



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO

LA STATALE

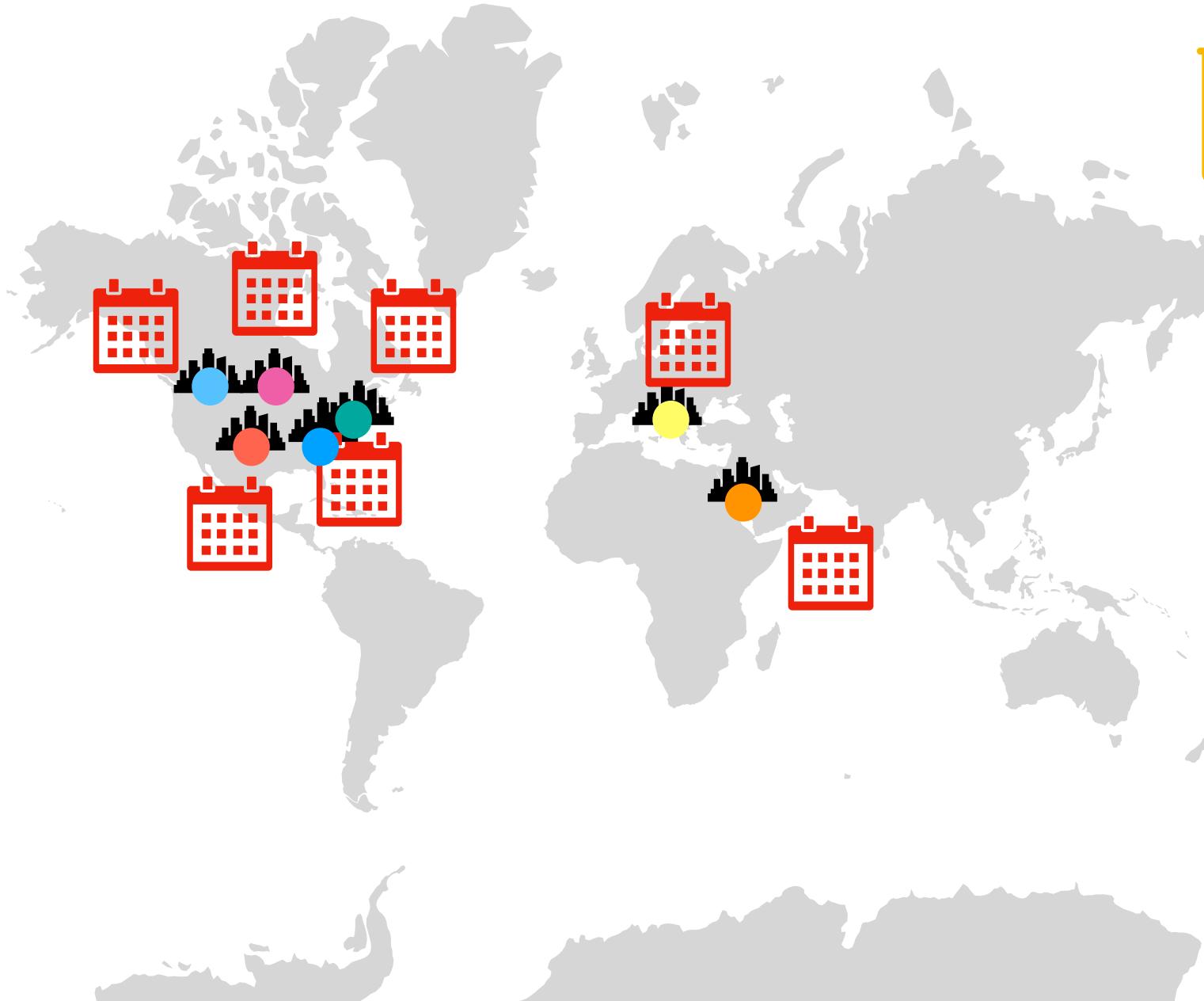
# Dati e informazioni

Problemi e requisiti della gestione dei dati

Prof. Stefano Montanelli

# Dati vs Informazione

Un **dato** è la registrazione simbolica di un elemento della realtà



	Month	Day	Year	Region	Country	State	City	AvgTemperature
2737324	6	21	2007	North America	US	Texas	San Antonio	76.6
1812220	7	18	2017	North America	US	Hawaii	Honolulu	83.8
631345	10	24	2016	Europe	Albania	NaN	Tirana	63.8
1781722	2	22	2010	North America	US	Georgia	Columbus	57.7
1077867	2	29	2004	Middle East	Saudi Arabia	NaN	Riyadh	63.8
1666628	7	21	2005	North America	US	Delaware	Wilmington	80.1
1620691	8	13	2006	North America	US	Colorado	Denver	70.6
1863000	5	17	2004	North America	US	Illinois	Springfield	69.8
2297985	8	1	2012	North America	US	New York	New York City	75.2
2318473	12	11	2017	North America	US	New York	Syracuse	26.4

In una **base di dati** i dati sono **correlati**. Correlare i dati ha lo scopo primario di fornirne un'interpretazione e trasformarli pertanto in **informazione**.

## **ROGNONE TRIFOLATO**

### **INGREDIENTI**

- 150 g rognone di vitello
- 60 g spinaci mondati
- 20 g vino bianco secco
- 10 g prezzemolo mondato
- uno spicchio d'aglio
- due cucchiaini di olio extravergine d'oliva
- sale
- pepe bianco in grani

Per la ricetta del rognone trifolato, sciacquate bene il rognone, accomodatelo in una padella con un cucchiaino d'olio e lo spicchio d'aglio non pelato, incoperchiatelo e fatelo stufare lentamente per circa 10', quindi bagnatelo con il vino e lasciatelo ridurre. A cottura ultimata (dovrà restare leggermente rosato all'interno), salate il rognone, completatelo con il prezzemolo tritato, tagliatelo a fettine e sistematelo nel piatto con il suo sughetto. Accompagnatelo con gli spinaci lessati, caldi, conditi con un pizzico di sale, il resto dell'olio e una macinata di pepe

## **Dati non strutturati**

Disponibili tipicamente in formato testuale (ma anche immagini, video o audio)

Memorizzati su file o in sistemi di gestione documentale

Possono presentare informazione strutturata, ma in cui la struttura non è esplicita e richiede di essere estratta dal file di origine

```
▼<items termsofuse="https://boardgamegeek.com/xmlapi/termsofuse">
