



*A Scuola di*  
**OPENCOESIONE**

**COMPRENDERE E RAFFINARE I DATI**

# COMPRENDERE I DATI PER USARLI AL MEGLIO

## COMPRENDERE I DATI

I dati che avete a disposizione potrebbero dirvi tante cose del fenomeno che state osservando.



## COMPRENDERE I DATI

Per farli “parlare”, cercate di capirli a pieno.

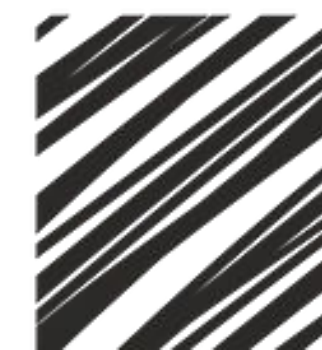
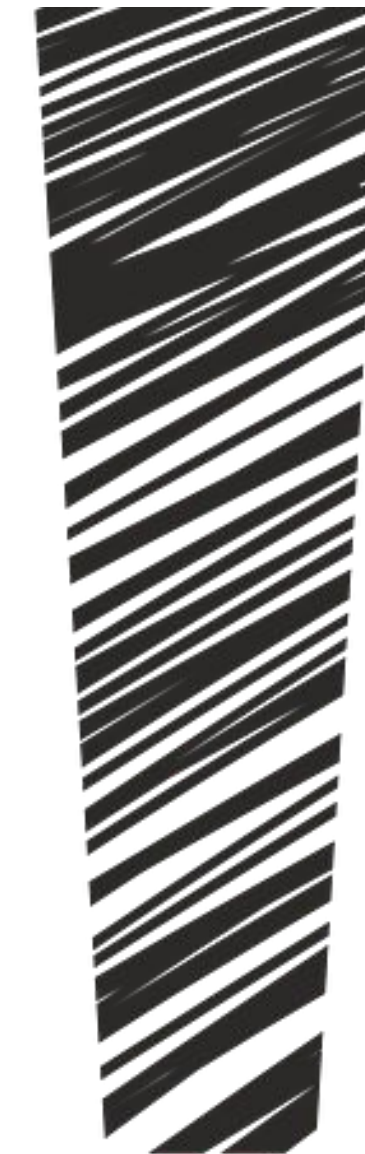
Fatevi queste domande:

- Che tipo di informazioni definiscono?
- Quali sono le loro caratteristiche principali?
- Che impatto hanno sulla realtà?

## COMPRENDERE I DATI

E' importante:

- capire bene la natura dei dati
- capire come analizzarli  
attraverso alcune semplici  
operazioni



# **LE CARATTERISTICHE DEI VOSTRI DATI: GEOGRAFICHE, TEMPORALI, DESCRITTIVE**



## DATI GEOGRAFICI



I dati che state utilizzando per i vostri lavori potrebbero contenere informazioni di tipo geografico.

Dirvi, cioè, qual è la misura di un fenomeno su un determinato territorio.

**DATI GEOGRAFICI**

- A quale territorio sono riferiti? Nazionale, regionale, provinciale o comunale?
- Qual è il nesso tra questo territorio e il progetto da voi scelto su OpenCoesione e che state approfondendo?
- Quali informazioni aggiuntive potreste raccogliere analizzando le informazioni geografiche dei vostri dati?



I dati possono permettere di analizzare l'andamento di un fenomeno nel tempo e misurarne le variazioni (per esempio da un mese all'altro o da un anno all'altro).



