



Informazione Statistica Informazione Geografica e Banche Dati Online: Sviluppi Metodologici ed Applicativi

Maurizio Vichi

Presidente Società Italiana di Statistica

e

Dpt. Statistica Probabilità e Statistiche Applicate

Università "La Sapienza" di Roma

em: maurizio.vichi@uniroma1.it



DSPSA



SAPIENZA
UNIVERSITÀ DI ROMA

Dipartimento di Statistica, Probabilità
e Statistiche Applicate





La Statistica: tra Diffusione e Comunicazione

La Statistica tra cattiva immagine e grande successo in Italia (ma anche all'estero)

- La funzione della Statistica è in generale di acquisire informazione “*soggettiva*” (ovvero dati su singoli individui) e restituire informazione “*oggettiva*”, (conoscenza) legata alla collettività;
- Con la *Società dell'Informazione* (basata su radio, televisione, telefonia e internet), l'informazione soggettiva è enormemente aumentata;
- La percezione pubblica della statistica è in parte negativa. Da che dipende?
- I media spesso usano informazione soggettiva e la pubblicano come oggettiva;
- La gente ha una percezione di dipendenza della Statistica ufficiale dal Governo. Necessità di favorire l'indipendenza e l'autorevolezza (credibilità);

Che cosa si può fare per migliorare l'immagine pubblica della Statistica.

Due punti in discussione nel workshop: la condivisione e la diffusione dell'informazione statistica

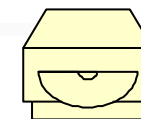
- Organizzare l'informazione statistica online per renderla universalmente accessibile, integrabile e facile da leggere ed interpretare;
- Completare la modernizzazione della Statistica. Il Datawarehouse non è un problema informatico ma è statistico;

La comunicazione dell'informazione statistica. La Statistica non si deve fermare alle tabelle e grafici ma deve comunicare l'informazione rilevante dandone l'interpretazione oggettiva, ovvero la descrizione ed l'inferenza su ciò che è stato osservato.

- Analizziamo gli sviluppi metodologici ed applicativi che sono necessari per migliorare la condivisione e la diffusione dell'informazione statistica

SIS Sistema Informativo Statistico online

Costruzione di un Sistema Informativo Statistico online

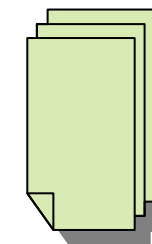


Quali sono le caratteristiche generali di un SIS online



Per gli aspetti di Organizzazione

- + Accessibilità dei dati e dei metadati online;
- + Organizzazione dei dati per l'analisi statistica;
- + Integrazione tra dati (modularità) mediante internet



Per gli aspetti di Comunicazione

- + Ridurre i tempi di pubblicazione dei dati;
- + Trasparenza e certificazione (qualità dei dati);
- + Predisposizione di strumenti statistici di elaborazione;
- + Interpretazione oggettiva delle elaborazioni (information mining).





Basi Dati Online

Analizziamo tre rilevanti DB Online per esaminare le loro caratteristiche

- UN data (<http://data.un.org/>)
 - La Divisione Statistica delle Nazioni Unite (UNSD) del DESA ha lanciato un nuovo servizio internet che permette di trovare e scaricare una varietà di dati statistici delle agenzie delle Nazioni Unite.
 - Databases
 - Education [UNESCO UIS Data](#)
 - Employment [ILO Data](#)
 - Energy [UNSD Energy Statistics](#)
 - Environment [UNFCCC Greenhouse Gas Inventory](#)
 - Food and Agriculture [FAO Data](#) , [FAO Fertilizer Data](#)
 - Health [WHO Data](#) , HIV/AIDS [UNAIDS Data](#)
 - Human Development [UNDP Human Development Report](#)
 - Indicator databases [UNSD Key Global Indicators](#), [UNSD Gender Info 2007](#) , [UNSD Indicators on Women and Men](#) ,[UNSD Millennium Development Goals](#) , [UNICEF The State of the World's Children 2009](#)
 - Industry [UNSD Industrial Commodity Statistics](#)
 - Information and Communication Technology [TU Data](#)
 - National Accounts [UNSD Official Country Data](#) , [UNSD Estimates of Main Aggregates](#)
 - Population [UNSD Demographic Statistics](#) , [UNPD World Population Prospects](#)
 - Refugees [UNHCR Statistical Yearbook](#)
 - Trade [UNSD Commodity Trade Statistics](#)
 - Tourism [UNWTO Data](#)