  ▼<item type="boardgame" id="308119">
    <thumbnail>https://cf.geekdo-images.com/y1UmpxJNUJXMyz6Kg1YqEw__thumb/img/9NMhr6hW50vEmGyJUTWWin/200x150/filters:strip_icc()/pic6039229.jpg</thumbnail>
    <image>https://cf.geekdo-images.com/y1UmpxJNUJXMyz6Kg1YqEw__original/img/40uHX_Z6J00FHvVqI</image>
    <name type="primary" sortindex="1" value="Pax Renaissance: 2nd Edition"/>
    <name type="alternate" sortindex="1" value="Pax Renaissance 2"/>
    <description>Pax Renaissance 2nd Edition is a game for 1 to 4 players where you play as a banker in Europe at the height of the Renaissance during the late 15th and early 16th-century. A time when European nobility and warlords declined and the arts and technology advanced in the wake of a new economic force: merchants and bankers. These bankers became the wealthiest private individuals ever as they guided Europe into the modern age. As a Renaissance banker, you will finance kings or republics, sponsor voyages of discovery, join secret cabals, or unleash jihads and inquisitions. Your choices determine if Europe is elevated into the bright modern era or remains festering in dark feudalism. Four victories determine the future course of Western Society: will it be towards imperialism, trade globalization, religious totalitarianism, or enlightened art and science? New components include punchouts for pirate ships, pawns for concessions and serfs, placards for player boards, and empire squares allowing storage of repressed tokens and queens even if the King is not in a player Tableau. Extra game Map Cards are included, allowing the mounted gameboard to be left behind for small game portability. Useful if you have limited space in your game bag. Expanded rules include Dalmatians, Cryptography, Condottiere, apostasy, and the solitaire game. Strategy and Frequently-Asked Questions are included.</description>
    <yearpublished value="2021"/>
    <minplayers value="1"/>
    <maxplayers value="4"/>
  ▼<poll name="suggested_numplayers" title="User Suggested Number of Players" totalvotes="20">
    ▼<results numplayers="1">
      <result value="Best" numvotes="1"/>
      <result value="Recommended" numvotes="6"/>
      <result value="Not Recommended" numvotes="8"/>
    </results>
```

