

# NEXA

Center for Internet & Society

## L'enorme archivio di dati: il Web

Giuseppe Rizzo

Dottorando di Ricerca

Dipartimento di Automatica ed Informatica

Politecnico di Torino

Email: [giuseppe.rizzo@polito.it](mailto:giuseppe.rizzo@polito.it)

29° Mercoledì di Nexa

Centro NEXA su Internet & Società

Politecnico di Torino

Torino, 9 Marzo 2011



# Agenda

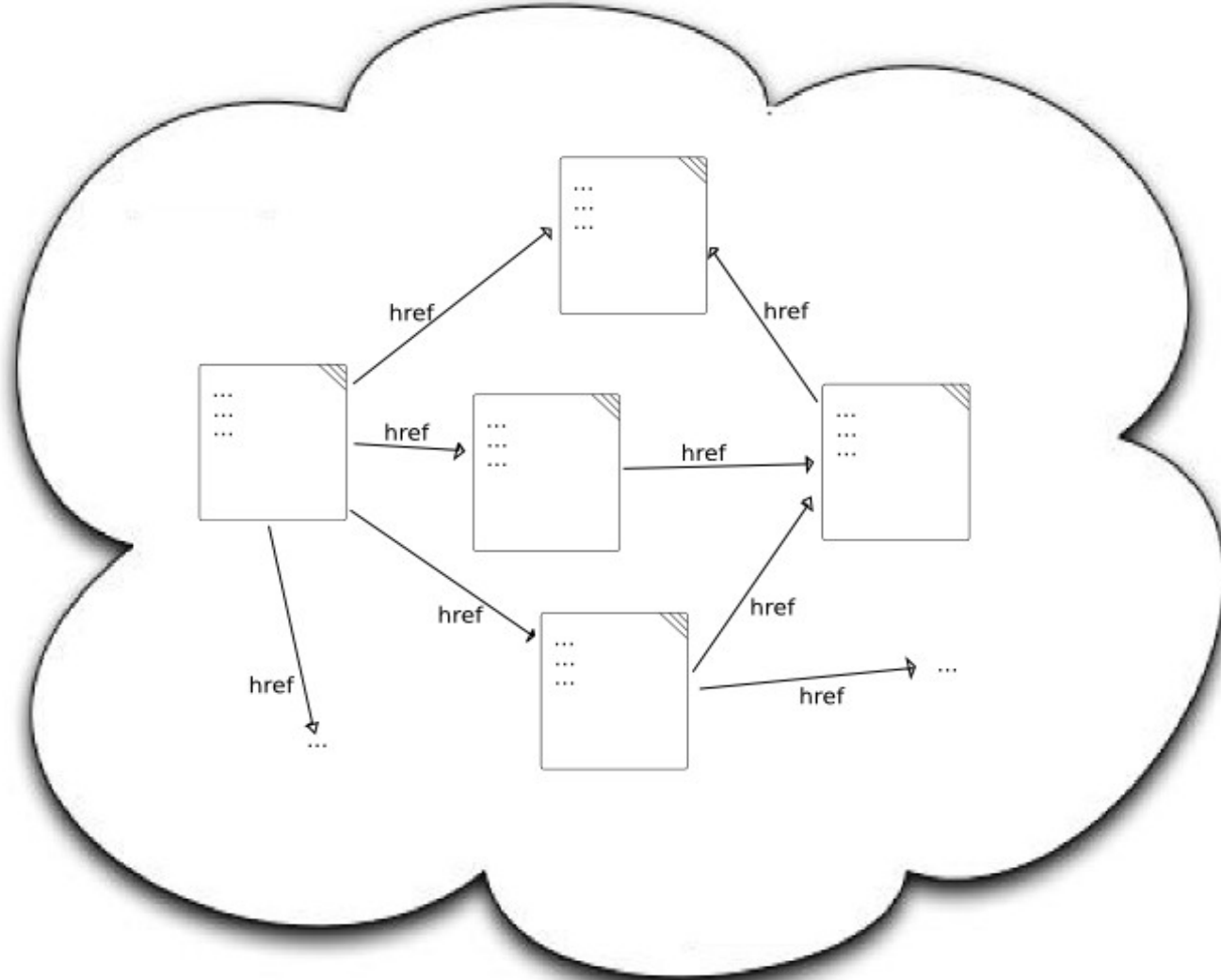
- ◆ Idea
- ◆ Il Web dei documenti
- ◆ Il Web dei dati
- ◆ Linked Data
- ◆ Come condividere i dati
- ◆ Conclusioni



# Idea



Il Web è uno spazio di informazione globale,  
un enorme archivio distribuito di dati





- ◆ E' l'elemento atomico di scambio dell'informazione
- ◆ E' il contenitore di differenti tipi di informazione:
  - codice, usato dalle applicazioni per creare le viste o interfacce di presentazione
  - testo, immagini e video, comprensibili all'uomo



# Esempio: Wikipedia

Log in / create account

Article **Discussion** Read Edit View history Search

**Turin** ←

From Wikipedia, the free encyclopedia Coordinates: 45°04'N 07°42'E  
 (Redirected from [Torino](#))

For other uses, see [Turin \(disambiguation\)](#) ←  
 "Torino" redirects here. For other uses, see [Torino \(disambiguation\)](#).