Basi Dati Online 2

- **Eurostat** (<http://epp.eurostat.ec.europa.eu>)
- Disponibile sul sito ed esplorata con un albero
- **Euroindicators database**
- **Economy and finance** [Main economic indicators](#), [National accounts \(incl.GDP\)](#), [Government statistics](#), [Financial accounts](#), [Exchange rates](#), [Interest rates](#), [Monetary and other financial statistics](#), [Prices](#), [Balance of payments](#)
- **Population and social conditions** [Population](#), [Health](#), [Education and training](#), [Labour market](#), [Living conditions and welfare](#), [Crime and criminal justice](#),
- **Industry, trade and services** [Short-term business statistics](#), [Structural business statistics](#), [Information society statistics](#), [Tourism](#), [Statistics on the production of manufactured goods](#),
- **General and regional statistics** [Regions](#), [Urban audit](#), [Non EU countries](#),
- **Agriculture and fisheries** [Agriculture Forestry](#), [Fisheries](#), [Food](#)
- **External trade** [External trade](#)
- **Transport**
- **Environment and energy** [Environment](#), [Energy](#)
- **Science and technology** [Science and technology](#)



Basi Dati Online 3

- **Statistics Portal [OECD.Stat Extracts](#)**

- General Statistics
- Agriculture and Fisheries
- Demography and Population
- Development
- Economic Projections, Education and Training
- Environment
- Finance
- Globalisation
- Health
- Industry and Services
- International Trade and Balance of Payments
- Labour
- Monthly Economic Indicators
- National Accounts
- Prices and Purchasing Power Parities
- Productivity
- Public Sector, Taxation and Market Regulation
- Regional Statistics
- Science, Technology and Patents
- Social and Welfare Statistics
- Non-member Economies
- Others



Vista Univariata o tabellare dei DB

- In UNDATA sono riportate le serie storiche degli indicatori ed ha un motore di mappe tematiche;
- In EUROSTAT c'è una applet Java che permette di costruire tabelle;
- In OECD Portal c'è eXplorer che è uno strumento per costruire tabelle e grafici e fare confronti territoriali;
- In un SIS i dati sono logicamente strutturati in uno o più "CUBI" di informazioni:
- **unità territoriali** x **indicatori** x **tempi**, che sono le dimensioni tipiche di ogni fenomeno statistico con struttura complessa.

Struttura Multiway dei Dati organizzati in

uno o più data-set a tre vie (Three-way Data Set) **X**

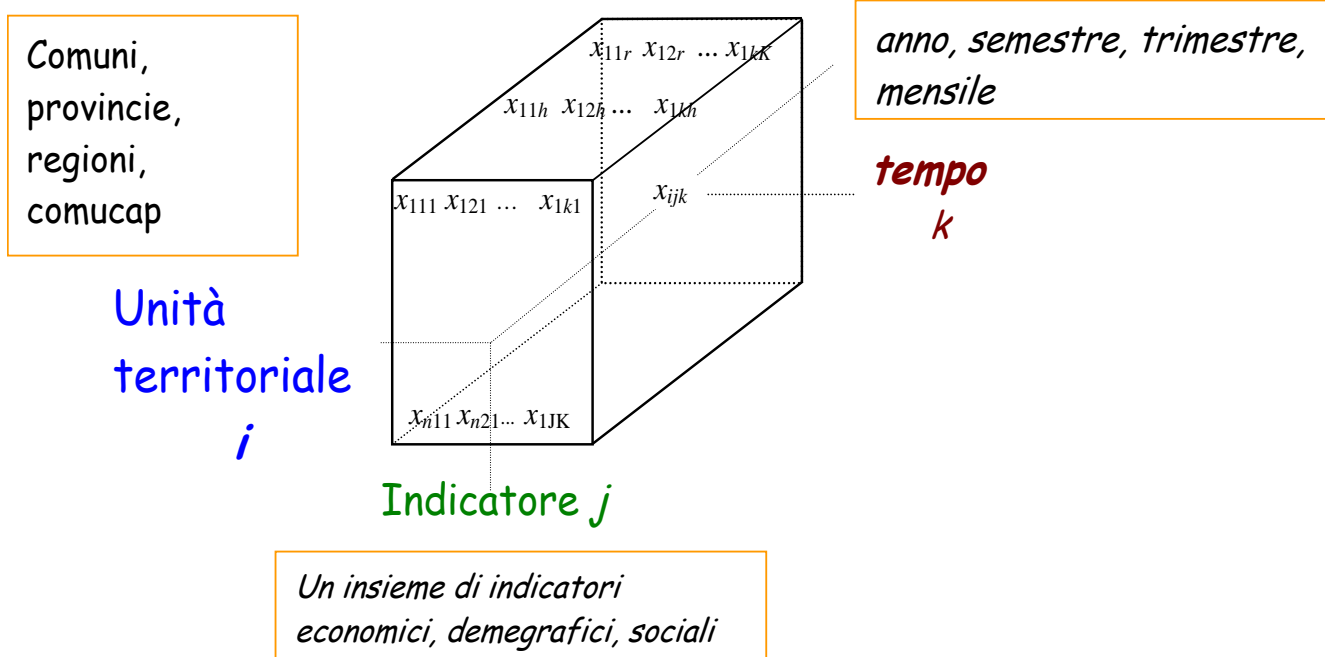
un insieme X di $I \times J \times K$ attributi relativi a:

J **Indicatori** (variabili) misurati, (osservati, stimati) su

I **Unità Territoriali** (oggetti geografici, aggregazioni amministrative) osservati,

K **Occasioni** (tempi, diverse fonti di dati)

Si ha un Cubo X o 3-Way Array





Analisi statistica dei dati di un SIS

I dati presentano tre diverse tipologie di relazione di cui si deve tener conto

❖ **RELAZIONE TERRITORIALI**, tra le unità territoriali rispetto agli indicatori in un istante di tempo (*correlazione spaziale*, (dis)similarità fra oggetti spaziali, vincoli di contiguità, ecc.);

TIPO DI ANALISI: Statistica per il territorio (Spaziale) ed uso dei GIS; Indici di associazione spaziale, regressione e segmentazione vincolata al territorio.

❖ **RELAZIONE MULTIVARIATE**, ovvero tra gli indicatori sulla base di unità territoriali osservate in un istante di tempo (associazione e correlazione fra variabili);

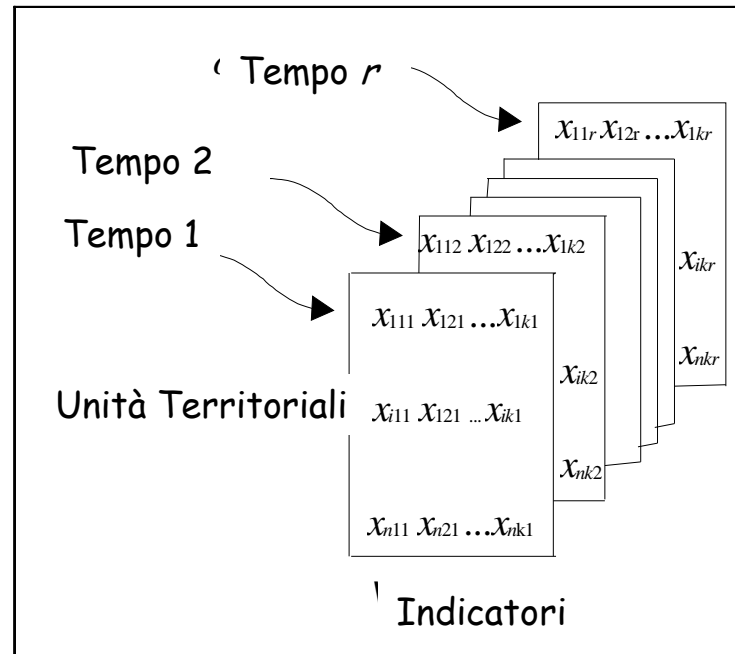
TIPO DI ANALISI : Statistica Multivariata. Costruzione di indicatori complessi, segmentazione del mercato. Regressione tra indicatori

❖ **RELAZIONI TEMPORALI**, ovvero in ciascun indicatore che si riferiscono a indicatori e/o unità territoriali (auto-correlazione temporale);

TIPO DI ANALISI: Analisi delle Serie Storiche ed uso della modellistica per le Previsioni.

