

Web Semantico: Tecnologie e Applicazioni

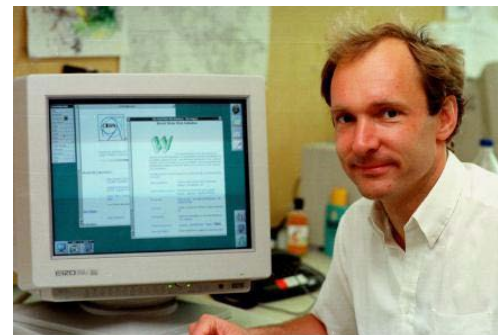
Ing. Irina Trubitsyna

Outline

- Introduzione
 - Web semantico:
 - Principi e Principali Tecnologie
- Markup dei dati strutturati
 - RDFa, Microdati, JSON-LD

Web Semantico

- Il termine Web Semantico è stato proposto da Tim Berners-Lee nel 2001, nell'ambito di un articolo pubblicato dalla rivista *Scientific American*.
- L'ideatore del WWW avanzava l'ambizione di creare una sorta di Internet “pensante”, in cui le informazioni siano processabili e comprensibili direttamente alle macchine.
- Nella concezione di Tim Berners-Lee, il termine Web Semantico è associato all'idea di un Web nel quale agiscano agenti intelligenti, cioè applicazioni in grado di comprendere il significato dei contenuti presenti on line e dunque in grado di guidare l'utente direttamente verso l'informazione ricercata, così come di sostituirsi all'utente stesso nello svolgimento di alcune operazioni.



Tim Berners Lee (2001)

«Fino ad oggi il Web si è sviluppato più come strumento di contenuti per uso ‘umano’, piuttosto che come mezzo di informazione per l’elaborazione automatizzata. Servono dati specificatamente progettati per computer, dati comprensibili alle macchine [...]

I computer devono aver accesso a serie strutturate di informazioni e a regole di deduzione che permettano un ‘ragionamento’ automatizzato»

« Il Web fu disegnato come uno spazio di informazioni, non solo per la comunicazione uomo-uomo, ma affinché anche le macchine potessero avere la possibilità di partecipare e dare il loro contributo. Uno dei maggiori ostacoli è stato il fatto che la maggior parte dell'informazione sul Web è disegnata per essere fruita dall'uomo (...), la struttura dei dati non è riconoscibile per un robot che naviga il Web.

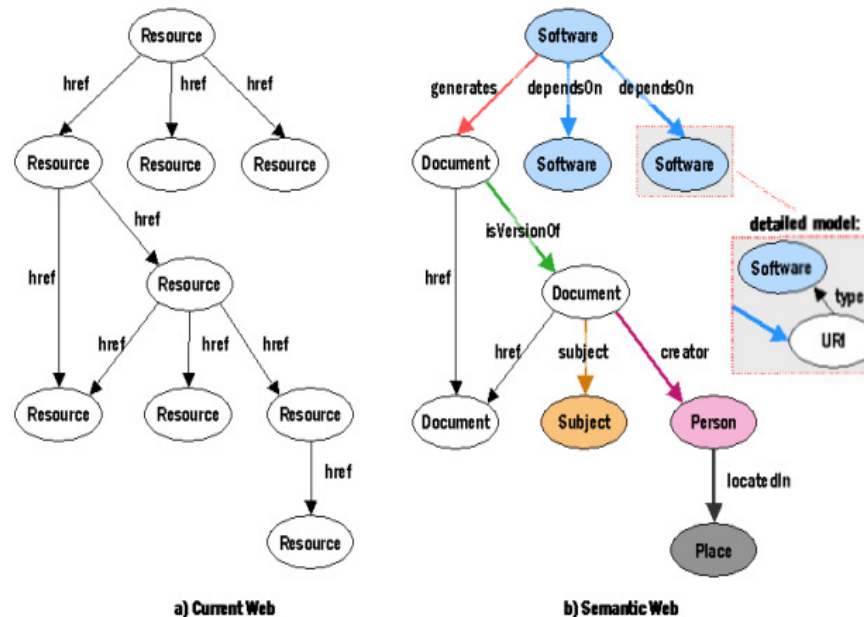
L'approccio del Web Semantico, invece, sviluppa linguaggi per esprimere le informazioni in una forma accessibile e processabile da una macchina».

Principi del Web Semantico (2002)

- **Ogni cosa può essere identificata attraverso una URI (Uniform Resource Identifier).** Qualsiasi oggetto, proprietà e stato presente in rete dovrebbero essere identificati in maniera univoca e non ambigua.

- **Le risorse e i link possono avere un “tipo”.**

L'idea è di avere una meta-informazione che associa un concetto a ciascuna risorsa e a ciascun link; tale associazione può consentire ai sistemi di comprendere le relazioni logiche e semantiche tra le risorse, esplicitate attraverso link.



Principi del Web Semantico (2002)

- **L'informazione parziale è tollerata.** Gli strumenti e le tecnologie del Web Semantico devono essere in grado di tollerare parzialità e incompletezza delle informazioni on-line (es. collegamenti ipertestuali soltanto “parziali” e incompleti) e capaci, quindi, di garantire alle macchine di comprendere il significato delle relazioni, anche se incomplete e inesatte dal punto di vista logico e semantico.
- **Non c'è bisogno della verità assoluta.** La veridicità delle informazioni disponibili sul Web non è dimostrabile in assoluto. Tuttavia, il Web Semantico si basa su applicazioni potenzialmente in grado di creare una «macchina della fiducia», i.e. verificare in automatico, senza l'intervento degli utenti, l'affidabilità dei documenti e dei contenuti presenti on line attraverso interpretazioni e inferenze.
- **L'evoluzione è tollerata.** Nel Web documenti, informazioni e contenuti sono soggetti a evoluzioni e mutamenti di significato. L'infrastruttura deve essere in grado di gestire il processo di evoluzione e aggiornamento delle informazioni in rete e dei relativi significati, non sostituendo quelli nuovi a quelli vecchi, ma integrandoli e coordinandoli.
- **Design minimalista.** Nella definizione tecnica del Web Semantico, il W3C ha inteso proporre una infrastruttura meno standardizzata possibile del Web Semantico, con l'obiettivo di facilitare lo sviluppo di uno spazio aperto alle innovazioni e alle nuove applicazioni on line. Tale infrastruttura fa riferimento ai quattro livelli (dati, schema, ontologico e logico), che disegnano un'architettura minimalista del Web Semantico.