**Turin** (Italian: *Torino*, pronounced [toˈriːno] listen<sup>ⓘ</sup>; Piedmontese: *Turin*, pronounced [tyˈrin]) is a major city which is a business and cultural centre in northern Italy, capital of the Piedmont region, located mainly on the left bank of the Po River surrounded by the Alpine arch. The population of the city proper is 909,193 (November 2008) while the population of the urban area is estimated by Eurostat to be 1.7 million inhabitants; the Turin metropolitan area is estimated by OECD to have a population of 2.2 million.<sup>[2]</sup>

Turin is a flourishing, industrious and cosmopolitan<sup>[citation needed]</sup> European city, which enjoys state-of-the-art technology and architectural developments.<sup>[3][4][unreliable source?]</sup>The city has a rich culture and history, and is known for its numerous art galleries, restaurants, churches, palaces, operahouses, piazzas, parks, gardens, theatres, libraries, museums and other venues. Turin is well-known for its baroque, rococo, neo-classical, and Art


**Turin**  
Torino

← **Comune** →

**Comune di Torino**



A collage of Turin: in the top left is the Mole Antonelliana, followed by a view of the city under the snow, the Piazza Vittorio Veneto, the Royal Palace of Turin and the Museo del Risorgimento (Palazzo Carignano)



**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact Wikipedia

Toolbox

Print/export

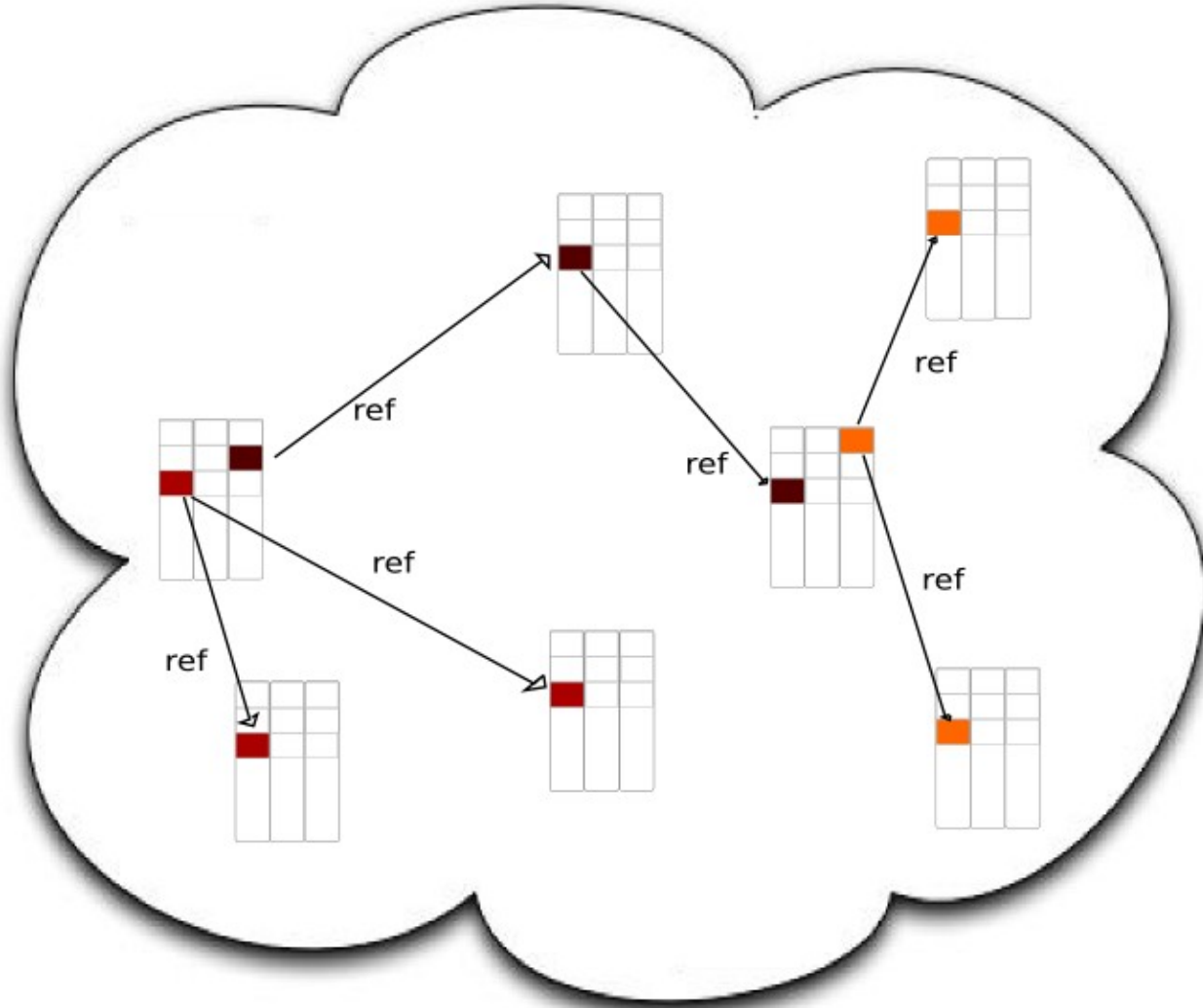
Languages  
Afrikaans  
العربية  
Armãneashce  
Arpetan  
Aymar aru  
Azərbaycanca  
বাংলা  
Bân-lâm-gú  
Беларуская  
Беларуская (тарашкевіца)  
Беларускі