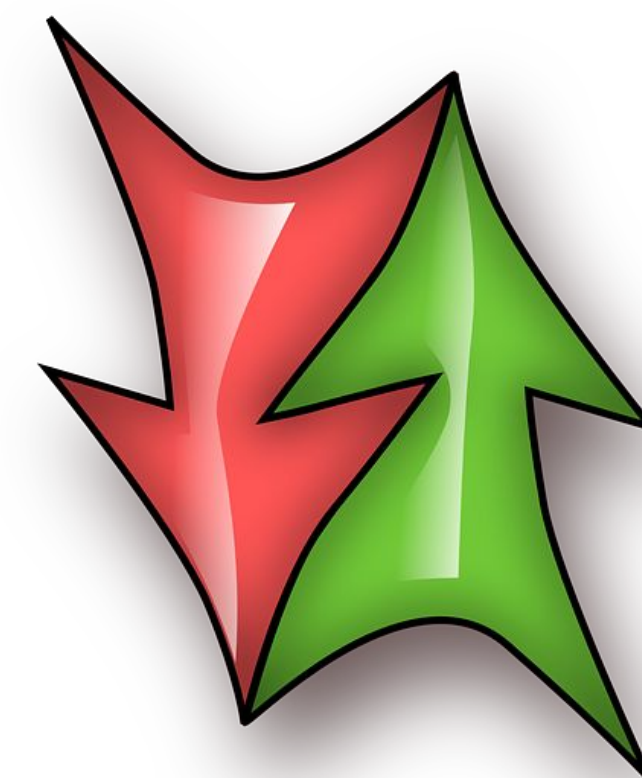
## SERIE STORICHE: ESEMPI



La raccolta differenziata dei rifiuti è aumentata, diminuita o rimasta stabile?



Il numero di turisti sul un territorio è costante di anno o anno o ci sono delle variazioni?



Il fenomeno in esame è in crescita o in decrescita (per esempio, di mese in mese)?

# SERIE STORICHE: ESEMPI (ISTAT)

ENGLISH HOME

**Istat** Istituto Nazionale di Statistica

POPOLAZIONE E FAMIGLIE SOCIETÀ E ISTITUZIONI ISTRUZIONE E LAVORO ECONOMIA AMBIENTE E TERRITORIO CERCA NEL SITO Statistiche A-Z Glossario

HOME > DATI ANALISI E PRODOTTI > BANCHE DATI [ENGLISH]

## BANCHE DATI E SISTEMI INFORMATIVI

**DATI ANALISI E PRODOTTI**

**BANCHE DATI**

StatBase

**TAVOLE DI DATI**

Bollettino mensile di statistica on line

**MICRODATI**

Riconoscimento  
Normativa

**COMUNICATI STAMPA**

**AUDIZIONI E CONTRIBUTI AL PARLAMENTO**

**PUBBLICAZIONI**

Rivista statistica ufficiale  
Istat working papers

**CALCOLATORI**

Contanomi  
Rivalutazioni  
Soglia di povertà  
Matrimoni in Italia

**OPEN DATA**

**STATISTICHE SMART DA BIG DATA**

**STATISTICHE A-Z PAROLE**

L'accesso, la navigazione e il download di tutti i dati e le informazioni sono gratuiti.

