenza

- Informazione applicata da parte dell'osservatore che "sa" quando e come usarla
- Deriva da un insieme di percezioni ed aspettative che l'osservatore ha nei confronti di un evento
- È una predisposizione ad agire in un determinato modo

# Conoscenza

- Nel nostro caso la conoscenza deriva dalla aggregazione, descrizione ed elaborazione dei dati
- Questi sono processi non solo “tecnici” ma anche e soprattutto “mentali” perché implicano sempre una interpretazione finale

# What's different about knowledge?



Data

"1840KL0617"



Information

"The KLM flight leaves Detroit at 18.40"



Knowledge

"That's not a good flight; often busy and delayed"

Decontextualization

Contextualization



Marketing Intellig

Joep Wijman (Philips)

Luciano Pilotti  
Università degli Studi di Milano  
[luciano.pilotti@unimi.it](mailto:luciano.pilotti@unimi.it)

SPSS



Teradata  
a division of HCL

THE  
TECHNOLOGY  
PARTNERS

VALUE LAB  
Marketing Sales Training Solutions





## IL DATO

**019FR230**





## L'INFORMAZIONE

**019** = è il codice del comune

**FR** = è il codice provincia

**230** = è il progressivo dell'azienda

*Azienda agricola "xyz" nel comune di Cassino -  
Provincia di Frosinone - Lazio*



# LA CONOSCENZA

## DA Archivi:

azienda "xyz" di proprietà del sig. "Taldetali" in comune di Cassino - Provincia di Frosinone – Lazio opera come stalla di sosta in Via YYYZZZZ numero ##  
coord. Geografiche Long. EST ##,## Lat Nord##,##



# LA CONOSCENZA

Da altre fonti:

La provincia ha elevata densità di allevamenti familiari  
In passato notevoli flussi di partite suine provenienti  
dall'UMBRIA

Le strutture in cui vengono detenuti gli animali sono 2  
di vecchia costruzione

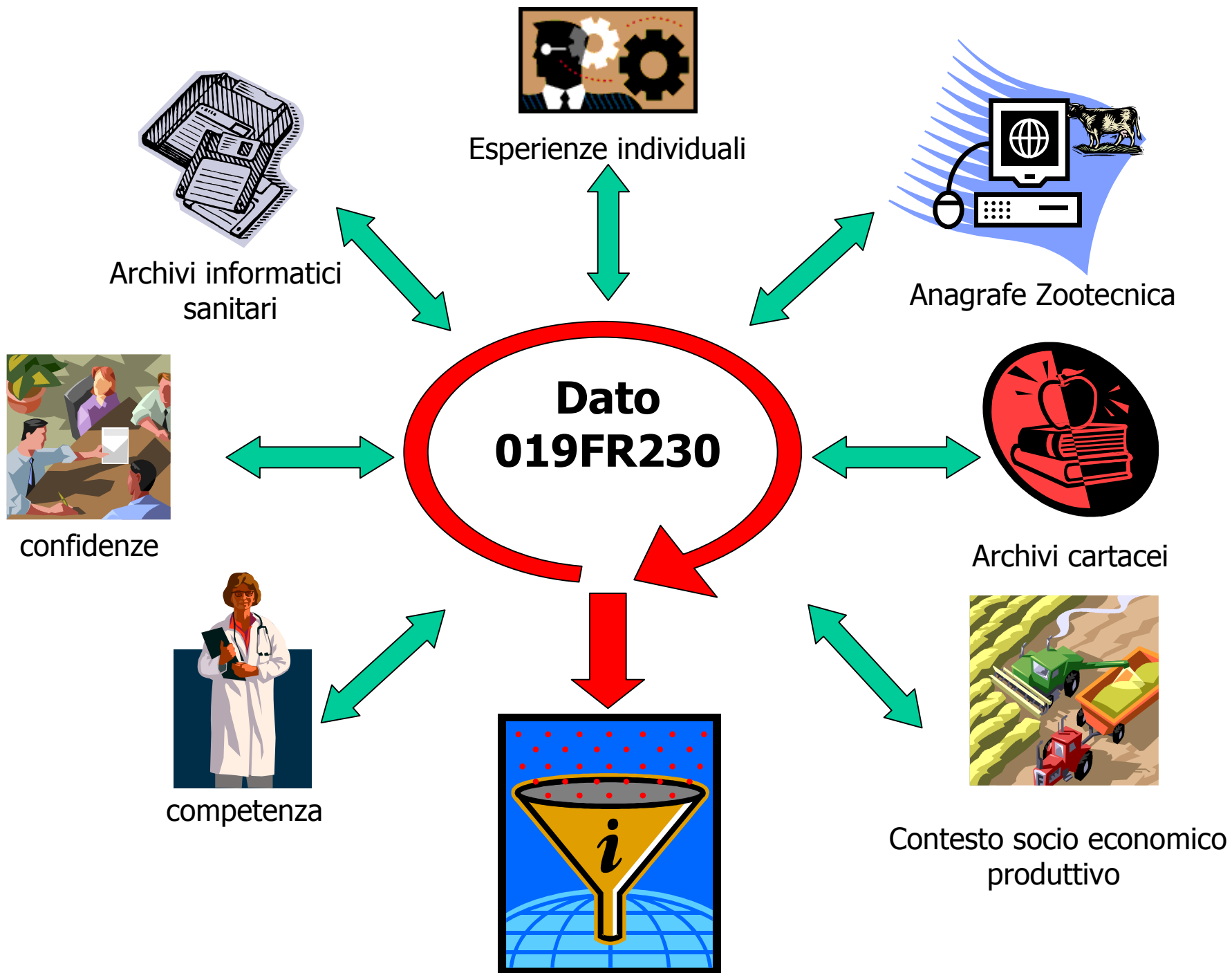
Il proprietario effettua anche trasporto di animali conto  
terzi