I dati e le informazioni di formattazione sono aggregati e presentati all'utilizzatore

I riferimenti ipertestuali sono annegati nel documento



# Web dei dati





- ◆ L'elemento atomico di informazione è il dato
- ◆ Al dato non è associata la formattazione: il dato è grezzo, “raw”
- ◆ La vista viene creata dal fruitore dell'informazione





# Esempio: DBpedia

## Wikipedia

**WIKIPEDIA**  
The Free Encyclopedia

**Turin**  
From Wikipedia, the free encyclopedia

*For other uses, see **Turin (disambiguation)**.*  
*"Torino" redirects here. For other uses, see **Torino (disambiguation)**.*

**Turin** (Italian: *Torino*, pronounced [toˈriːno]  (help·listen); Piedmontese: *Turin*, pronounced [tyˈɾiŋ]) is a major city which is a business and cultural centre in northern Italy, capital of the Piedmont region, located mainly on the left bank of the **Po River** surrounded by the **Alpine arch**. The population of the city proper is 909,193 (November 2008) while the population of the urban area is estimated by Eurostat to be 1.7 million inhabitants; the Turin metropolitan area is estimated by OECD to have a population of 2.2 million.<sup>[2]</sup>


Turin is a flourishing, industrious and cosmopolitan European city, which enjoys state-of-the-art technology and architectural developments.<sup>[3][4][unreliable source?]</sup> The city has a rich culture and history, and is known for its numerous art galleries, restaurants, churches, palaces, operahouses, piazzas, parks, gardens, theatres, libraries, museums and other venues. Turin is well-known for its baroque, rococo, neo-classical, and Art Nouveau architecture. Much of the city's public squares, castles, gardens and elegant *palazzi* (such as Palazzo Madama), were built by Sicilian architect Filippo Juvarra, who modelled these buildings on the Baroque and classical style of Versailles.<sup>[5]</sup> Examples of these French-themed edifices include the Royal Palace of Turin, the Palazzina di caccia of Stupinigi and the Basilica di Superga. Turin is sometimes called the "cradle of Italian liberty", due to its having been the birthplace and home of notable politicians and people who contributed to the *Risorgimento*, such as Cavour.<sup>[6]</sup> The city currently hosts some of Italy's best universities, colleges, academies, lycea and gymnasia, such as the Polytechnic University of Turin. Prestigious and important museums, such as the Museo Egizio<sup>[7]</sup> and the Mole Antonelliana are also found in the city. Turin's several monuments and sights make it one of the world's top 250 tourist destinations, and the tenth most visited city in Italy in 2008.<sup>[8]</sup>

Turin used to be a major European political centre, being Italy's first capital city in 1861 and being home to the House of Savoy, Italy's royal family.<sup>[9]</sup> Even though much of its political significance and importance had been lost by World War II, it became a major European crossroad for industry, commerce and trade, and currently is one of Italy's main industrial centres, being part of the famous "industrial triangle", along with Milan and Genoa. Turin is ranked third, after Rome and Milan, for economic strength.<sup>[10]</sup> With a GDP of \$58 billion, Turin is the world's 78th richest city by purchasing power,<sup>[11][12]</sup> and even though the city was unable to become a "world city", unlike Milan or Rome, it was ranked by GaWC as "economically efficient", along with Jerusalem, Genoa, Macau, Marseille, Liverpool, Strasbourg, Salt Lake City, Seville and Tijuana, to name a few.<sup>[13]</sup> Turin is also home to much of the Italian automotive industry.<sup>[14][15]</sup>