Prima vista del SIS CONFRONTI TERRITORIALI

(sezionando verticalmente X)

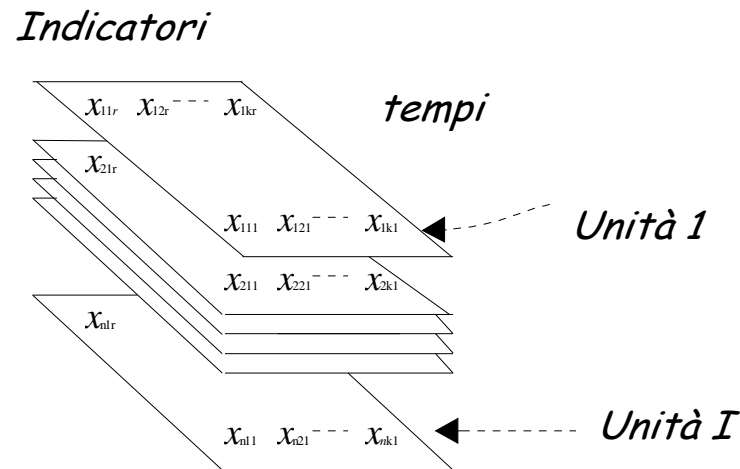


Unità territoriali multivariate

- Confronti territoriali tra indicatori in uno stesso istante di tempo (Analisi cross-section);
- Analisi delle posizioni relative tra le unità territoriali su indicatori o spazi fattoriali (indicatori composti);
- Segmentazione del territorio in aree omogenee.

Seconda Vista PREVISIONI STATISTICHE

(sezionando orizzontalmente X) n Serie Storiche Multiple

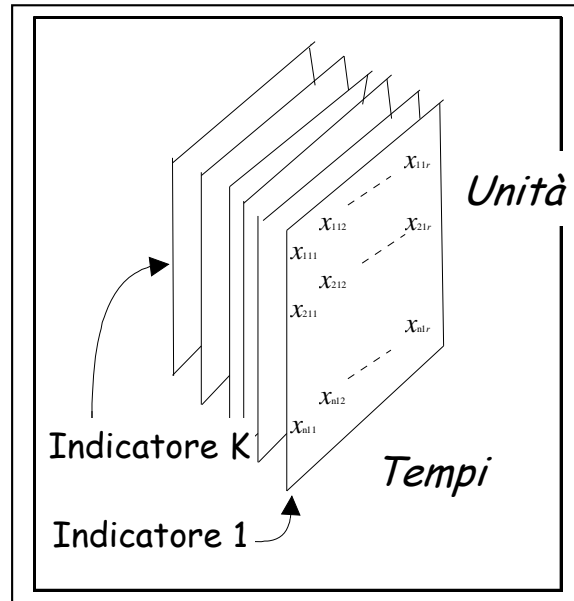


Un insieme di dati Panel

- Confronti tra n Serie Storiche multivariate (Panel di dati)
- Analisi delle "storie" tra più unità territoriali sulla base di un set di indicatori;
- Previsioni su un territorio sulla base di un set di indicatori;
- Analisi di convergenza e divergenza territoriale;

Terza vista **CONVERGENZA** di **INDICATORI**

(sezionando lateralmente x) un insieme di K **Serie Storiche Multiple**



Un insieme di dati Panel

- Confronti "storici" tra indicatori per le unità territoriali;
- Previsioni di indicatori in un territorio;
- Analisi di convergenza e divergenza degli indicatori sul territorio;

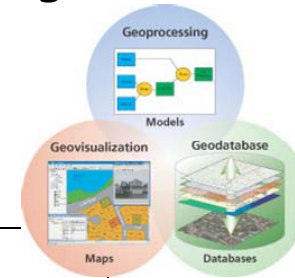
Altra modalità "geografica" di vedere i dati

GIS - Geographical Information System

- Un GIS è un sistema orientato alla Geografia, che permette, la gestione, l'analisi e la visualizzazione di informazioni legate al territorio