Azienda a rischio introduzione MVS



## La conoscenza

- Nel nostro settore la conoscenza si avvale del **processo** di integrazione di più livelli di informazione grazie alla possibilità di “incrociare” il singolo dato di partenza con molte fonti di altri dati e informazioni





## La raccolta dati è una “rete di informazione”

- Collaborazione tra le componenti (nodi) della rete di informazione
- Uso dello stesso linguaggio
- Interpretazione univoca-convenzionale degli elementi del linguaggio (variabili-dati-informazioni)

## Il Patto implicito

- Il significato della variabile
- La natura del dato
- Tipo e qualità dell'informazione



non variano lungo la "filiera" informativa

- Non ci deve essere interpretazione ma  
COERENZA di senso

*"stabilità del dato"*



**DATO**



**INFORMAZIONE**





# Essere pronti

preparedness

- Il primo punto necessario alla gestione di una raccolta dati è essere pronti alla raccolta dati

# Essere pronti

- La raccolta dati non è solo un atto formale
- Non serve a niente se eseguita male
- Dobbiamo pensare subito che i primi fruitori siamo noi

# Essere pronti

Raccolta



Archivio



Uso



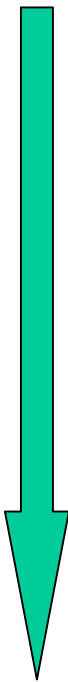
- Garbage In ... ..  
Garbage Out
- quello che ci metti ci  
ritrovi
- tutti i nodi vengono al  
pettine

# Essere pronti

- Conoscere gli strumenti della raccolta dati
  - Schede di prelievo
  - Indagini Epidemiologiche
  - Questionari
  - Estrazioni da archivi informatici (SIL ecc..)
- Già Pronti o Fai da te

# Passo 1

Arrivare preparati alla raccolta dati

- 
1. Chiarirsi molto bene lo scopo
  2. Identificare i dati critici per garantire completezza e appropriatezza nella raccolta
  3. Immaginare quali dati servono
  4. Pensare a come organizzarli in un dataset
  5. Pensare, sapere o farsi dire come devono essere elaborati

# Passo 2

## Le fonti dati necessarie

- E' un atto fondamentale per la raccolta dati
- interviste sulla base della scheda (anamnesi, segnalamento)
- verbali di prelievo
- anagrafe Zootecnica
- Esiti diagnostici
- Ecc..

# Passo 3

## Raccogliere i dati

→assicurare:

- Coerenza tra chi raccoglie i dati (training)
- Procedure comuni per la raccolta
- Definire supporto di inserimento dati (excel, access, eccc...)