Turin is well known as the home of the Shroud of Turin, the football teams Juventus F.C. and Torino F.C., the headquarters of automobile manufacturers Fiat, Lancia and Alfa Romeo, and as host of the 2006

Coordinates: 45°04′N 07°42′E﻿ / ﻿45.067°N 7.7°E﻿ / 45.067; 7.7

**Turin**  
Torino  
— Comune —  
Comune di Torino



A collage of Turin: in the top left is the Mole Antonelliana, followed by a view of the city under the sun, the Piazza Vittorio Veneto, the top of Palace of Turin and the Museo del Risorgimento (Palazzo Caviglioglio)



Coat of arms

Location of Turin in Italy

Coordinates: 45°04′N 07°42′E﻿ / ﻿45.067°N 7.7°E﻿ / 45.067; 7.7

<b>Country</b>	<span><span><span></span></span><span> </span></span> Italy
<b>Region</b>	Piedmont
<b>Province</b>	Turin (TO)
<b>Government</b>	
<span> </span> - <b>Mayor</b>	Sergio Chiamparino (Democratic Party)
<b>Area</b>	
<span> </span> - <b>Total</b>	130.17 <span> </span> km <sup>2</sup> (50.3 <span> </span> sq <span> </span> mi)
<b>Elevation</b>	229 <span> </span> m (754 <span> </span> ft)
<b>Population</b> (30 April 2009) <sup>[1]</sup>	
<span> </span> - <b>Total</b>	910,188
<span> </span> - <b>Density</b>	6,992.3/km <sup>2</sup> (18,110/sq <span> </span> mi)
<b>Demonym</b>	Torinese
<b>Time zone</b>	CET (UTC+1)
<span> </span> - <b>Summer (DST)</b>	CEST (UTC+2)
<b>Postal code</b>	10100, 10121-10156
<b>Dialing code</b>	011
<b>Patron saint</b>	John the Baptist
<b>Saint day</b>	24 June
<b>Website</b>	<i>Official website</i>



## DBpedia

dbpedia-owl:areaCode	011
dbpedia-owl:areaTotal	130170000.000000 (xsd:double)
dbpedia-owl:elevation	239.000000 (xsd:double)
dbpedia-owl:leaderName	dbpedia:Sergio_Chiamparino
dbpedia-owl:populationAsOf	2009-04-30 (xsd:date)
dbpedia-owl:populationTotal	910188 (xsd:integer)
dbpedia-owl:postalCode	10100, 10121-10156
dbpedia-owl:province	dbpedia:Province_of_Turin
dbpedia-owl:region	dbpedia:Piedmont
dbpedia-owl:saint	dbpedia:John_the_Baptist
dbpedia-owl:thumbnail	http://upload.wikimedia.org/wikipedia/co
dbpedia-owl:wikiPageExternalLink	<ul style="list-style-type: none"> <li>http://www.fieralibro.it/</li> <li>http://www.worldstatesmen.org/Italy_sta</li> <li>http://torino.cittametropolitana.com</li> <li>http://www.buddies.it</li> <li>http://www.universiadetorino2007.org/ET</li> <li>http://www.aboutturin.com/en/</li> <li>http://www.museonazionaledelcinema.o</li> <li>http://www.torinofilmfest.org/index.php?</li> <li>http://www.comune.torino.it</li> <li>http://www.flickr.com/photos/wildshutter</li> <li>http://www.witchology.com/contents/inte</li> <li>http://www.museoegizio.it/pages/hp_en.</li> <li>http://www.torino2006.org/</li> <li>http://www.turismotorino.org/</li> <li>http://citymayors.com/interviews/turin_i</li> <li>http://www.comune.torino.it/en/</li> </ul>
dbpprop:areaCode	11 (xsd:integer)
dbpprop:areaTotalKm	130 (xsd:integer)
dbpprop:coordinatesDisplay	title
dbpprop:criteria	i, ii, iv, v
dbpprop:day	--06-24
dbpprop:elevationM	239 (xsd:integer)
dbpprop:hasPhotoCollection	http://www4.wiwiss.fu-berlin.de/flickrwrap