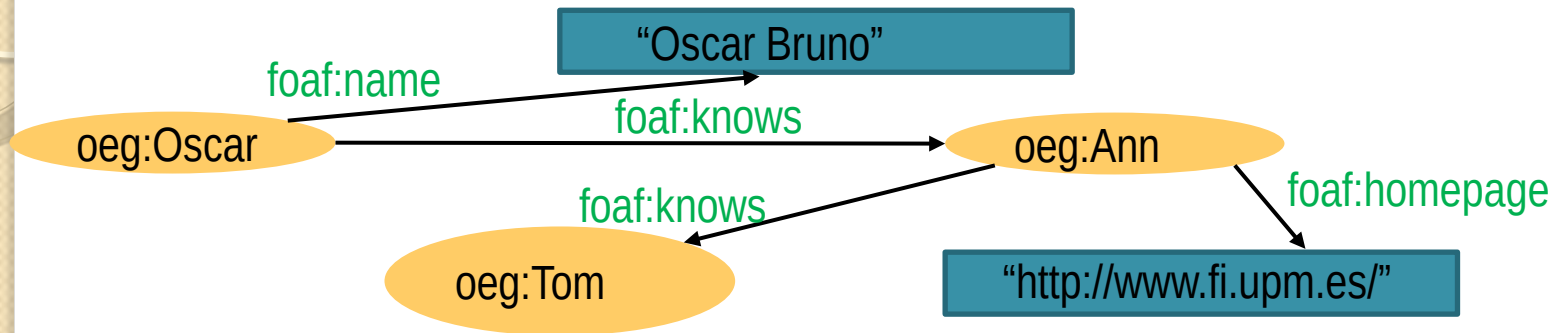
Principali Tecnologie

- ***RDF (Resource Description Framework):***
Il linguaggio di modellazione dei dati per il Web Semantico.
- ***SPARQL (SPARQL Protocol e RDF Query Language):***
Il linguaggio d'interrogazione del Web Semantico, progettato specificamente per interrogare i dati tra i vari sistemi.
- ***OWL (Web Ontology Language):***
Il linguaggio di schema, o il linguaggio di rappresentazione della conoscenza del Web Semantico.

RDF

- RDF è il modello di dati del Web Semantico ed è alla base della flessibilità innata di queste tecnologie.
- Tutti i dati del Web Semantico, compreso lo schema che descrive i dati RDF stessi, sono rappresentati in RDF.
- RDF è un grafo orientato etichettato.

Grafo RDF

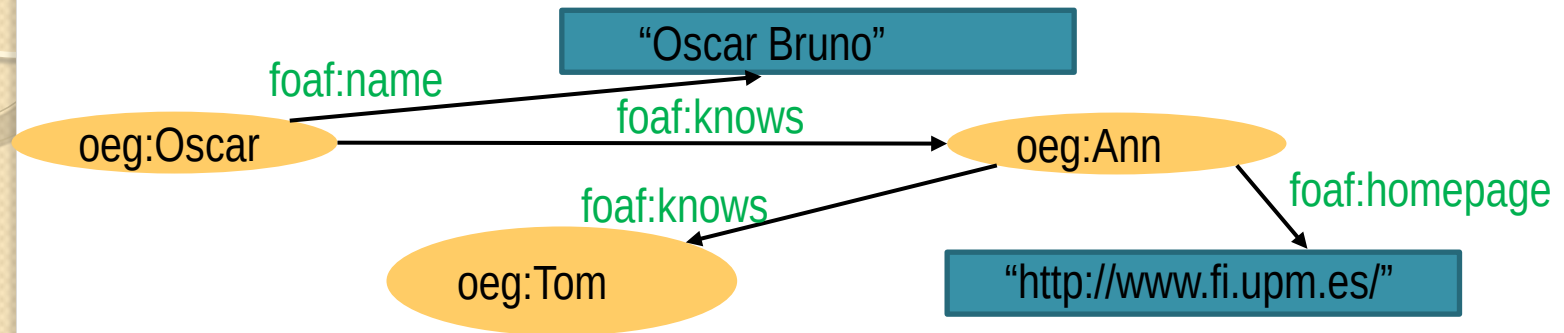


Ci sono tre tipi di nodi in un grafo RDF:

- **risorse (resource nodes):** rappresentate da ovali, sono identificate da URI e possono avere associate delle proprietà
- **letterali (literal nodes):** descrivono valori delle proprietà, sono rappresentati da rettangoli etichettati mediante le stringhe o anche URI
- **nodi bianchi (blank nodes):** sono risorse senza URI.

Per identificare risorse e le loro proprietà spesso si utilizza una forma ridotta di URI che chiameremo *CURIE (Compact URI)*. Un CURIE si presenta nella forma **prefisso:stringa**, dove il prefisso è una stringa che identifica l'URI di un namespace.

Grafo RDF



- Le *asserzioni (statement)* o *triple* che rappresentano gli archi del grafo RDF hanno la forma **(soggetto, predicato, oggetto)**.

Es: (oeg:Oscar, foaf:name, "Oscar Bruno")
(oeg:Oscar, foaf:knows, oeg:Ann)
(oeg:Ann, foaf:knows, oeg:Tom)
(oeg:Ann, foaf:homepage, "http://www.fi.upm.es/")

- Il grafo RDF è una semplice collezione di triple
- Una base di dati per la gestione di RDF è spesso chiamata *RDF Database* o *triplestore*.
- Un **Grafo Denominato (Named Graph)** è un insieme di triple identificato da un URI.
- Per riferirsi a una tripla in un grafo denominato, si utilizza la nozione di *quadrupla*, che ha la forma **(grrafo_denominato, soggetto, predicato, oggetto)**.

Vocabolari

- Il concetto di universalità, compreso e condiviso da tutti a livello teorico, è però difficilmente realizzabile in pratica. Per risolvere (almeno in parte) questo problema sono stati introdotti i vocabolari RDF, che definiscono insiemi di URI per gli archi dei grafi RDF.
- I vocabolari sono lo strumento fondamentale per dare semantica alle informazioni. Il vocabolario (universale o, più spesso, specifico per un determinato dominio di conoscenza) si basa sulla concettualizzazione condivisa di un dominio, fornendo gli elementi per rappresentarlo in modo formale ed esplicito.
- Per condividere i dati, le applicazioni devono
 - condividere un vocabolario comune o
 - tradurre un grafo RDF scritto utilizzando un dato vocabolario in un secondo grafo RDF definito su un altro vocabolario (utilizzando SPARQL).
- I vocabolari si creano per mezzo di RDF Schema o OWL.

Vocabolario FOAF

- **FOAF** (acronimo di **Friend of a friend** - Amico di un amico)
- <http://www.foaf-project.org/>
- descrivere persone, con le loro attività e le relazioni con altre persone e oggetti.

FOAF (2000-2015+)

Friend of a Friend (FOAF): an experimental linked information system.

FOAF is a computer language defining a dictionary of people-related terms that can be used in structured data (e.g. [RDFa](#), [JSON-LD](#), [Linked Data](#)).



Status: in stasis

- [FOAF specification](#) - a computer language for people and machines.
- The [original writeup](#) from 2000 explains what we thought we were doing.
- [Linked Data vocabularies using FOAF](#).
- [Linked Data datasets using FOAF](#).
- [Wikipedia entry](#).
- [FOAF for researchers, archivists](#).
- [foaf-dev](#) mailing list.
- [TimBL Linked Data notes](#)
- New stuff: see also [schema.org](#)
- Old: [Guardian article](#), [Galway Workshop](#), [ETech talk](#) (2004)
- Retro: [Dagstuhl talk](#)
- Nearby: [2002 fiction...](#)

Contacts: [Dan Brickley](#) (@danbri), [Libby Miller](#) (@libbymiller)

Vocabolario *Dublin Core*

- Il progetto del Dublin Core (nome completo: **Dublin Core Metadata Initiative**, in acronimo **DCMI**) si è sviluppato in ambito OCLC (**O**n line **C**omputer **L**ibrary **C**enter), la grande rete di servizi americana per le biblioteche.
- Un nucleo di elementi essenziali ai fini della descrizione di qualsiasi materiale digitale accessibile via web (titolo, creatore, etc).
- Nucleo proposto nel dic. 1996, dal 2005 anche in formato RDF.
- <http://dublincore.org/>

The Metadata Community — Supporting Innovation in Metadata Design, Implementation & Best Practices

METADATA
Dublin Core® Metadata Initiative
INNOVATION

join us!