Prima di passare alla progettazione di un Dataset occorre conoscere lo scopo per il quale siamo chiamati a gestire i dati:

- **Conoscere gli obiettivi:**
  - Diagnosi / refertazione
  - Presenza/Assenza della malattia/condizione
  - Esclusione dell'esistenza di una malattia/condizione
  - Prevalenza della malattia/condizione
  - Studio di "efficacia" di una procedura/trattamento  
(vaccinazioni-selezione genetica-procedure ambientali-trattamenti)
  - Reclutamento casi e controllo dei focolai
  - Individuare trend



Prima di passare alla progettazione di un Dataset occorre conoscere lo scopo per il quale siamo chiamati a gestire i dati:

- **Porsi/porre le domande giuste:**

- Qual è la mole dei dati da gestire?
- La base dati deve essere implementata nel tempo (dati dinamici/statici)?
- I dati devono essere gestiti da più utenti?
- Che tipo di report o analisi devono essere prodotti?
  - Software statistico utilizzato

Si passa alla progettazione del Database



- **Selezione delle informazioni (dati)**  
necessarie al perseguimento degli obiettivi (Es. cod\_az, matrice animale, specie, data, asl, esito..) che devono essere raccolte in modo strutturato (tabelle)

- **Scelta delle fonti dati:** (NSIL, SIEV, BDN, enti esterni, archivi interni, risorse web .....

# Definizioni di Database/dataset

Nelle definizioni di Database alcuni includono il SW altri definiscono questa inclusione impropria

Liceocuneo.it

## Definizione di Database

- Un **Database** può essere definito come un **insieme di dati tra loro correlati**, memorizzati su un supporto di memoria di massa, costituenti un tutt'uno, che possono essere manipolati da più programmi applicativi

Doc.studenti.it

## Database

Archivio elettronico di dati, organizzati in modo integrato attraverso tecniche di modellazione di dati e gestiti sulle memorie di massa dei computer **attraverso appositi software**

Dizionari.corriere.it

Database: “*Insieme di grandi quantità di informazioni tra loro omogenee, organizzato in modo da permettere una rapida ricerca al suo interno; anche, **impropriamente**, il programma applicativo che lo gestisce*”

# Parole chiave per un DATASET

- Memorizzazione
- Archiviazione
- Manipolazione/elaborazione
- Organizzazione/integrazione
- Omogeneità

# Il dataset minimo: Tabella

Ogni insieme di dati che si vuole memorizzare deve essere ricondotto a una o più tabelle – noi tratteremo in questo corso la raccolta “semplice” in una tabella (o file di calcolo).

La tabella per essere funzionale alla gestione dei dati deve avere una **struttura semplice e universale** formata da CAMPI e RIGHE

CAMPI

RIGHE

	num_reg	descrizione m	cod_az	Comune deter	prova	esito
▶	11081200	LATTE DI MAS	048FR053	PATRICA	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11081201	LATTE DI MAS	055FR391	POFI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11081208	LATTE DI MAS	038FR106	FROSINONE	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11085483	LATTE DI MAS	046FR027	PALIANO	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11085484	LATTE DI MAS	071FR016	SERRONE	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11085485	LATTE DI MAS	035FR004	FIUGGI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11086111	LATTE OVINO	034RM001	COLLEFERRO	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11086112	LATTE OVINO	034RM039	COLLEFERRO	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11087564	LATTE OVINO	102RM019	SEGNI	2-4' DDD	INFERIORE AL LIMITE DI QUANTIFICAZIONE
	11087567	LATTE OVINO	102RM161	SEGNI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11087572	LATTE OVINO	102RM048	SEGNI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11088639	LATTE DI MAS	035FR040	FIUGGI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11088645	LATTE DI MAS	035FR032	FIUGGI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11088647	LATTE DI MAS	042FR016	GUARCINO	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11088757	LATTE OVINO	102RM031	SEGNI	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11088759	LATTE OVINO	020RM073	COLLEFERRO	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	11090337	LATTE DI MAS	046FR062	PALIANO	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	12001427	LATTE DI MAS	078FR027	TORRE CAJET	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA
	12003215	LATTE DI MAS	039FR009	FUMONE	2-4' DDD	INFERIORE AL LIMITE DI RILEVABILITA

# Organizzare una Tabella

Righe = record = unità minima di archiviazione univoca

Si tratta dell'unità a cui si riferiscono tutte le variabili e quindi i dati inseriti