**BANCHE DATI GENERALI**

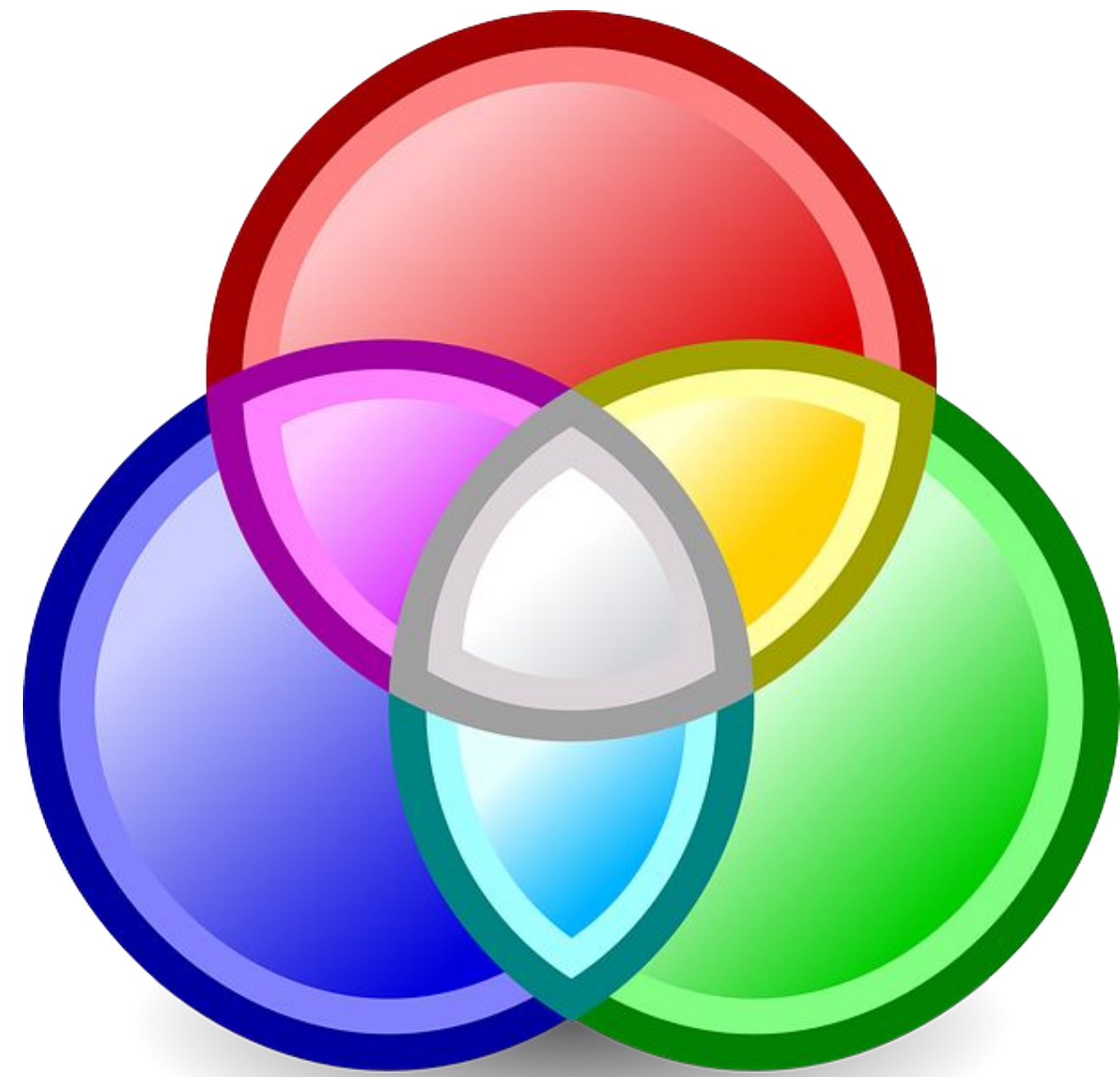
- IstatData**  
 IstatData è il nuovo data warehouse di diffusione (in sostituzione di I.Stat) che permette l'accesso ai dati aggregati prodotti dall'Istat. [Leggi tutte le informazioni relative a IstatData.](#)
- I.STAT**  
 La banca dati completa per gli esperti. Un [web service](#) consente l'interrogazione diretta machine-to-machine
- STAT.BASE**  
 L'accesso diretto ai dati più richiesti
- SERIE STORICHE**  
 Oltre 1.500 serie storiche per raccontare i mutamenti ambientali, sociali ed economici di cui l'Italia è stata protagonista dalla costituzione dello Stato unitario

Per capire quanto possano essere importanti le informazioni contenute nelle serie storiche, andate anche a curiosare nelle banche dati Istat...



## DATI QUALITATIVI

Si tratta di dati che definiscono le qualità e le caratteristiche intrinseche di uno o più oggetti.



## DATI QUALITATIVI: ESEMPIO



Un esempio: i prodotti tipici trentini (da OpenData Trentino, <http://dati.trentino.it/dataset/prodotti-tradizionali-trentini>). Ciascun prodotto viene definito per tipologia, luogo di produzione, link di approfondimento. Si tratta di qualità aggiuntive del dato disponibili per il fruitore.