Home Metadata Basics DCMI Specifications Community and Events Join/Support About Us

Enter keyword

 Metadata and Ubiquitous Access to Culture, Science and Digital Humanities • DC 1-4 Sep 2015 São Paulo, Brazil

Chair's
Community Message

Current | 12 January 2015 | 28 November 2014 | 24 October 2014 | 29 August 2014 | 22 July 2014 | 18 June 2014 | 14 May 2014 | 16 April 2014 | 19 March 2014

Announcements

National Diet Library of Japan publishes translations of key DCMI specifications

2015-03-30, DCMI is please to announce that the National Diet Library, the sole national library in Japan, has translated the DCMI [Metadata Terms](#) and the [Singapore Framework for Dublin Core Application Profiles](#) into Japanese. The links to the new Japanese translations, as well as others are available on the DCMI

Vocabolario *GoodRelations*

- <http://www.heppnetz.de/projects/goodrelations/>
- nato nel 2008 e affermatosi come standard de facto nel settore degli e-shop.
- permette di attribuire una descrizione semantica alle informazioni su azienda, prodotti/servizi, prezzi e logistica presenti in un sito di commercio elettronico.



The screenshot shows the GoodRelations website interface. The header is orange with the text "GoodRelations The Web Vocabulary for E-Commerce" and navigation links: Main, Quickstart, Cookbook, User's Guide, Specification, Tools, Community. The main content area features a section titled "Automotive" with a background image of cars. The text describes how manufacturers and dealers can publish rich details of new and used vehicles. A "Read more" button is visible. To the right, there is a paragraph titled "The most powerful Web vocabulary for e-commerce" and another paragraph explaining the benefits of GoodRelations for e-commerce sites.

GoodRelations
The Web Vocabulary for E-Commerce

Main Quickstart Cookbook User's Guide Specification Tools Community

Automotive
Manufacturers and dealers can publish rich details of new and used vehicles so that search engines and browser extensions present them all to your potential customers.
[Read more](#)

The most powerful Web vocabulary for e-commerce
A paradigm shift for e-commerce. Since 2008.

Only 5% of all potential visitors of your site will actually see your offers in their original beauty. 95% will never get beyond a reduced preview of your great products and services as provided by a Web search engine.

GoodRelations is the most powerful vocabulary for publishing all of the details of your products and services in a way friendly to search engines, mobile applications, and browser extensions. By adding a bit of extra code to your Web content, you make sure that potential customers realize all the great features and services and the benefits of doing business with you, because their computers can extract and present this information with ease.

Video



Who uses GoodRelations?

Google
Yahoo!
BestBuy
sears.com
kmart.com

... and 10,000 more

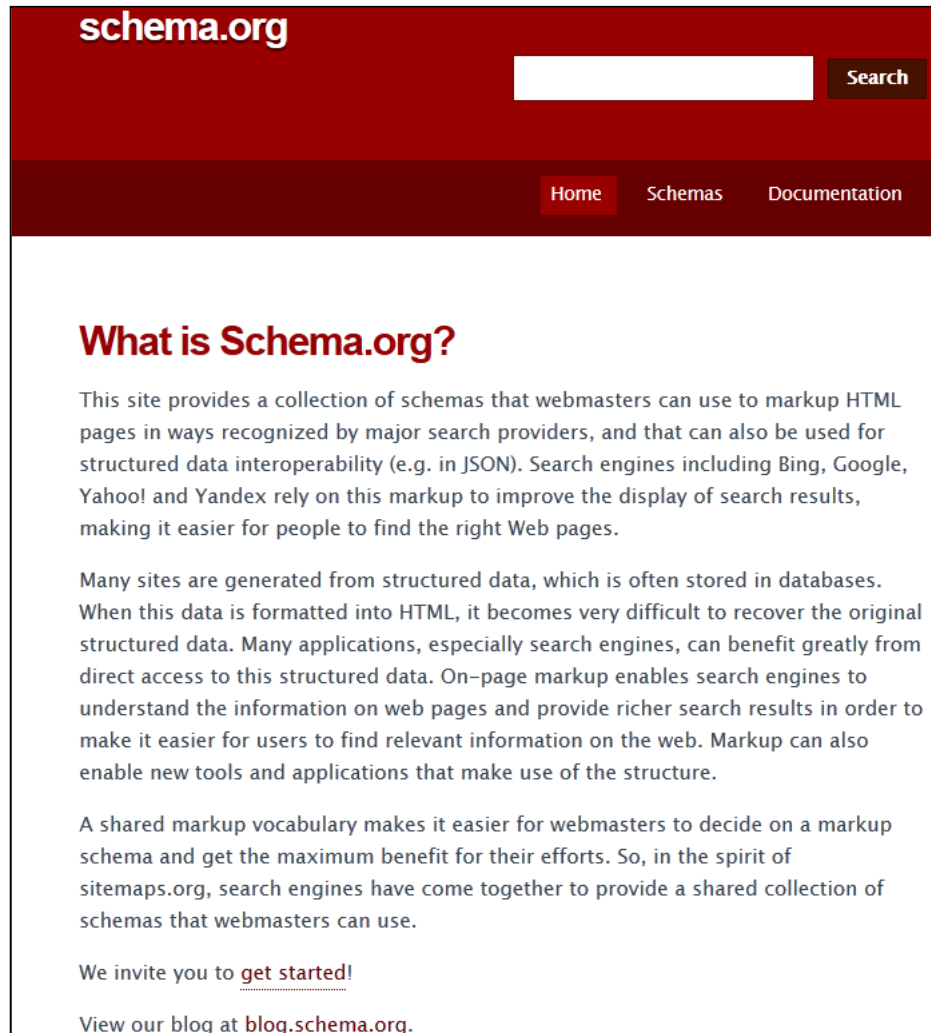
See [here](#) for additional references.

News from Twitter

 **martin hepp** @mfhepp 27 Apr
@TrevorBramble here is background re why #schema.org is strategically important
wiki.goodrelations-vocabulary.org/GoodRelations...
#goodrelations

 **martin hepp** @mfhepp 16 Apr

Schema.org

A screenshot of the Schema.org website. The header is dark red with the 'schema.org' logo in white. Below the logo is a white search bar and a 'Search' button. A navigation bar below the header contains 'Home', 'Schemas', and 'Documentation' links. The main content area is white and features the heading 'What is Schema.org?'. The text explains that the site provides schemas for marking up HTML pages for search engines like Bing, Google, Yahoo!, and Yandex. It also mentions that structured data can be used for JSON and that search engines use this markup to improve search results. A paragraph explains that many sites are generated from structured data in databases, and that formatting this into HTML makes it difficult to recover the original data. It states that on-page markup enables search engines to understand web page information and provide richer search results. Another paragraph notes that a shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit. The page concludes with an invitation to 'get started' and a link to the blog at 'blog.schema.org'.

- Iniziativa lanciata a giugno 2011 da Google, Yahoo e Bing (da novembre 2011 collabora anche Jandex) con obiettivo di definire un vocabolario comune, senza sostituirsi, ma integrando gli strumenti esistenti e stimolando lo sviluppo di pagine web dotate di markup semantico.
- Definisce un insieme standard di schemi (vocabolari) che i webmaster possono utilizzare per formattare le pagine HTML in modo riconosciuto dai principali motori di ricerca, e che può essere utilizzato anche per l'interoperabilità dei dati strutturati (ad esempio in JSON).

Schema.org

schema.org

Home

Organization of Schemas

The schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy.