Il record può essere rappresentato (a seconda delle esigenze) da diversi livelli di aggregazione:

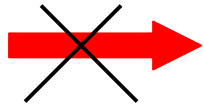
- Campione (aliquota numerata-ceppo-ecc..)
- Individuo (animale-alveare)
- Gruppo (azienda-capannone-partita-apiario)

# Organizzare una Tabella

Righe = record = unità minima di archiviazione univoca

Minore è il livello di aggregazione maggiore è il dettaglio informativo della tabella

Le informazioni possono essere aggregate ad un livello superiore (es. individuo  azienda) anche successivamente

**Attenzione:** se iniziate la tabella con un livello di aggregazione alto (gruppo) non potrete riorganizzare i dati per un livello di aggregazione inferiore (es. azienda  individuo)

# Organizzare una Tabella

Righe = record = unità minima di archiviazione univoca

Inoltre un iniziale livello di aggregazione alto (gruppo) comporta una perdita di capacità di analisi

*Es. **record=azienda** non si può valutare l'effetto dell'età o del sesso o del numero di parti sulla condizione in esame*

Quindi....è sempre consigliato di impostare una tabella le cui righe sono costituite dal livello di aggregazione minimo



# Organizzare una Tabella

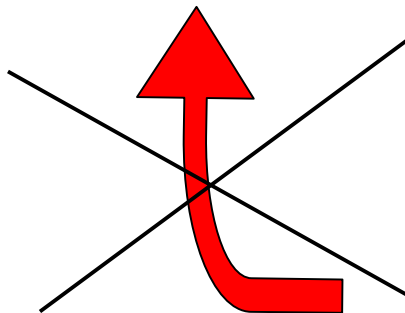
Righe = record = unità minima di archiviazione univoca

Matricola	cod Azienda	indirizzo produttivo azienda	specie	età mesi	Sesso	esito brucellosi
00000001	000XX001	carne	bovino	36	F	negativo
00000002	000XX002	latte	bufalino	30	F	positivo
00000003	000XX003	latte	bovino	24	F	negativo
00000004	000XX001	carne	bovino	48	F	negativo
00000005	000XX002	latte	bufalino	55	F	negativo
00000006	000XX003	latte	bovino	18	F	negativo
00000007	000XX003	latte	bovino	36	M	negativo
00000008	000XX002	latte	bufalino	30	F	negativo
00000009	000XX001	carne	bovino	61	F	negativo
....	....	....	....	....	....	....

Individuo - liv. Aggregazione basso  
(max dettaglio)

OK

gruppo - liv. Aggregazione alto



cod Azienda	indirizzo produttivo	specie	numero capi	stato sanitario brucellosi	Num maschi	Num femmine	≤36 mesi	> 36 mesi
000XX001	carne	bovino	50	negativa	0	3	1	2
000XX002	latte	bufalino	100	positiva	0	3	2	1
000XX003	latte	bovino	150	negativa	1	2	3	0
....	....	....	....	....	....	....	....	....

Perdita di informazioni = sesso, età

# Organizzare una Tabella

Colonna = campo/variabile = attributi del record

Si tratta di tutti gli attributi del record considerati necessari per i nostri scopi.

Come per noi...un nome ed un cognome sono associati ad una descrizione della nostra condizione

Un attributo si riferisce ad un solo campo nel quale viene valorizzato univocamente per ciascun record

..ossia lo stesso dato non deve essere riportato in più campi

# Organizzare una Tabella

Colonna = campo/variabile = attributi del record

Contengono le informazioni necessarie a

1. Identificare univocamente i record (num\_registro-matricola ....)
2. Descrivere i record (anagrafiche: sesso, razza, età...)
3. Definire provenienza (comune, provincia, regione, ASL...)
4. Identificare i gruppi di appartenenza (cod-azi, indirizzo produttivo, trattamento si/no.....)
5. Indicare l'ambito di attività (motivi del prelievo, categorie di rischio, tecniche ecc)
6. tracciare gli esiti

# Organizzare una Tabella

Colonna = campo/variabile = attributi del record

Quando si crea una tabella di raccolta dati i campi sono definiti a priori in funzione dell'obiettivo

Regola: creare solo il numero di campi necessario per l'indagine ed evitare l'ipertrofia del dataset

Deriva da un'altra regola generale degli studi ..ossia..  
raccogliere il set minimo di dati necessari

Adottare il principio della massima parsimonia

# Organizzare una Tabella

Colonna = campo/variabile = attributi del record

Ad eccezione dei campi che identificano univocamente il record TUTTI GLI ALTRI possono essere utilizzati per definire successivi raggruppamenti di record per l'analisi descrittiva o analitica dei risultati

I campi si definiscono quindi anche **variabili di raggruppamento**

Es. Classi d'età, sesso, categorie di rischio, provenienza....