Diversi modi di vedere un GIS

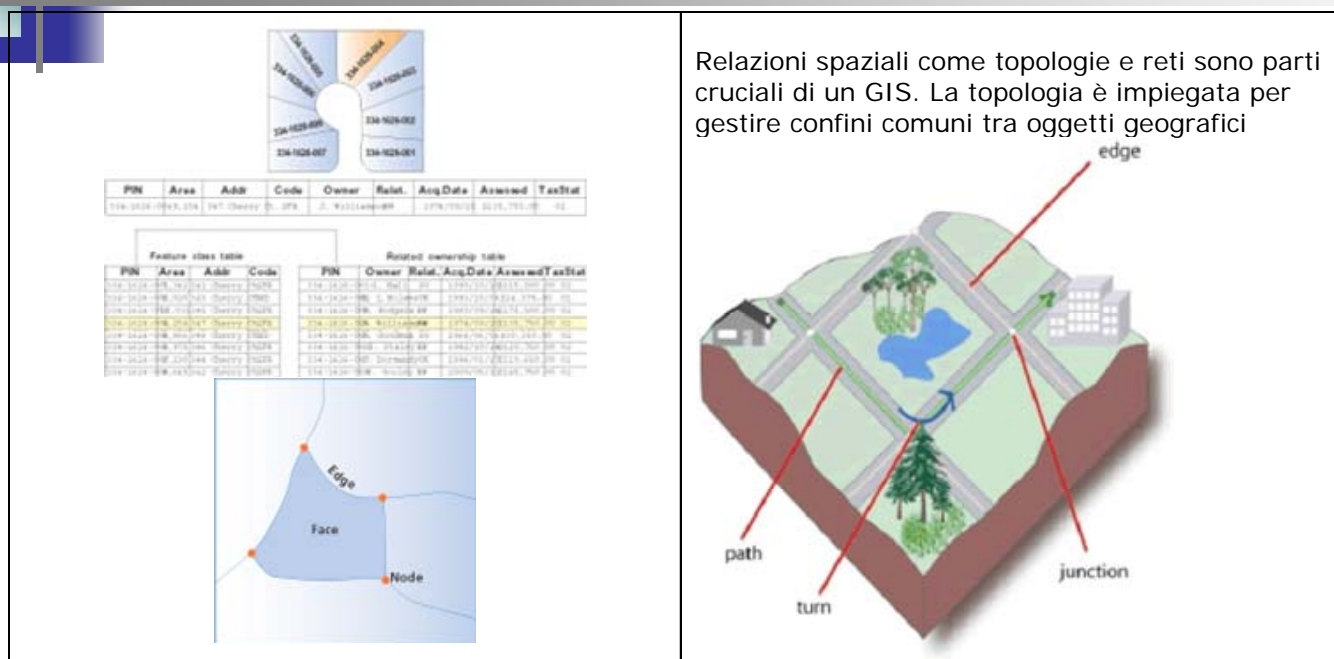
La vista del Database: dati geografici



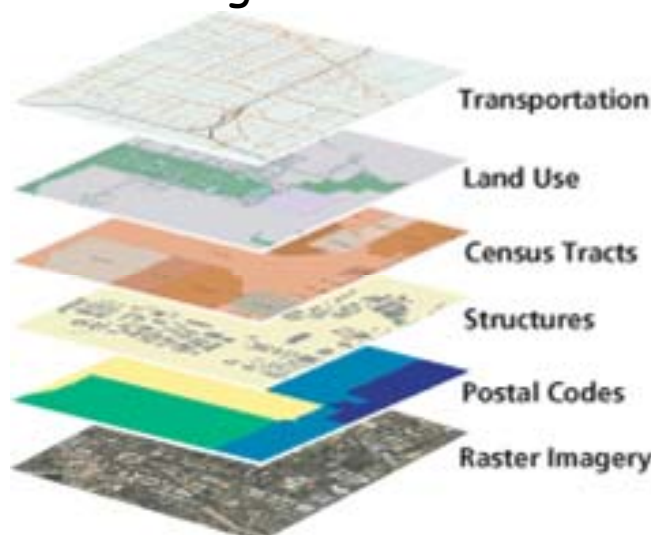
<p>Collezioni ordinate di caratteristiche vettoriali (insiemi di punti, linee e poligoni)</p>	<p>modelli digitali di elevazione del territorio</p>																									
<p>Collegamenti tra oggetti geografici</p>	<p>Terreni e altre superfici</p>																									
<table border="1"><thead><tr><th>From</th><th>To</th><th>Type</th><th>Direction</th><th>Distance</th></tr></thead><tbody><tr><td>A</td><td>B</td><td>Direction</td><td>N7E</td><td>-</td></tr><tr><td>A</td><td>C</td><td>Angle</td><td>300.56</td><td>201</td></tr><tr><td>A</td><td>C</td><td>Distance</td><td>-</td><td>20.50</td></tr></tbody></table> <p>Insiemi di dati di ricerche geografiche</p>	From	To	Type	Direction	Distance	A	B	Direction	N7E	-	A	C	Angle	300.56	201	A	C	Distance	-	20.50	<p>Addresses</p> <table border="1"><tbody><tr><td>3350 45th Ave NE</td></tr><tr><td>3383 30th Ave NE</td></tr><tr><td>2459 Country Rd. 9 NE</td></tr><tr><td>3383 30th A ve N E</td></tr><tr><td>2458 C ountry Rd. 9 N E</td></tr></tbody></table> <p>altre informazioni geografiche</p>	3350 45th Ave NE	3383 30th Ave NE	2459 Country Rd. 9 NE	3383 30th A ve N E	2458 C ountry Rd. 9 N E
From	To	Type	Direction	Distance																						
A	B	Direction	N7E	-																						
A	C	Angle	300.56	201																						
A	C	Distance	-	20.50																						
3350 45th Ave NE																										
3383 30th Ave NE																										
2459 Country Rd. 9 NE																										
3383 30th A ve N E																										
2458 C ountry Rd. 9 N E																										

Secondo e Terzo modo di vedere un GIS

Dati di relazione tra gli oggetti spaziali e i dati di flusso.



Organizzazione tematica dei dati (**livelli tematici**)





Integrazione tra GIS e SIS

Tre tipologie di vista dei dati

Visualizzazione cartografica (tipica GIS)

Gli indicatori che definiscono una serie territoriale originano un cartogramma.

Visualizzazione tabellare (tipica SIS)

Tabelle unità territoriali x indicatori;

Tabelle unità territoriali x tempi;

Tabelle indicatori x tempi

Visualizza grafica (tipica SIS e GIS + cartografia)