Browse the full hierarchy:

- [One page per type](#)
- [Full list of types, shown on one page](#)

Or you can jump directly to a commonly used type:

- Creative works: [CreativeWork](#), [Book](#), [Movie](#), [MusicRecording](#), [Recipe](#), [TVSeries](#) ...
- Embedded non-text objects: [AudioObject](#), [ImageObject](#), [VideoObject](#)
- [Event](#)
- Health and medical types: notes on the health and medical types under [MedicalEntity](#).
- [Organization](#)
- [Person](#)
- [Place](#), [LocalBusiness](#), [Restaurant](#) ...
- [Product](#), [Offer](#), [AggregateOffer](#)
- [Review](#), [AggregateRating](#)
- [Action](#)

We also have a small set of [primitive data types](#) for numbers, text, etc. More details about the data model, etc. are available [here](#).

Es. schema.org/MusicEvent

Thing > Event > MusicEvent

Event type: Music event.

Property	Expected Type	Description
Properties from Event		
attendee	Organization or Person	A person or organization attending the event. Supersedes attendees .
doorTime	DateTime	The time admission will commence.
duration	Duration	The duration of the item (movie, audio recording, event, etc.) in ISO 8601 date format .
endDate	Date	The end date and time of the item (in ISO 8601 date format).
eventStatus	EventStatusType	An eventStatus of an event represents its status; particularly useful when an event is cancelled or rescheduled.
location	PostalAddress or Place	The location of the event, organization or action.
offers	Offer	An offer to provide this item—for example, an offer to sell a product, rent the DVD of a movie, or give away tickets to an event.
organizer	Organization or Person	An organizer of an Event.
performer	Organization or Person	A performer at the event—for example, a presenter, musician, musical group or actor. Supersedes performers .
	Date	Used in conjunction with eventStatus for rescheduled or cancelled events. This property contains the previously

Permette descrivere un concerto, con startDate, località ed altre proprietà che descrivono i informazioni principali relativi al concerto.

Es. schema.org/Person

Thing > Person

A person (alive, dead, undead, or fictional).

Property	Expected Type	Description
Properties from <i>Person</i>		
<i>additionalName</i>	Text	An additional name for a Person, can be used for a middle name.
<i>address</i>	PostalAddress	Physical address of the item.
<i>affiliation</i>	Organization	An organization that this person is affiliated with. For example, a school/university, a club, or a team.
<i>alumniOf</i>	EducationalOrganization	An educational organizations that the person is an alumni of. Inverse property: <i>alumni</i> .
<i>award</i>	Text	An award won by this person or for this creative work. Supersedes <i>awards</i> .
<i>birthDate</i>	Date	Date of birth.
<i>birthPlace</i>	Place	The place where the person was born.
<i>brand</i>	Brand or Organization	The brand(s) associated with a product or service, or the brand(s) maintained by an organization or business person.
<i>children</i>	Person	A child of the person.
<i>colleague</i>	Person	A colleague of the person. Supersedes <i>colleagues</i> .
<i>contactPoint</i>	ContactPoint	A contact point for a person or organization. Supersedes <i>contactPoints</i> .
<i>deathDate</i>	Date	Date of death.
<i>deathPlace</i>	Place	The place where the person died.
<i>duns</i>	Text	The Dun & Bradstreet DUNS number for identifying an organization or business person.
<i>email</i>	Text	Email address.
<i>familyName</i>	Text	Family name. In the U.S., the last name of an Person. This can be used along with givenName instead of the name property.
<i>faxNumber</i>	Text	The fax number.
<i>follows</i>	Person	The most generic uni-directional social relation.
<i>gender</i>	Text	Gender of the person.
<i>givenName</i>	Text	Given name. In the U.S., the first name of a Person. This can be used along with familyName instead of the

Markup dei dati strutturati

Google e altri motori di ricerca supportano il vocabolario *schema.org* per i dati strutturati, che possono essere incorporati in una pagina HTML utilizzando uno qualsiasi dei tre formati alternativi: *microdati*, *RDFa*, e *JSON-LD*.

- *Microdata* e *RDFa* definiscono nuovi attributi HTML che consentono di indicare la corrispondenza tra i nomi dei campi definiti nello *schema.org* e il testo visibile all'utente nella pagina.
- *JSON-LD (JSON for Linked Data)* è il più nuovo e il più semplice formato di markup: esso consente di incorporare un blocco di dati JSON all'interno di un tag script in qualsiasi parte del codice HTML. Poiché i dati non devono essere intercalati con il testo visibile all'utente, permette di esprimere molto più facilmente elementi di dati nidificati (ad es., il Paese di un *PostalAddress* di un *MusicVenue* di un evento). Inoltre, Google può leggere i dati JSON-LD, anche quando viene iniettato in modo dinamico nel contenuto della pagina, come ad esempio per codice Javascript.

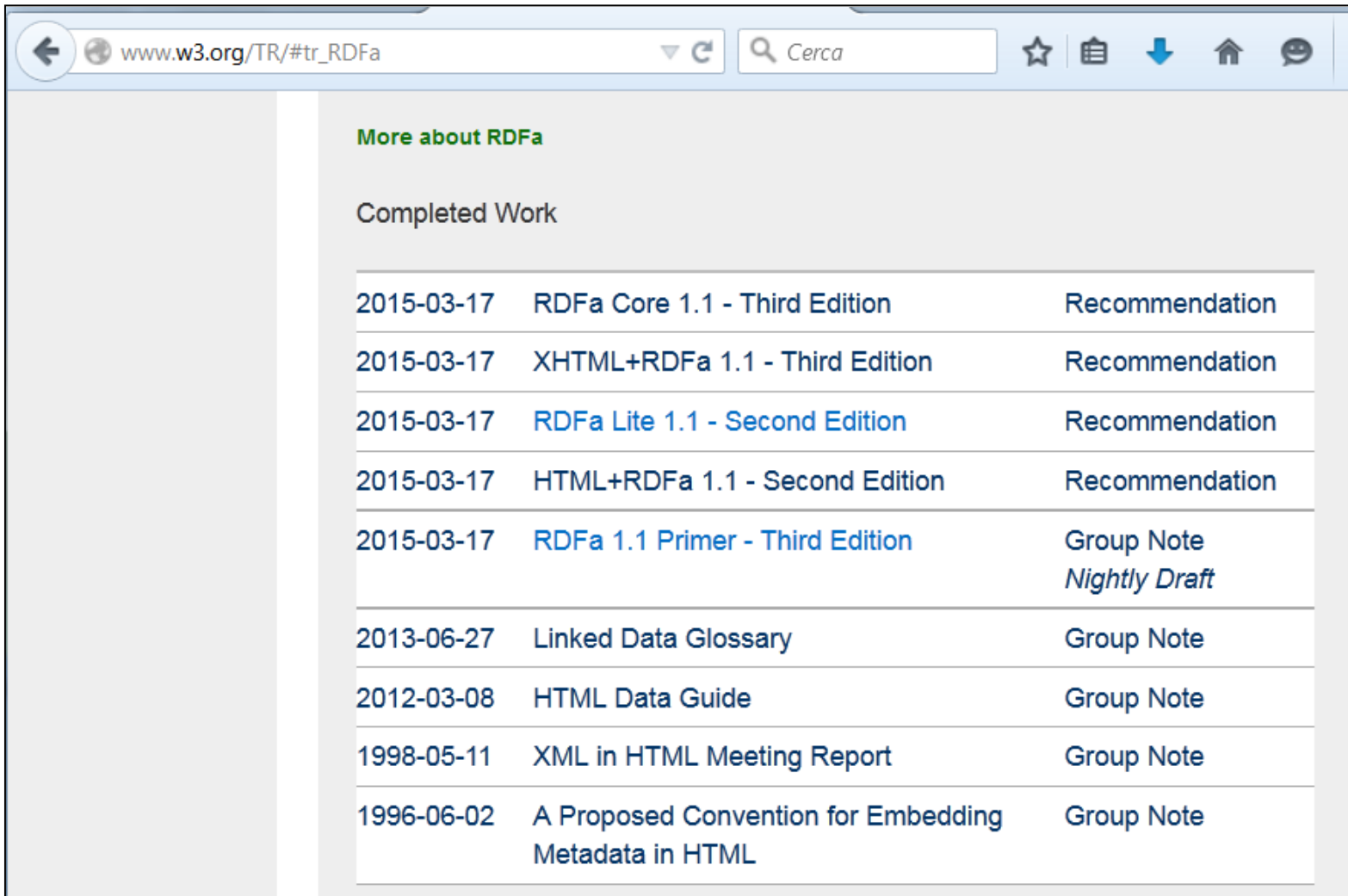
Google è in procinto di aggiungere il supporto JSON-LD a più funzioni di markup alimentato. Finora, JSON-LD è supportata per tutte le funzioni Knowledge Graph, sitelink search boxes e Event Rich Snippets; Google consiglia l'uso di JSON-LD per quelle caratteristiche.