# Organizzare una Tabella

Colonna = campo/variabile = attributi del record

Ad ogni campo deve essere attribuito un nome sintetico ma rappresentativo del suo contenuto

In sostanza evitare nomi astrusi o tecnicismi che possono impedire a voi (in un secondo momento) o ad altri che maneggeranno la tabella (per es. l'OES) di comprenderne il contenuto

E' quindi sempre utile creare e tenersi da parte un documento che descrive il contenuto di ogni campo a mo' di legenda-promemoria

# Organizzare una Tabella

Colonna = campo/variabile = attributi del record

I dati contenuti nei campi devono essere

**Appropriati** rispetto al formato necessario

**Completi** in tutte le celle (no missing)

**Pertinenti** ossia relativi alla variabile a cui si riferiscono

I campi contengono dati omogenei che si riferiscono alle righe e che consentono il raggruppamento dei dati

**...questa NON è una tabella funzionale , non è un dataset.....** Le righe non si riferiscono alla stessa entità e i campi non contengono informazioni univoche (non si possono eseguire filtri, ordinamento, interrogazioni e .....tanto altro)

alveare 1		alveare 2		alveare 3	
mese	% mort	mese	% mort	mese	% mort
gen-15	20	gen-15	8	gen-15	15
feb-15	30	feb-15	5	feb-15	18
mar-15	15	mar-15	6	mar-15	65
mag-15	18	mag-15	15	mag-15	47
giu-15	10	giu-15	26	giu-15	78
alveare 4		alveare 5		alveare 6	
mese	% mort	mese	% mort	mese	% mort
gen-15	18	gen-15	80	gen-15	22
feb-15	25	feb-15	24	feb-15	15
mar-15	36	mar-15	18	mar-15	14
mag-15	21	mag-15	33	mag-15	2
giu-15	7	giu-15	12	giu-15	50

Mortalità &  
alveare in  
campi  
diversi

**Questa invece è una tabella funzionale**  
si tratta degli stessi dati disposti in modo funzionale a  
un dataset

Alveare	mese	% mort
1	gen-15	20
1	feb-15	30
1	mar-15	15
1	mag-15	18
1	giu-15	10
2	gen-15	8
2	feb-15	5
2	mar-15	6
2	mag-15	15
2	giu-15	26
3	gen-15	15
3	feb-15	18
3	mar-15	65
3	mag-15	47
3	giu-15	78
4	gen-15	18
4	feb-15	25
4	mar-15	36
4	mag-15	21
4	giu-15	7
5	gen-15	80
5	feb-15	24
5	mar-15	18
5	mag-15	33
5	giu-15	12
6	gen-15	22
6	feb-15	15
6	mar-15	14
6	mag-15	2
6	giu-15	50