Per variabili qualitative

1 indicatore: grafico a colonne

1 indicatore x anni: grafico a colonne contrapposte

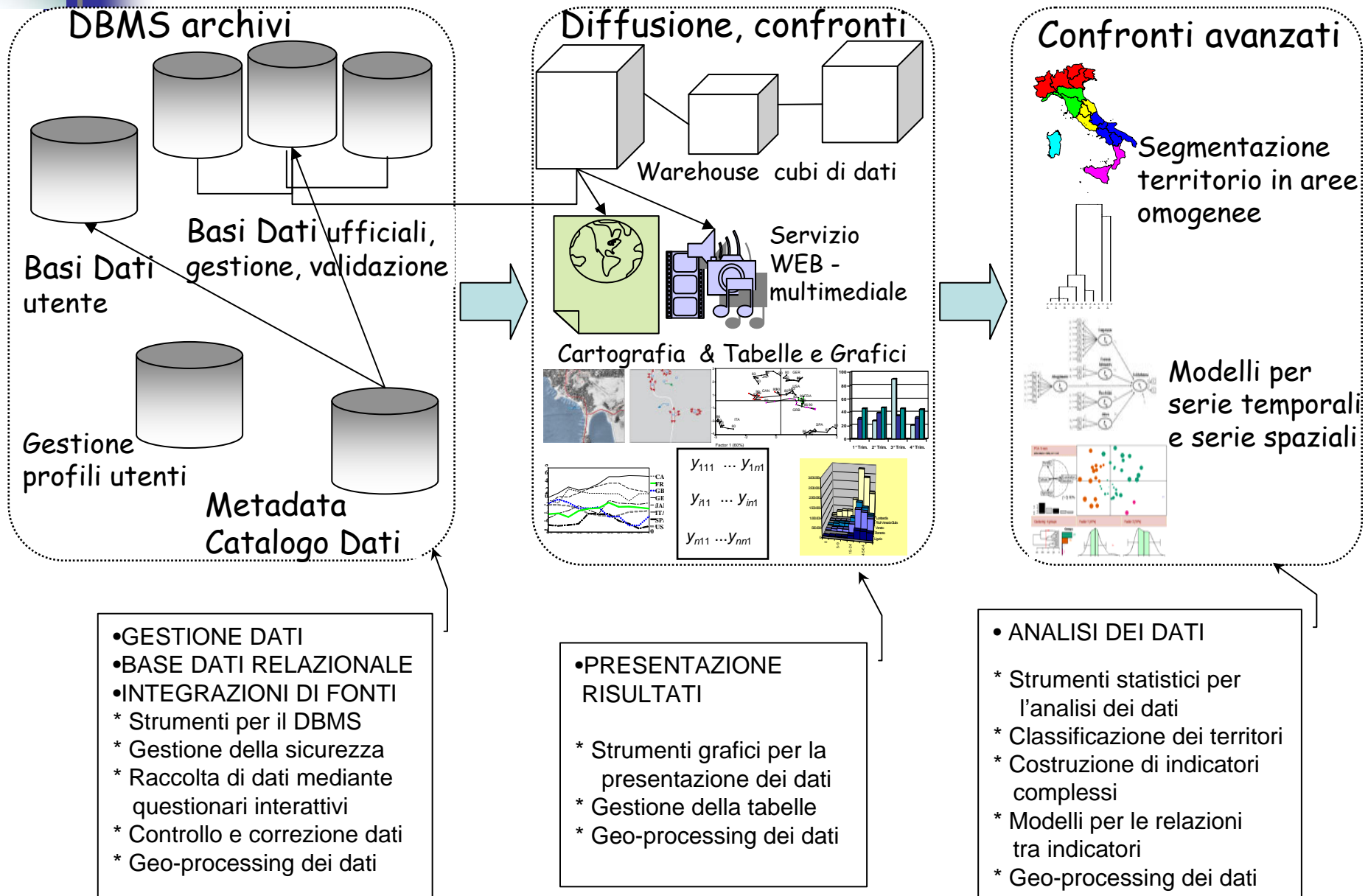
Per variabili quantitative

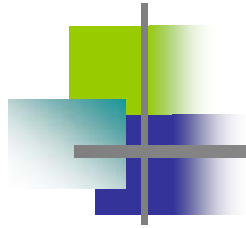
Istogramma

1 indicatore x anni: grafico cartesiano

m indicatori x anni: grafici cartesiani a confronto

Integrazione tra GIS e SIS: Geo Web Starter





Software Statistico (Open Source)

DESC FREQ PROB ANOVA1 ANOVA+ EXPER SLR MLR LOG LOGIT PROBIT GLM ANCOVA NONPAR LOGLIN TIME SURV PCA FACT CCA CA DISCR CLUST

	DESC	FREQ	PROB	ANOVA1	ANOVA+	EXPER	SLR	MLR	LOG	LOGIT	PROBIT	GLM	ANCOVA	NONPAR	LOGLIN	TIME	SURV	PCA	FACT	CCA	CA	DISCR	CLUST	
ADE 4	
DATAPLOT		
EASYREG							
GRETL						
INSTAT +							
MACANOVA	
MATRIXER								
MICROSIRIS	
OPENSTAT	
R
TANAGRA	
VISTA	
WINIDAMS	



Software Online

- **SOCR** (Statistics Online Computational Resource)
- (<http://www.socr.ucla.edu/>) Database di risorse computazionali online
 - include:
 - [Distributions](#) (interactive graphs and calculators),
 - [Experiments](#) (virtual computer-generated analogs of popular games and processes),
 - [Analyses](#) (collection of common web-accessible tools for statistical data analysis),
 - [Games](#) (interfaces and simulations to real-life processes),
 - [Modeler](#) (tools for distribution, polynomial and spectral model-fitting and simulation),
 - [Graphs, Plots and Charts](#) (comprehensive web-based tools for exploratory data analysis),
 - [Additional Tools](#) (other statistical tools and resources),
 - [SOCR Wiki](#) (collaborative Wiki resource),
 - [Educational Materials and Hands-on Activities](#) (varieties of SOCR educational materials),
 - [SOCR Statistical Consulting](#)
 - [Statistical Computing Libraries](#)