[https://www.boardgamegeek.com/xmlapi2/thing?  
id=308119,235555&stats=1&marketplace=1&comments=1&ratingcomments=1&pagesize=100&page=1](https://www.boardgamegeek.com/xmlapi2/thing?id=308119,235555&stats=1&marketplace=1&comments=1&ratingcomments=1&pagesize=100&page=1)

# Dati semi strutturati

Disponibili in formati diversi (e.g., XML, HTML, Json, RDF)

Tipicamente memorizzati in DBMS specifici (come ad esempio DBMS NoSql) o su file

Presentano prevalentemente informazione strutturata, ma con pochi **vincoli di formato e coerenza dei dati e frequente ridondanza**

# Dati (apparentemente) strutturati

scopus

Authors	Author(s) ID	Title	Year	Source title	Volume	Issue	Art. No.	Page st	Page	Page	Cited	DOI	Link	Document Type	Publication Stage	Open Access	Source	EID	
Castano, S., Falduti, M., Fe	56267928600	A know	2021	Information Systems			101842					10.10	<a href="https://">https:/</a> Article	Article in Press			Scopus	2-s2.0-851	
Bursic, S., Boccignone, G.,	57217259822	Improv	2020	Applied Scier	10	11	4002				2	10.33	<a href="https://">https:/</a> Article	Final	All Open Acces	Scopus	2-s2.0-850		
Castano, S., Falduti, M., Fe	56267928600	The LA	2020	CEUR Works	2690			15	26				<a href="https://">https:/</a> Conference Paper	Final			Scopus	2-s2.0-850	
Castano, S., Falduti, M., Fe	56267928600	A Boot	2020	CEUR Works	2646			1	11				<a href="https://">https:/</a> Conference Paper	Final			Scopus	2-s2.0-850	
Faria, D., Ferrara, A., Jimér	35247987600	Crowd	2020	Knowledge Engineering Review	e33							10.10	<a href="https://">https:/</a> Article	Article in Press	All Open Acces	Scopus	2-s2.0-850		
Castano, S., Ferrara, A., Fa	56267928600	Crime	2019	Proceedings of the 17th International C			179	183			4	10.11	<a href="https://">https:/</a> Conference Paper	Final			Scopus	2-s2.0-850	
Castano, S., Ferrara, A., Ma	56267928600	Levera	2019	Future Gener	95			790	801			1	10.10	<a href="https://">https:/</a> Article	Final			Scopus	2-s2.0-850

Disponibili prevalentemente in formato tabellare

Tipicamente memorizzati in DBMS relazionali, ma anche in fogli di calcolo come Excel

La struttura è rigida, definita da un insieme di campi che si applicano a tutti gli oggetti descritti, ma non è detto che siano rispettati **vincoli di formato e coerenza dei dati** e che non vi sia **ridondanza**

# Schema e istanza

Il livello di strutturazione di una base di dati dipende dalla presenza o meno di uno **schema** predefinito dei dati e dal livello di **normatività di tale schema in merito alle regole (vincoli)** che definiscono un'una descrizione degli oggetti rappresentati (**istanza**) come valida rispetto allo schema

## Nessuno schema

**ROGNONE TRIFOLATO**

**INGREDIENTI**

- 150 g rognone di vitello
- 60 g spinaci mondati
- 20 g vino bianco secco
- 10 g prezzemolo mondato
- uno spicchio d'aglio
- due cucchiaini di olio extravergine d'oliva
- sale
- pepe bianco in grani