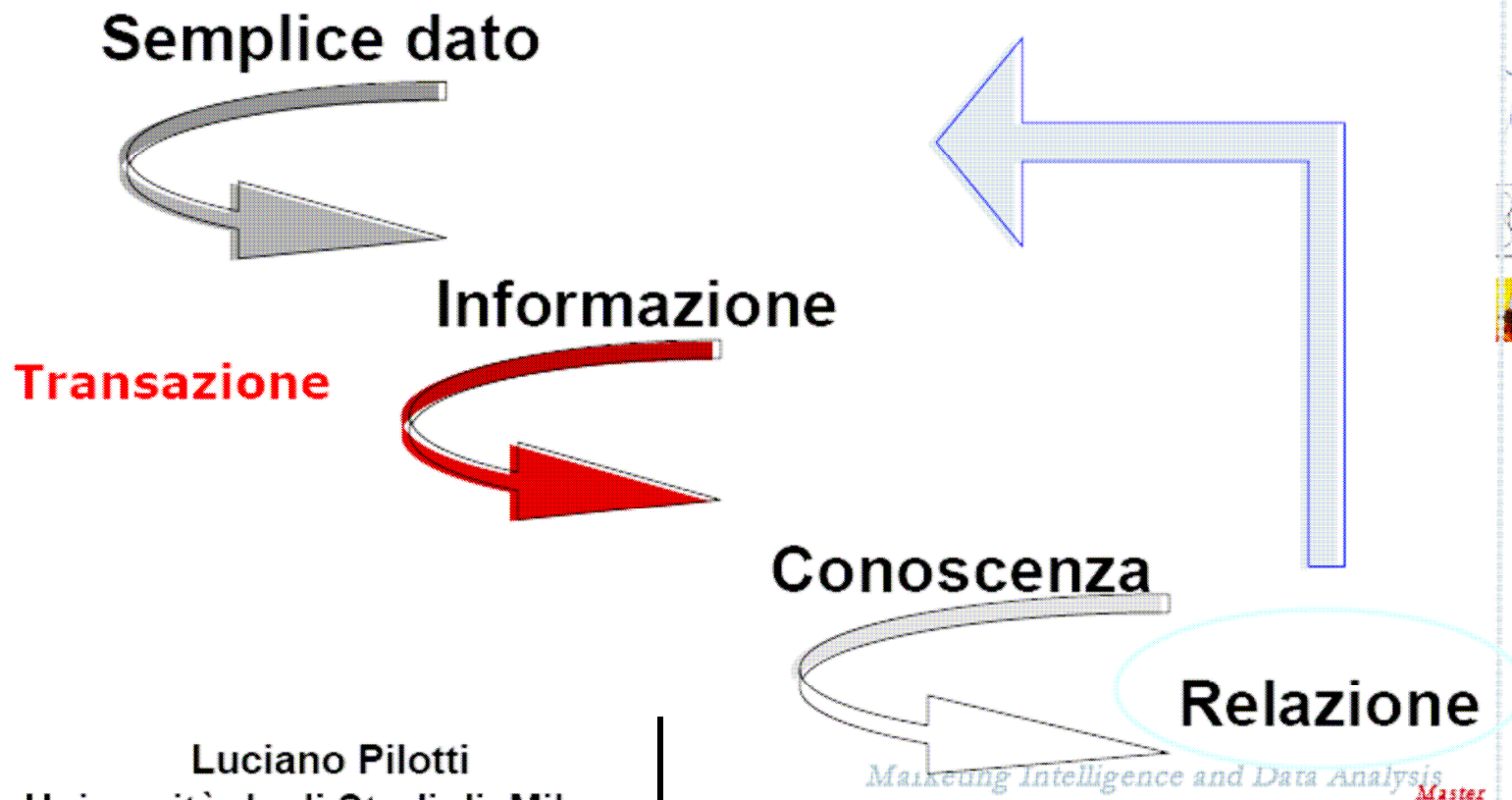
# Esempio di creazione di una tabella

- **Obiettivo:** stimare la sieroprevalenza per West Nile in cavalli del padule del Fucecchio per coorte di nascita per verificare la circolazione virale successiva al 1998
- Righe-record=individui
- Campi-variabili necessari: identificativo record; data prelievo, anno nascita, comune,esito
- Formati: *identificativo* testo; *data* europeo; *anno* numerico, *comune* testo, *esito* testo (codifica: positivo/negativo)