Per i restanti tipi di Rich Snippets e breadcrumbs, Google consiglia l'uso di microdati o RDFa.

RDFa

- *RDFa (Resource Description Framework in attributes).*
- **RDFa** (RDF incorporato in HTML). Permette di incorporare dati RDF all'interno di normali pagine Web ai fini di esprimere concetti semantici associati ad elementi di una pagina XHTML.
- Es. di utilizzo: Best Buy, il rivenditore al dettaglio di elettronica di consumo negli Stati Uniti (21% del mercato), operante anche in Canada, Messico e in Cina.
- Arricchisce la sintassi XHTML attraverso una serie di attributi aggiuntivi che, associati ai tag già noti, simulano le triple RDF.
- Versione base: RDFa Lite
 - 5 semplici attributi: *vocab, typeof, property, resource, prefix*
- Versione completa: RDFa Core
 - Altri attributi: *about, content, datatype, href, inlist, rel, rev, src*

RDFa – situazione attuale



The screenshot shows a web browser window with the address bar displaying 'www.w3.org/TR/#tr_RDFa'. The page content includes a section titled 'More about RDFa' and a table of 'Completed Work'. The table lists various RDFa specifications and documents, including 'RDFa Core 1.1 - Third Edition', 'XHTML+RDFa 1.1 - Third Edition', 'RDFa Lite 1.1 - Second Edition', 'HTML+RDFa 1.1 - Second Edition', 'RDFa 1.1 Primer - Third Edition', 'Linked Data Glossary', 'HTML Data Guide', 'XML in HTML Meeting Report', and 'A Proposed Convention for Embedding Metadata in HTML'.

More about RDFa		
Completed Work		
2015-03-17	RDFa Core 1.1 - Third Edition	Recommendation
2015-03-17	XHTML+RDFa 1.1 - Third Edition	Recommendation
2015-03-17	RDFa Lite 1.1 - Second Edition	Recommendation
2015-03-17	HTML+RDFa 1.1 - Second Edition	Recommendation
2015-03-17	RDFa 1.1 Primer - Third Edition	Group Note <i>Nightly Draft</i>
2013-06-27	Linked Data Glossary	Group Note
2012-03-08	HTML Data Guide	Group Note
1998-05-11	XML in HTML Meeting Report	Group Note
1996-06-02	A Proposed Convention for Embedding Metadata in HTML	Group Note

RDFA editor

rdfa.info/play/



WARNING: This is beta software, there are [bugs](#). This tool is not yet ready for production use. That said, it's pretty useful already, and the bugs preventing production use will be sorted out soon.

Examples: [Person](#) [Social Network](#) [Event](#) [Place](#) [Product](#) [SVG](#)

```
<div prefix="
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
foaf: http://xmlns.com/foaf/0.1/
gr: http://purl.org/goodrelations/v1#
xsd: http://www.w3.org/2001/XMLSchema#"
typeof="gr:Offering">
<div>
<h1 property="gr:name">Canon Rebel T2i (EOS 550D)</h1>
<div rel="foaf:depiction">

The Canon Rebel T2i (EOS 550D) is Canon's top-of-the-line consumer digital SLR
It can shoot up to 18 megapixel resolution photos and features an ISO range of
</p>
<link rel="gr:hasBusinessFunction" href="http://purl.org/goodrelations/v1#Sell" />
<meta property="gr:hasEAN_UCC-13" content="013803123784" />
Sale price:
<span property="gr:hasPriceSpecification" typeof="gr:UnitPriceSpecification">
<span property="gr:hasCurrency" content="USD">$</span>
<span property="gr:hasCurrencyValue" datatype="xsd:float">899</span>
</div>
```

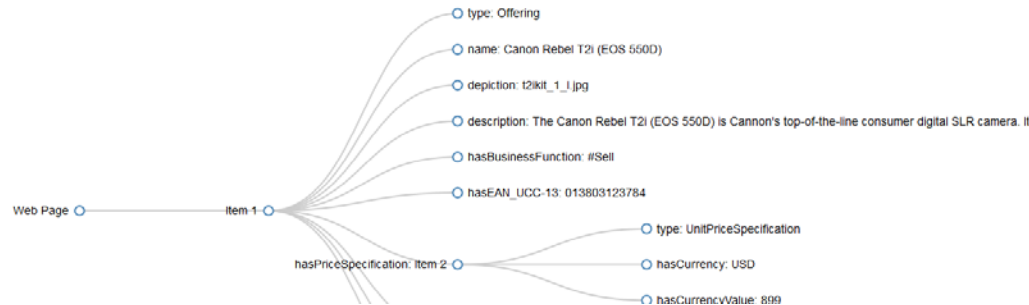
Canon Rebel T2i (EOS 550D)



The Canon Rebel T2i (EOS 550D) is Canon's top-of-the-line consumer digital SLR camera. It can shoot up to 18 megapixel resolution photos and features an ISO range of 100-6400.

Sale price: \$ 899 [\[more...\]](#)

Visualization [Raw Data](#)



RDFa Lite: attributo *vocab*

Es.1 *Specifichiamo il vocabolario che useremo*

```
<p vocab="http://schema.org/">
```

My name is Manu Sporny and you can give me a ring via 1-800-555-0199.

```
</p>
```

The screenshot shows a web browser window with the address bar displaying 'rdfa.info/play/'. The page has a dark navigation bar with links: 'RDFa', 'Play', 'Documentation', 'Tools', and 'Developers'. Below the navigation bar is a yellow warning box that reads: 'WARNING: This is beta software, there are bugs. This tool is not yet ready for production use. That said, it's pretty useful already, and the bugs preventing production use will be sorted out soon.' Under the warning box is a row of tabs: 'Examples:', 'Person', 'Social Network', 'Event', 'Place', 'Product', and 'SVG'. The 'Examples:' tab is selected. The main content area is divided into two columns. The left column contains the RDFa code:

```
<p vocab="http://schema.org/">
My name is Manu Sporny and you can give me a ring via 1-800-555-0199.
</p>
```

 The right column displays the rendered HTML: 'My name is Manu Sporny and you can give me a ring via 1-800-555-0199.' At the bottom of the main content area are two tabs: 'Visualization' and 'Raw Data'. The 'Raw Data' tab is selected. In the bottom left corner of the page, there is a 'Web Page' icon and a small circular icon.