Per la ricetta del rognone trifolato, sciacquate bene il rognone, accomodatelo in una padella con un cucchiaino d'olio e lo spicchio d'aglio non pelato, incoperchiatelo e fatelo stufare lentamente per circa 10', quindi bagnatelo con il vino e lasciatelo ridurre. A cottura ultimata (dovrà restare leggermente rosato all'interno), salate il rognone, completatelo con il prezzemolo tritato, tagliatelo a fettine e sistematelo nel piatto con il suo suggerito. Accompagnatelo con gli spinaci lessati, caldi, conditi con un pizzico di sale, il resto dell'olio e una macinata di pepe

```

<items termsfuse="https://boardgamegeek.com/xmlapi/termsofuse">
  <item type="boardgame" id="308119">
    <thumbnail>https://cf.geekdo-images.com/y1UmpxJNUJXMyz6Kg1YqEw
    in/200x150/filters:strip_icc()/pic6039229.jpg</thumbnail>
    <image>https://cf.geekdo-images.com/y1UmpxJNUJXMyz6Kg1YqEw</image>
    <name type="primary" sortindex="1" value="Pax Renaissance: 2nd
    <name type="alternate" sortindex="1" value="Pax Renaissance 2">
      <description>Pax Renaissance 2nd Edition is a game for 1 to 4
      and early 16th-century. A time when European nobility
      merchants and bankers. These bankers became the wealthiest pri
      you will finance kings or republics, sponsor voyages of discov
      elevated into the bright modern era or remains festering in da
      imperialism, trade globalization, religious totalitarianism, o
      concessions and serfs, placards for player boards, and empire
      Tableau. Extra game Map Cards are included, allowing
      in your game bag. Expanded rules include Dalmatians,
      are included.</description>
      <yearpublished value="2021"/>
      <minplayers value="1"/>
    </name>
  </item>
</items>
  
```

Schema

Istanza

Istanza

## Schema

Authors	Author(s) ID	Title	Year	Source title
Castano, S., Faldu, M., Fe	56267928600	A knov	2021	Information S
Bursic, S., Boccignone, G.,	57217259822	Improv	2020	Applied Scier
Castano, S., Faldu, M., Fe	56267928600	The LA	2020	CEUR Works
Castano, S., Faldu, M., Fe	56267928600	A Boot	2020	CEUR Works
Faria, D., Ferrara, A., Jimér	35247987600	Crowd	2020	Knowledge E
Castano, S., Ferrara, A., Fa	56267928600	Crime	2019	Proceedings
Castano, S., Ferrara, A., Mo	56267928600	Levera	2019	Future Gener

scopus

Istanze

# Sfide nella gestione dei dati

Gestire i dati richiede strumenti per affrontare efficacemente (e efficientemente) le seguenti sfide:

**Duplicazioni**

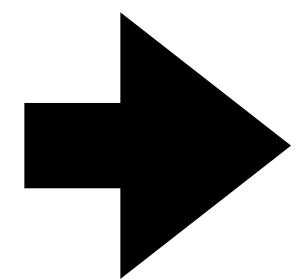
**Violazione dell'integrità**

**Errori di inserimento**

**Mancanza di controlli sui tipi**

**Difficoltà di collegamento fra file**

**Controllo dell'accesso**



**DBMS ( DataBase Management System)**



# Caratteristiche delle basi di dati

- Autodescrizione -> linguaggi DDL, DML
- Indipendenza (fisica, logica) -> Astrazione
- Viste multiple -> multi-utenza, linguaggio DQL
- Controllo della concorrenza -> condivisione
- Controllo dell'accesso -> linguaggio DCL

# Caratteristiche delle basi di dati

- In linea di principio, *ogni collezione di dati può considerarsi una base di dati*
- Tuttavia, si assume che una base di dati abbia alcune caratteristiche distinctive:
  - Rappresenta un aspetto limitato della realtà, riferito a un sottoinsieme dei dati disponibili denominato *universo del discorso*
  - Costituisce un insieme di dati logicamente coerenti e dotati di un significato intrinseco; una collezione casuale di dati non è una base di dati
  - Una base di dati è progettata, costruita e popolata per uno scopo specifico, rivolto a un gruppo di utenti determinato e a alcune specifiche applicazioni a cui gli utenti sono interessati (da cui discendono *vincoli* e *regole*)