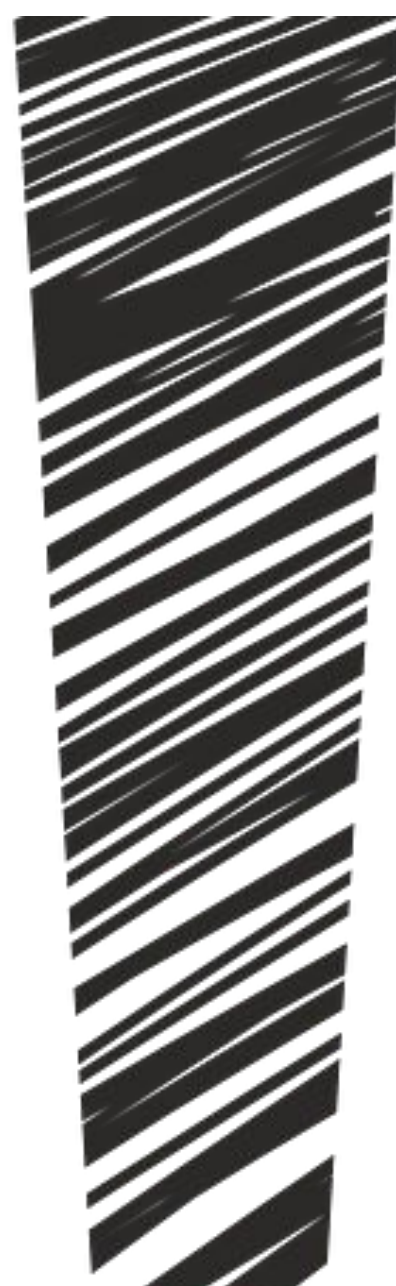


Quindi chiedetevi:  
quali caratteristiche hanno i dati  
che avete raccolto o state  
raccolgendo?  
Quali sono quelle che possono  
interessarvi di più?



# PULIRE I DATI

## PULIRE I DATI

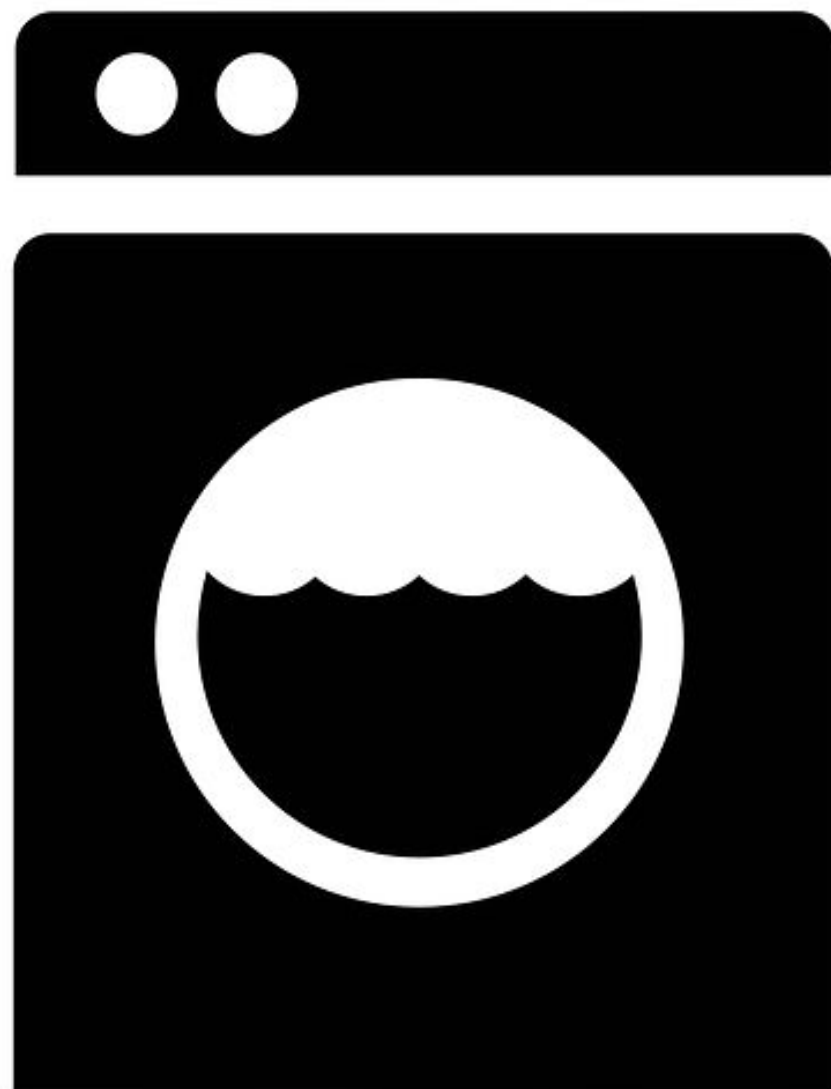


Prima di analizzare i dati è necessario fare alcune verifiche per renderli **coerenti** e sfruttarne al meglio tutte le **caratteristiche**.