RDFa Lite: attributo *typeof*

Es.2 *Specifichiamo il tipo della risorsa che descriviamo*

```
<p vocab="http://schema.org/" typeof="Person">
```

My name is Manu Sporny and you can give me a ring via 1-800-555-0199.

```
</p>
```

Web Page ○ ————— Item 1 ○ ————— ○ type: Person

RDFa Lite: attributo *property*

Es.3 *Specifichiamo le proprietà della risorsa che descriviamo*

```
<p vocab="http://schema.org/" typeof="Person">
```

My name is

```
<span property="name">Manu Sporny</span>
```

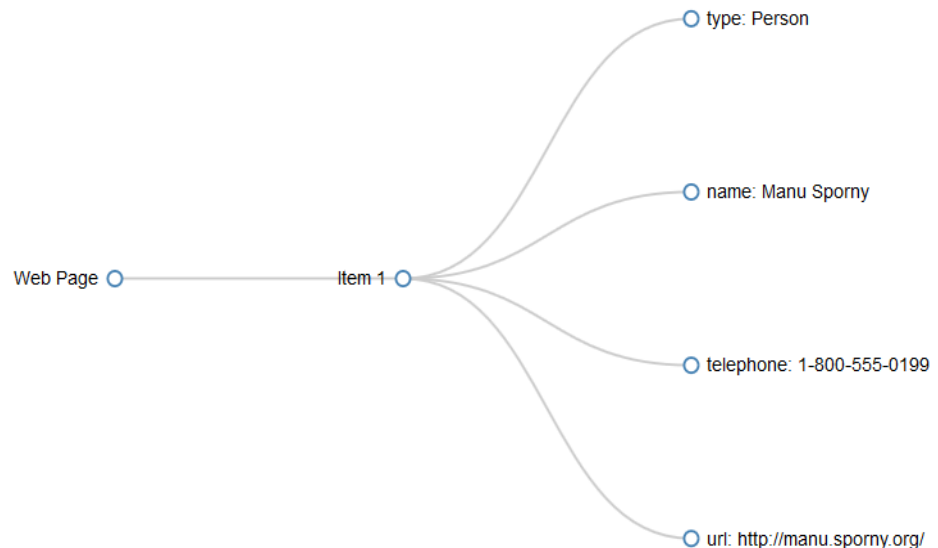
and you can give me a ring via

```
<span property="telephone">1-800-555-0199</span>
```

or visit

```
<a property="url" href="http://manu.sporny.org/">my homepage</a>.
```

```
</p>
```



RDFa Lite: attributo *resource*

Es.4 *Identifichiamo la risorsa che descriviamo*

```
<p vocab="http://schema.org/" resource="#manu" typeof="Person">
```

My name is

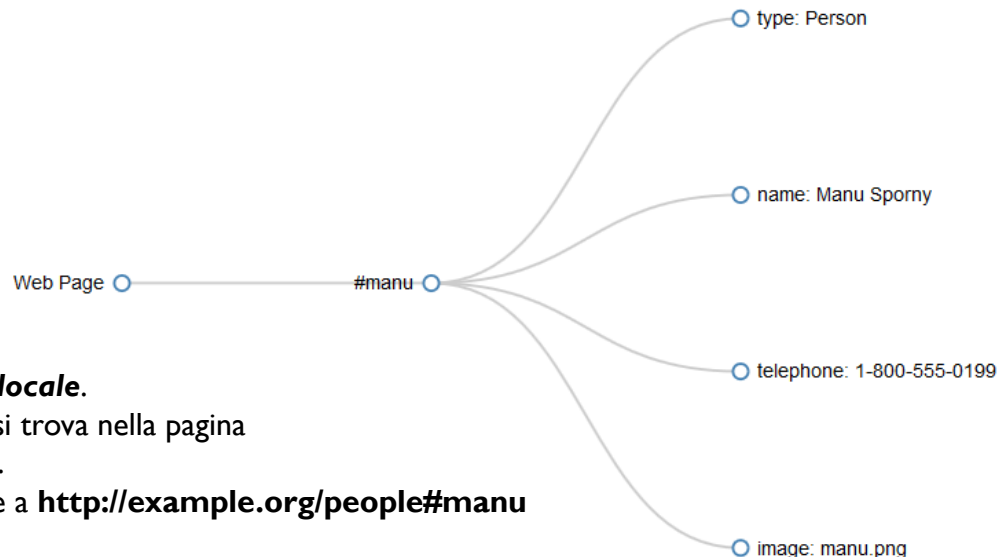
```
<span property="name">Manu Sporny</span>
```

and you can give me a ring via

```
<span property="telephone">1-800-555-0199</span>.
```

```

</p>
```



#manu è identificatore locale.

Supponiamo che markup si trova nella pagina

<http://example.org/people>.

Allora #manu corrisponde a **<http://example.org/people#manu>**

RDFa Lite: attributo *prefix*

Es.5 Assegnamo prefisso Ov ad OpenVocabulary

```
<p vocab="http://schema.org/" prefix="ov: http://open.vocab.org/terms/" resource="#manu"  
typeof="Person">
```

My name is

```
<span property="name">Manu Sporny</span>
```

and you can give me a ring via

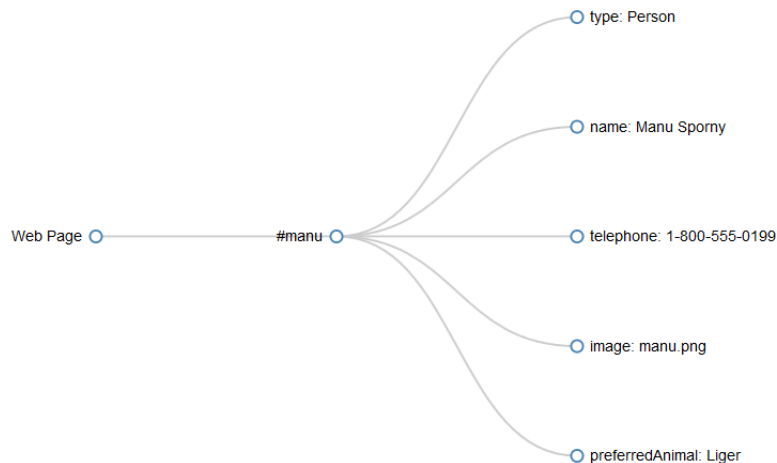
```
<span property="telephone">1-800-555-0199</span>.
```

```

```

My favorite animal is the Liger.**

```
</p>
```



RDFa Lite: proprietà "ripetute"

Es.6 La stessa proprietà può essere espressa utilizzando diversi vocabolari:

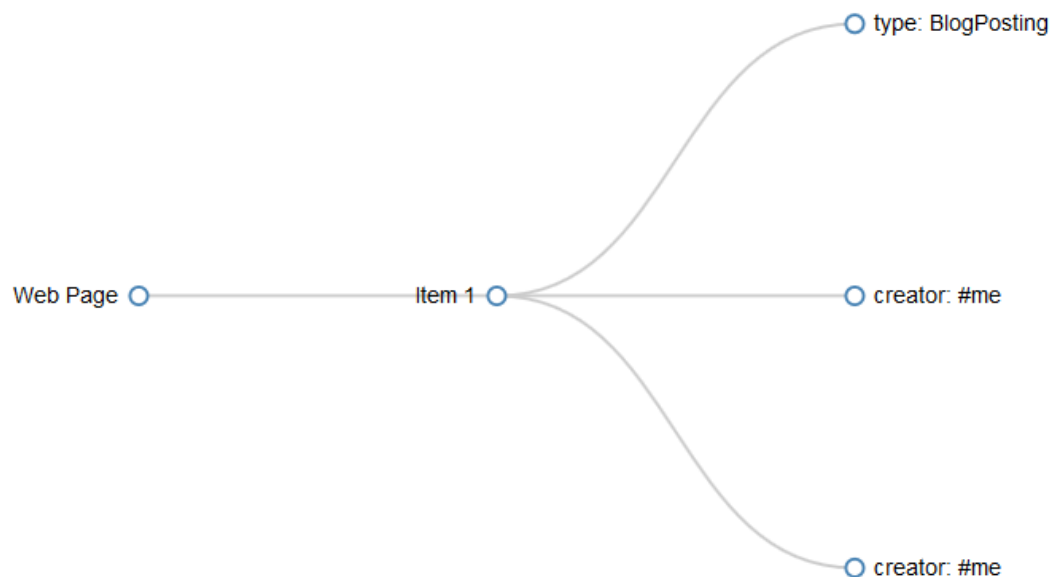
```
<body prefix="dc: http://purl.org/dc/terms/  schema: http://schema.org/">
```

```
<div typeof="schema:BlogPosting">
```

```
<h1 property="dc:creator schema:creator" resource="#me">Irina</h3>
```

```
</div>
```

```
</body>
```



Prefissi predefiniti

Widely used Vocabulary prefixes based on the vocabulary usage on the Semantic Web			
Prefix	URI	Description	For more details see
cc	http://creativecommons.org/ns#	ccREL	ccREL: The Creative Commons Rights Expression Language
ctag	http://commontag.org/ns#	Common Tag Ontology	Common Tag Specification
dc	http://purl.org/dc/terms/	Dublin Core Metadata Terms	DCMI Metadata Terms
dcterms	http://purl.org/dc/terms/	Dublin Core Metadata Terms	DCMI Metadata Terms
dc:1	http://purl.org/dc/elements/1.1/	Dublin Core Metadata Element Set, Version 1.1	Dublin Core Metadata Element Set, Version 1.1
foaf	http://xmlns.com/foaf/0.1/	FOAF	FOAF Vocabulary Specification
gr	http://purl.org/goodrelations/v1#	GoodRelations Ontology	GoodRelations Language Reference
ical	http://www.w3.org/2002/12/cal/icaltzd#	iCalendar terms in RDF	iCalendar terms in RDF
og	http://ogp.me/ns#	Facebook's Open Graph protocol	Open Graph Protocol
rev	http://purl.org/stuff/rev#	RDF Review Vocabulary	RDF Review Vocabulary
sioc	http://rdfs.org/sioc/ns#	SIOC Core Ontology	SIOC Core Ontology Specification
v	http://rdf.data-vocabulary.org/#	Google Rich Snippets' Vocabularies	Rich snippets (microdata, microformats, and RDFa)
vcard	http://www.w3.org/2006/vcard/ns#	vCard in RDF	Representing vCard Objects in RDF
schema	http://schema.org/	The Schema.org vocabulary	Full Schema.org documentation

Microdata

- Microdata è una specifica per incorporare i dati machine-readable in documenti HTML, introdotta dal HTML 5.
- Ad alto livello, i microdati consistono di un gruppo di coppie *nome-valore* definiti secondo un vocabolario.
 - I gruppi sono chiamati *item* (elementi),
 - Ogni coppia *nome-valore* è una proprietà.
- La sintassi di base include gli attributi:
 - *itemscope* per definire un elemento (item)
 - *itemtype* per specificare il tipo dell'elemento
 - *itemprop* per descrivere una delle proprietà dell'elemento.
- Altri attributi:
 - *itemid* per specificare l'identificatore globale dell'elemento
 - *itemref* – costrutto sintattico (non fa parte del *microdata model*)
permette di associare ad un item le proprietà definiti al di fuori del suo tag

Microdata: attributo *itemscope*

<section >

Hello, my name is John Doe, I am a Graduate research assistant at the University of Dreams. My friends call me Johnny. You can visit my homepage at www.example.com

</section>

identifico la sezione della pagina che descrive JohnDoe

<section **itemscope** itemtype="http://schema.org/Person">

Hello, my name is

John Doe,

I am a

Graduate research assistant

at the

University of Dreams

My friends call me

Johnny

You can visit my homepage at

www.example.com

</section>

Microdata: attributo *itemtype*

<section >

Hello, my name is John Doe, I am a Graduate research assistant at the University of Dreams. My friends call me Johnny. You can visit my homepage at www.example.com

</section>

<section itemscope **itemtype="http://schema.org/Person"**>

Hello, my name is

John Doe,

I am a

Graduate research assistant

at the

University of Dreams

My friends call me

Johnny

You can visit my homepage at

www.example.com

</section>

Specifico che item contenuto in *section* è una persona, come definito nella gerarchia di schema.org

Microdata: attributo *itemprop*

<section >

Hello, my name is John Doe, I am a Graduate research assistant at the University of Dreams. My friends call me Johnny. You can visit my homepage at www.example.com

</section>

<section itemscope itemtype="http://schema.org/Person">

Hello, my name is

John Doe,

I am a

Graduate research assistant

at the

University of Dreams

My friends call me

Johnny

You can visit my homepage at

www.example.com

</section>

descrivo le
proprietà

Microdata: attributo *itemprop*

<section itemscope itemtype="http://schema.org/Person">

Hello, my name is

****John Doe****,

I am a

****graduate research assistant****

at the

****University of Dreams****.

My friends call me

****Johnny****.

You can visit my homepage at

****www.JohnnyD.com****.

<section itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">

I live at

****1234 Peach Drive****,

****Warner Robins****,

****Georgia****.

</section>

</section>

Il valore della proprietà *name*
è il contenuto testuale del tag,
cioè John Doe

Il valore della proprietà *url* è il
valore di href, cioè
<http://www.JohnnyD.com>

Il valore della proprietà *address*
è un item di tipo *Postal Address*

Microdata: attributo *itemprop*

Valore della proprietà:

- **Item creato dall'elemento** se esiste un attributo *itemscope*
Es. precedente - item che descrive l'indirizzo postale di John
- **URL** nel caso degli elementi (tag):
a, area, link (valore dell'attributo *href*)
audio, embed, iframe, img, source, track, video (valore dell'attributo *src*)
objects (valore dell'attributo *time*)

Es. `<a href="http://www.example.com" itemprop="url"> www.example.com </a>`
Es. `<video src="esempio.webm" width="720" height="576" poster="immagine.jpg" controls autoplay loop preload="auto" muted itemprop="video" >`
Questo testo viene visualizzato se il browser non supporta l'HTML5.
`</video>`
- **Valore dell'attributo *value*** per gli elementi (tag) *data* e *meter* e dell'attributo *datetime* per *time*:
Es. `<time itemprop="startDate" datetime="2015-05-20T10:30">May 20, 10:30am</time>`
Es. `<meter itemprop="ratingValue" min=0 value=4 max=5> Rated 4/5 </meter>`
- **Valore dell'attributo *content*** per l'elemento *meta*:
Es. `<meta itemprop="birthDate" content="1961-08-05">`
- **Contenuto testuale** del tag, altrimenti:
Es. `<span itemprop="name">John Doe</span>`