N° REGISTRO	DATA PRELIEVO	COMUNE AZIENDA	RAZZA	NOME SOGGETTO	ANNO NASCITA	TECNICA	SESSO	ESITO
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	OKIE	2001	ELISA SIEROLOGICO IGG	C	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	DILLON	2000	ELISA SIEROLOGICO IGG	C	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	ISPANICO	2003	ELISA SIEROLOGICO IGG	M	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	TARRIS	2003	ELISA SIEROLOGICO IGG	C	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	T.CROSS	2000	ELISA SIEROLOGICO IGG	F	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	ICE COFFEE	2001	ELISA SIEROLOGICO IGG	C	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	CHARLIE	2003	ELISA SIEROLOGICO IGG	M	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	EISENOWER	2004	ELISA SIEROLOGICO IGG	M	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	DUN	2002	ELISA SIEROLOGICO IGG	F	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	ROSINA	2003	ELISA SIEROLOGICO IGG	F	negativo
18009	27/04/2006	S. CROCE S/ARNO (PI)	CAVALLO	WISE	2001	ELISA SIEROLOGICO IGG	F	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	REBEL O' LENA	2002	ELISA SIEROLOGICO IGG	C	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	RD MAJESTIC CATALISE	2003	ELISA SIEROLOGICO IGG	F	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	ARC KEITH ENTERPRISE	1996	ELISA SIEROLOGICO IGG	C	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	FROZEN SUNDAY	2001	ELISA SIEROLOGICO IGG	M	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	TONNY ALL DUN	2003	ELISA SIEROLOGICO IGG	M	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	STRATOSPANISH	2003	ELISA SIEROLOGICO IGG	C	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	GOLDEN HAIDER OAK	2002	ELISA SIEROLOGICO IGG	F	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	AR MR SPANISH	2003	ELISA SIEROLOGICO IGG	C	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	SPANISH SENORITA	2003	ELISA SIEROLOGICO IGG	F	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	AR DOCTOR JEKILL	2003	ELISA SIEROLOGICO IGG	C	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	FRECK LARCK OAK	2002	ELISA SIEROLOGICO IGG	C	negativo
19181	11/05/2006	FUCECCHIO (FI)	CAVALLO	JACK	2003	ELISA SIEROLOGICO IGG	C	positivo
19252	12/05/2006	MASSA E COZZILE (PT)	ASINO	ALBERTO	2005	ELISA SIEROLOGICO IGG	M	negativo
19252	12/05/2006	MASSA E COZZILE (PT)	ASINO	FRANCIA	2000	ELISA SIEROLOGICO IGG	F	negativo
19252	12/05/2006	MASSA E COZZILE (PT)	ASINO	GAVINO	2000	ELISA SIEROLOGICO IGG	M	negativo
19252	12/05/2006	MASSA E COZZILE (PT)	ASINO	ROSSA BETTINA	1988	ELISA SIEROLOGICO IGG	F	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	FAST RUNNER	2002	ELISA SIEROLOGICO IGG	M	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	INTERPOL SAS	2004	ELISA SIEROLOGICO IGG	M	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	ISAGORA	2004	ELISA SIEROLOGICO IGG	F	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	CORTE LOD	1999	ELISA SIEROLOGICO IGG	F	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	IOLANDO SF	2004	ELISA SIEROLOGICO IGG	M	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	IMPETUOSI RAL	2004	ELISA SIEROLOGICO IGG	F	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	INDIO DELL' EST	2004	ELISA SIEROLOGICO IGG	M	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	GAZZA D'ARMO	2003	ELISA SIEROLOGICO IGG	F	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	FAIR WAY	2002	ELISA SIEROLOGICO IGG	C	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	DEL SANTO	2000	ELISA SIEROLOGICO IGG	C	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	CASANOVA. RAL	1999	ELISA SIEROLOGICO IGG	M	positivo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	CARRACCI	1999	ELISA SIEROLOGICO IGG	M	negativo
19822	16/05/2006	FUCECCHIO(FI)	CAVALLO	DANUBIO RAL	2000	ELISA SIEROLOGICO IGG	M	negativo

Omogeneità e stabilità

# La catena del valore informativo di conoscenza e relazione



Luciano Pilotti  
Università degli Studi di Milano  
[luciano.pilotti@unimi.it](mailto:luciano.pilotti@unimi.it)

SPSS



Teradata  
a division of EMC

THE  
TECHNOLOGY  
PARTNERS

VALUE LAB  
Marketing - Sales - Training Solutions

IBM  
Business Value Analytics

Marketing Intelligence and Data Analysis  
Master

# Quindi cos'è un dato?

- E' un elemento complesso che deve avere caratteristiche, unità di misura e formato pre-determinate
- Deve essere **Appropriato** rispetto al formato necessario a descrivere la variabile (campo), **Completo e Pertinente** ossia rappresentare la variabile a cui si riferisce
- DEVE ESSERE STABILE ed omogeneo nel dataset per mantenere inalterata l'informazione derivante sia dal singolo dato che dal loro insieme

# Temi in Sospeso

Tipi di variabili e loro caratteristiche (Francesca Iaconi)

Fonti dati e dataset (Massimo Mari)

Come alimentare un dataset, codificare e transcodificare i dati (Marcello Sala)

Elaborazioni di Base – raggruppamenti (Marcello sala)