Esistono metodi specifici anche in questo caso.

Vediamo quali...

## PULIRE I DATI

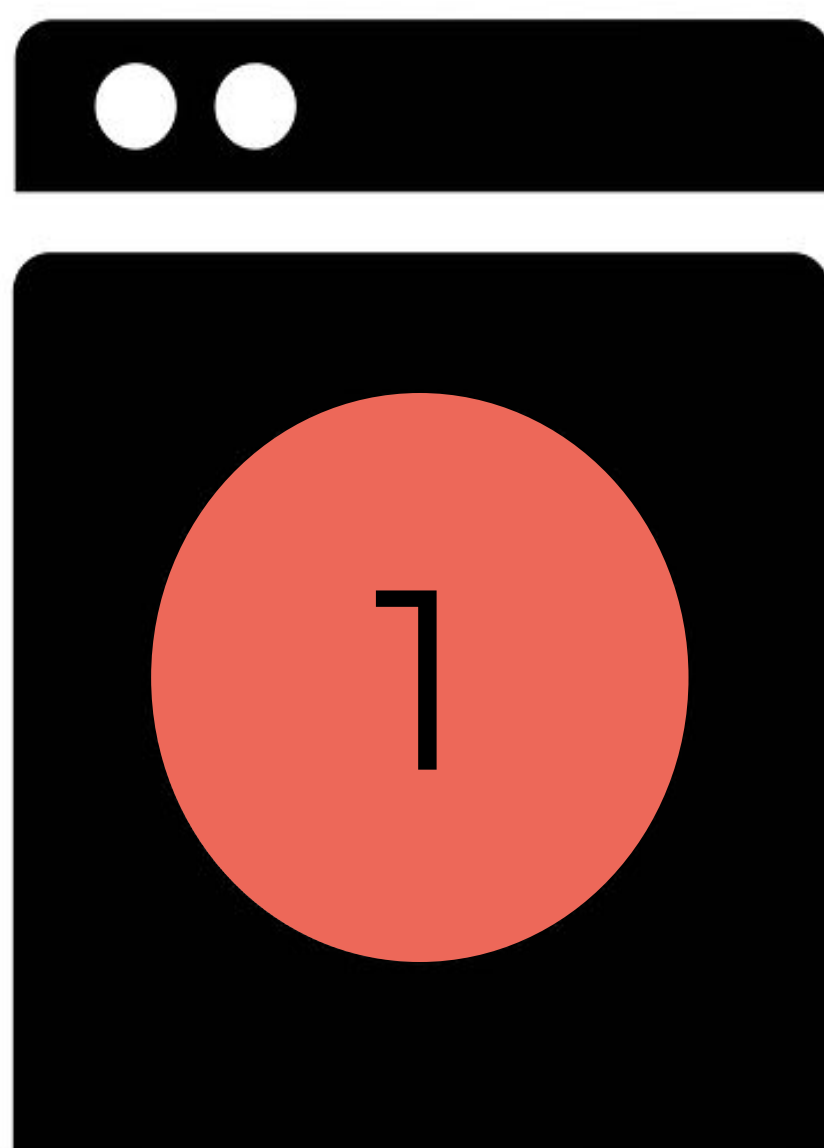


La pulizia dei dati è fondamentale per assicurarsi che siano **strutturati bene** all'interno delle **righe** e delle **colonne** e che i valori inseriti siano **coerenti**.

Analizzare i dati senza fare un attento data cleaning potrebbe portare successivamente a **gravi errori** o a difficoltà ingestibili.