- ◆ Le informazioni sono strutturate
- ◆ Una macchina può attingere all'informazione grezza senza doverla filtrare per rimuovere il rumore



- ◆ Spazio di informazione globale
- ◆ Archivi di dati sparsi per la rete
- ◆ Dati tra loro “linked”, collegati : Linked Data
- ◆ Il fruitore può aggregare e collezionare dati provenienti da archivi distribuiti





# Esempio di Linked Data



Archivio sui livelli  
di ozono nell'aria  
della città di Torino



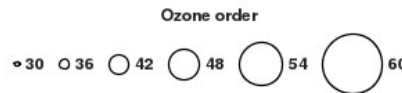
Archivio sui livelli  
di traffico della  
città di Torino



Quartiere	Ora	Livello ozono	Numero macchine	...



# App da data.gov



### Datasets Used

Dataset 8 : Clean Air Status and Trends Network (CASTNET): Ozone  
 Dataset 9 : Clean Air Status and Trends Network (CASTNET): Visibility

### SPARQL Query

Query 1  
 Query 2

<http://www.data.gov/semantic/Castnet/html/exhibit>

- <http://www.data.gov/raw/8>

- <http://www.data.gov/raw/9>



Rif. Tim Berners-Lee, "The Great Unveiling" in Long Beach, CA. USA, 4, Feb 2009.



- ◆ Utilizzo di URI<sup>1</sup> per l'identificazione delle risorse
- ◆ Utilizzo del protocollo HTTP per permettere di navigare le relazioni e consultare le risorse
- ◆ Uso di standard RDF<sup>2</sup> per la rappresentazione delle relazioni tra risorse ed un punto di accesso dove esplorare le inferenze
- ◆ Creare i collegamenti tra le varie URI al fine di scoprire nuovi dati

---

<sup>1</sup> URI : Uniform Resource Identifier, è l'indirizzo Web con il quale si identificano le risorse

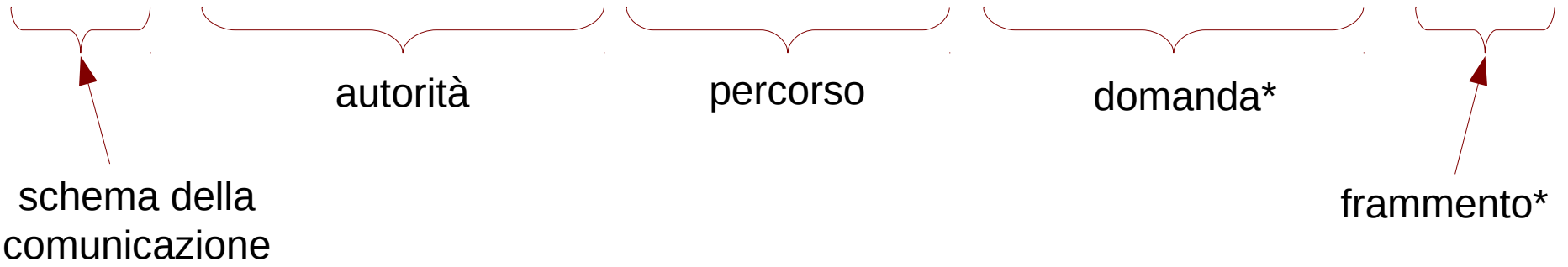
<sup>2</sup> RDF: Resource Description Framework, definisce come rappresentare le risorse





# URI

<http://example.com/over/there?name=ferret#nose>



\*: campi opzionali.



- ◆ E' un formalismo con il quale si definiscono le risorse e le relazioni tra esse
- ◆ Si basa sul concetto di dichiarazione
- ◆ La dichiarazione è composta da tre elementi, tripla: il soggetto, il predicato, l'oggetto
- ◆ Ogni elemento è descritto mediante una URI. L'oggetto inoltre può essere descritto da un “literal” ossia da una sequenza di caratteri



- ◆ Il Web dei dati permette di collegare informazioni tra archivi di dati non strettamente collegati
- ◆ Dati aperti permettono il libero consumo e la libera navigazione da parte del fruitore
- ◆ Dati aperti creano l'infrastruttura su cui il Linked Data si basa
- ◆ Il Web ha bisogno di “Raw data now” ma anche di dati facili da usare, riusare e distribuire



- ◆ Affidabilità dei dati: il Web è uno spazio aperto in cui tutti possono aggiungere informazioni. Si demanda al fornitore la responsabilità su quello che viene pubblicato
- ◆ Dati con riferimenti temporali e spaziali
- ◆ Freschezza dei dati: aggiornati in tempo reale



Grazie per il vostro tempo e per la vostra  
attenzione!