Esistono alcuni strumenti che aiutano la pulizia dei dati, come **Open Refine**

(<http://openrefine.org>)

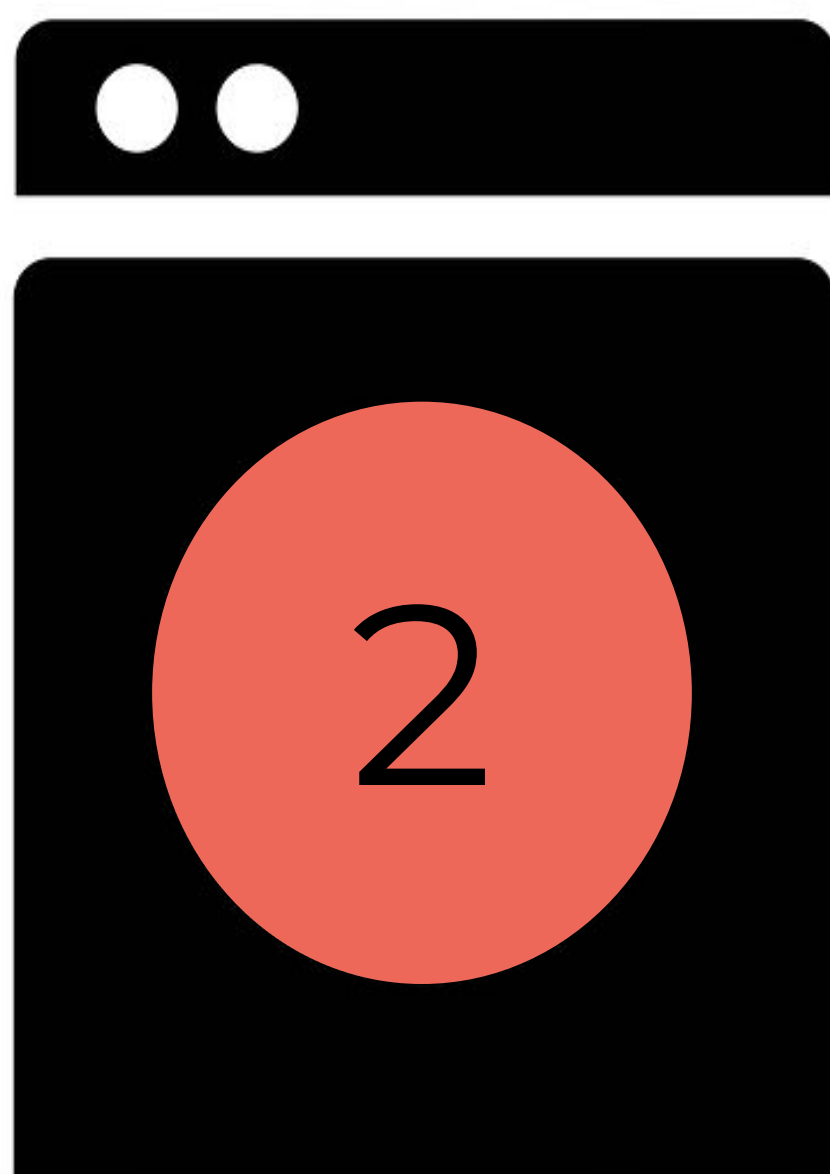


**Controllate che gli stessi valori**

**siano scritti allo stesso modo.**

Per esempio, se il valore “Ambiente” è scritto anche “ambiente” e “AMBIENTE”, allora è un errore, perché i valori devono essere scritti tutti in maniera analoga.





**Controllate le date** (se presenti): anche queste devono essere scritte in maniera omogenea.

È un errore se alcune date sono inserite in formato numerico (esempio: “01/03/2015”) e altre in formato alfanumerico (esempio: “10 marzo 2015”). È necessario avere le date in un solo formato: verificatelo andando su “Formato celle”.



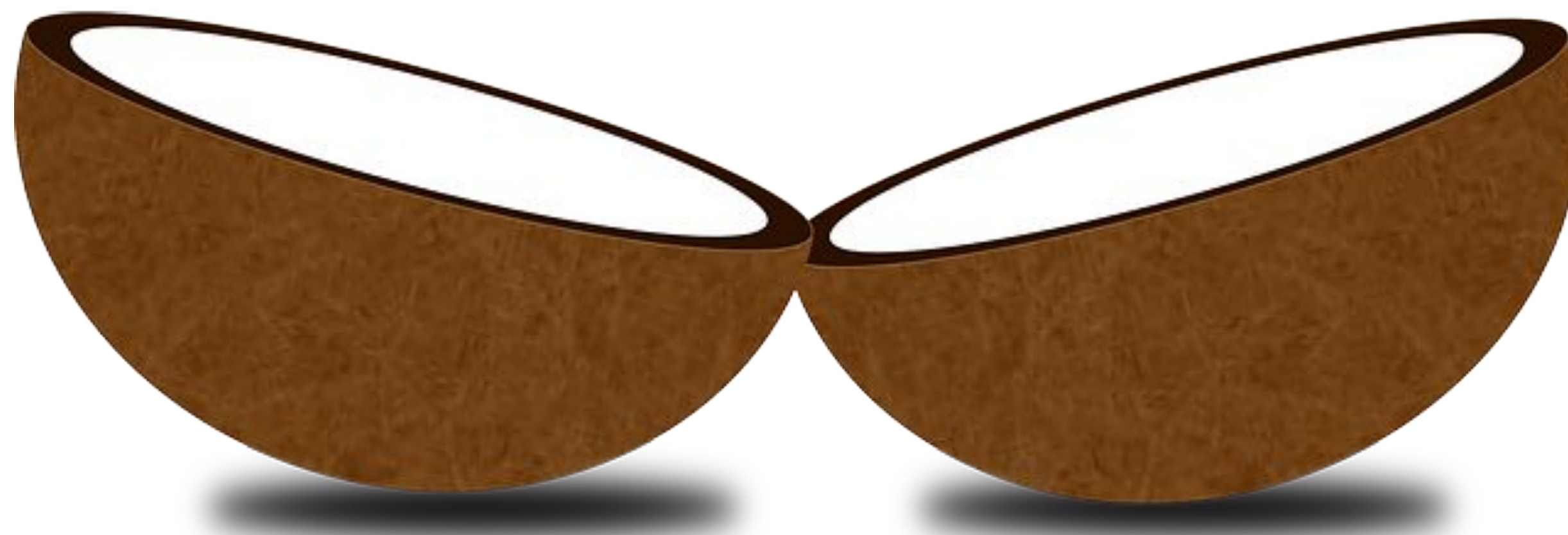
### **Verificate la numerazione.**

Quella italiana è diversa da quella anglosassone: in Italia i **punti** (“.”) indicano le migliaia, le **virgole** (“,”) indicano i decimali, mentre per gli anglosassoni è il **contrario**.

Assicuratevi che i numeri siano rappresentati correttamente.

Verificate il sistema di numerazione su “Impostazioni foglio di lavoro”.

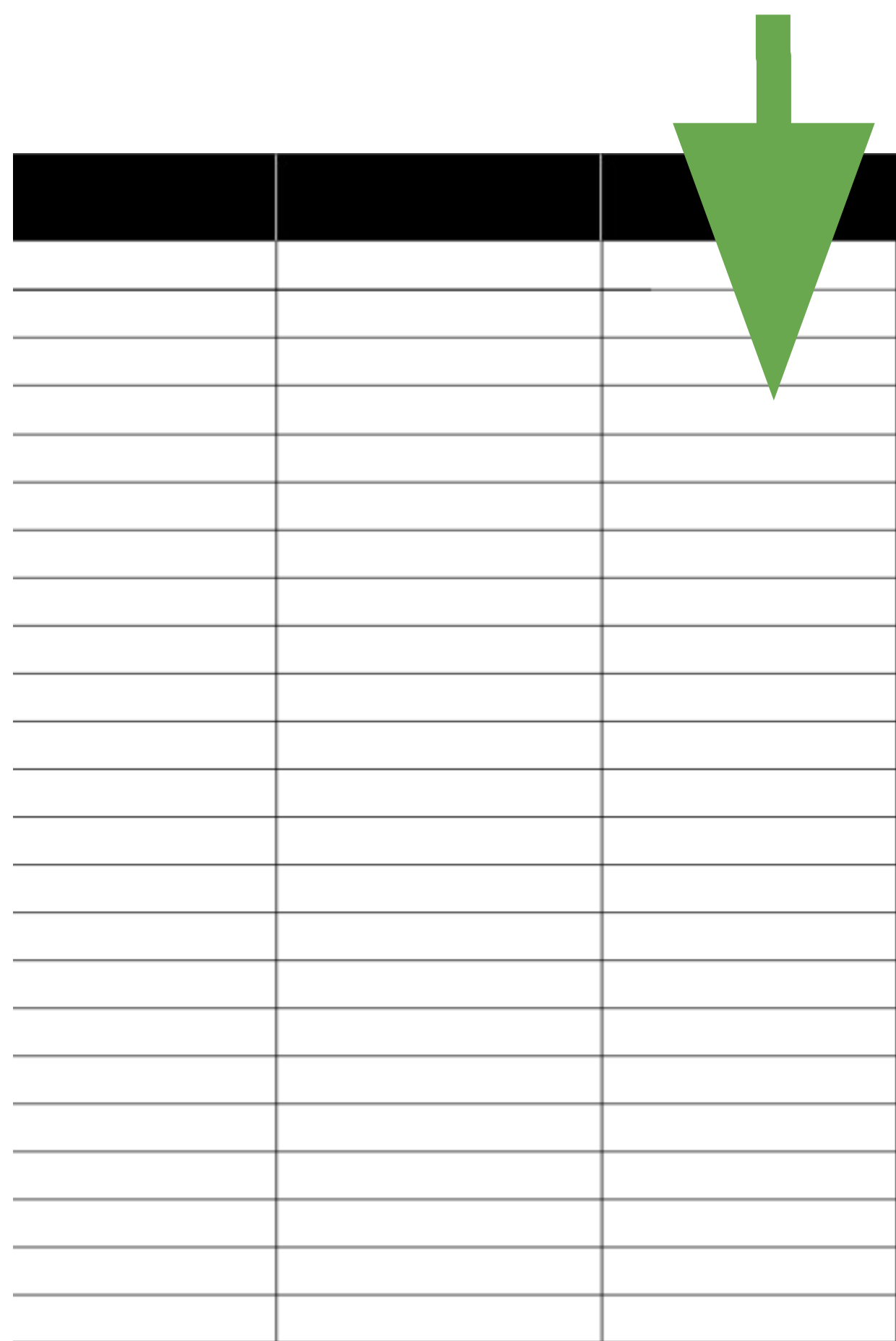
# ARRICCHIRE I DATI

**DIVIDERE**

I dati devono essere “granulari”, come se fossero degli atomi.

Per esempio: se i vostri dati hanno una colonna di **luoghi** e all’interno hanno un valore come questo: “*Cologno Monzese, provincia di Milano (Lombardia)*”, è meglio **dividere** la colonna in altre **tre colonne** (Comune + Provincia + Regione). Così potete **analizzarli** per **Comune, Provincia, Regione**.

## AGGIUNGERE



The diagram shows a table with a header row and 15 data rows. The header row is black. The first two columns of the data rows are also black. A large green arrow points downwards from the top of the third column, indicating the addition of a new column to the dataset.



Non è detto che il vostro dataset vi fornisca tutte le informazioni necessarie. In certi casi potreste essere voi ad arricchirlo, **aggiungendo nuove colonne** con specifiche informazioni. In parte potreste averlo già fatto, seguendo le indicazioni precedenti, ma possono esserci altri modi per arricchire un dataset...



## GEOCODING

	Latitudine	Longitudine

Usando lo stesso esempio, dopo aver diviso la colonna “Luogo” in tre colonne (Comune, Provincia, Regione), è utile inserire “Latitudine” e “Longitudine” per costruire una mappa.

Per farlo, inserirete due nuove colonne (Latitudine e Longitudine) e risalirete alle coordinate geografiche attraverso il luogo di riferimento.

## GEOCODING



Potete ricavare online le coordinate geografiche.

In rete ci sono tanti tool per farlo: cercateli!

Assicuratevi sempre che i dati siano **riusabili** (con OpenStreetMap si può, con GoogleMap con limiti).



*A Scuola di*  
**OPENCOESIONE**

**COMPRENDERE E RAFFINARE I DATI**



**COESIONE  
ITALIA**



**Presidenza del Consiglio dei Ministri**  
**Dipartimento per le politiche di coesione e per il sud**

In collaborazione con



**MIM**  
Ministero dell'Istruzione  
e del Merito



Progetto attualmente finanziato  
con il sostegno di

**POC** **PROGRAMMA  
OPERATIVO  
COMPLEMENTARE**

**POIN** **GOVERNANCE  
E CAPACITÀ  
ISTITUZIONALE  
2014-2020**