Microdata: *itemid*

- Definisce l'identificatore globale di un item
- L'elemento che ha l'attributo `itemid` deve anche avere gli attributi `itemscope` e `itemtype`
- Il vocabolario specificato in `itemtype` deve supportare gli identificatori globali degli item
- Es.

```
<dl itemscope
  itemtype="http://vocab.example.net/book"
  itemid="urn:isbn:0-330-34032-8">
  <dt>Title
  <dd itemprop="title">The Reality Dysfunction
  <dt>Author
  <dd itemprop="author">Peter F. Hamilton
  <dt>Publication date
  <dd><time itemprop="pubdate" datetime="1996-01-26">26 January 1996</time>
</dl>
```

Microdata: *itemref*

- Costrutto sintattico (non fa parte del *microdata model*)
- Permette di associare ad un item le proprietà definiti al di fuori del suo tag specificando l'elenco di identificatori (**id**) dei tag contenenti le proprietà, separati con lo spazio.
- Es.

```
<div itemscope id="amanda" itemref="a b"></div>
```

```
<p id="a">Name: <span itemprop="name">Amanda</span></p>
```

```
<div id="b" itemprop="band" itemscope itemref="c"></div>
```

```
<div id="c">
```

```
<p>Band: <span itemprop="name">Jazz Band</span></p>
```

```
<p>Size: <span itemprop="size">12</span> players</p>
```

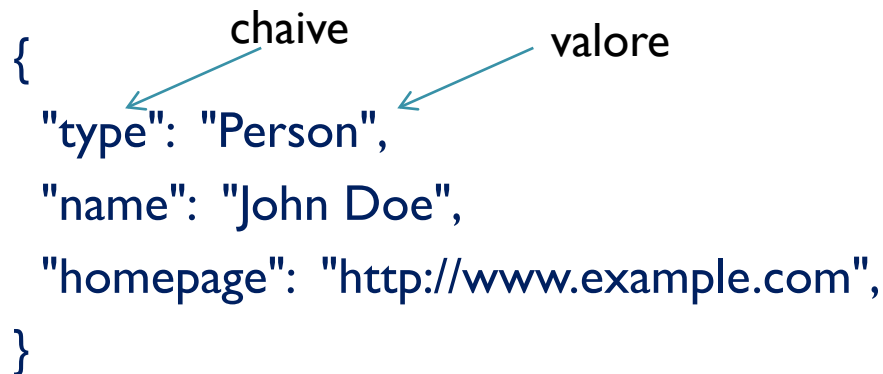
```
</div>
```

JSON

- *JSON (JavaScript Object Notation)*
- un formato adatto all'interscambio di dati fra applicazioni client-server
 - Semplicità d'utilizzo
 - Sia *machine-readable* che *human-readable*

Es.

```
{  
  "type": "Person",  
  "name": "John Doe",  
  "homepage": "http://www.example.com",  
}
```



- Problema: ambiguità (quando i dati arrivano da sorgenti diversi)

JSON-LD

- *JSON-LD (Javascript Object notation for Linked Data)*
- JSON-LD è un formato per la serializzazione di *Linked Data* basato su JSON
- Può essere incorporato nei documenti HTML dentro `<script type="application/ld+json">`

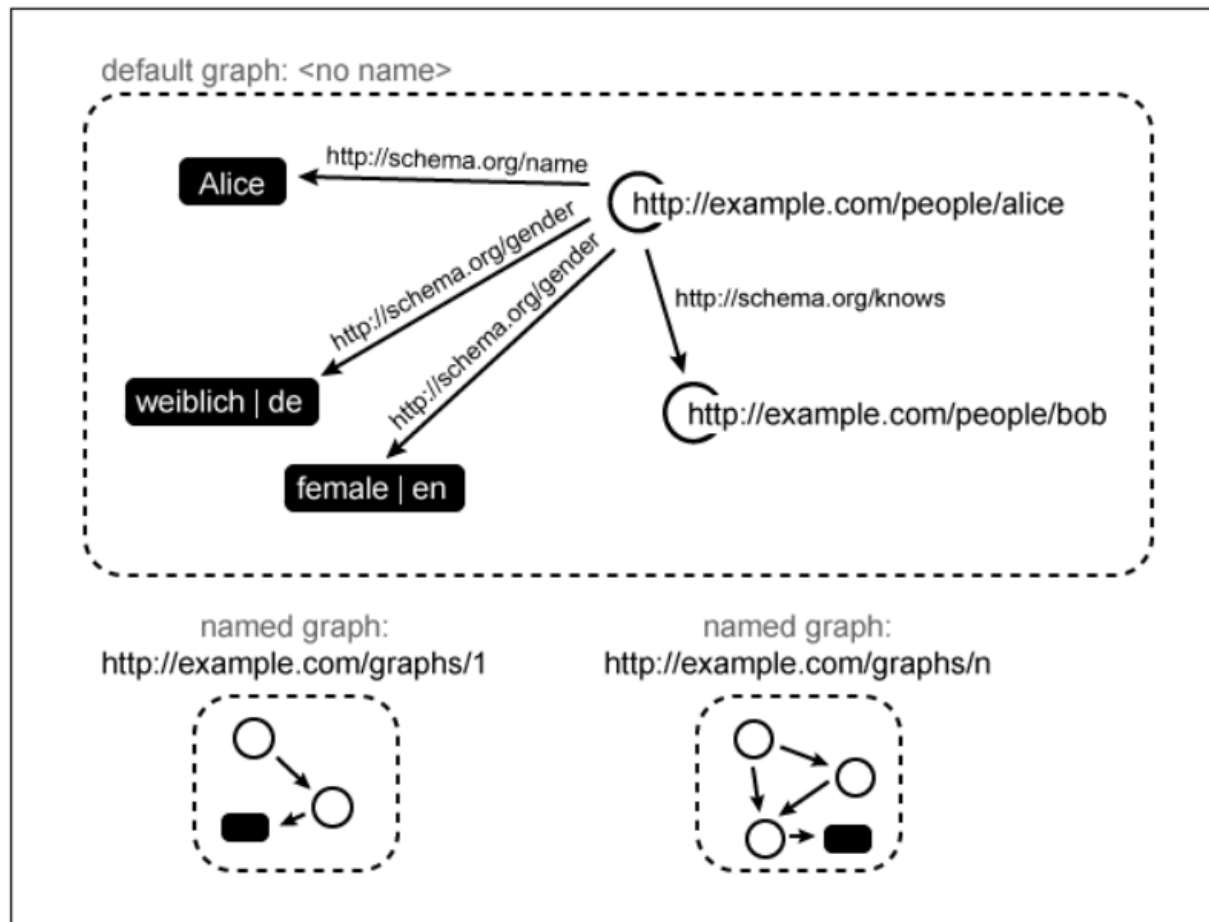
Es. `<script type="application/ld+json">`

```
{
  "@context": "http://schema.org",
  "@type": "Person",
  "name": "John Doe",
  "jobTitle": "Graduate research assistant",
  "affiliation": "University of Dreams",
  "additionalName": "Johnny",
  "url": "http://www.example.com",
  "address": {
    "@type": "PostalAddress",
    "streetAddress": "1234 Peach Drive",
    "addressLocality": "Wonderland",
    "addressRegion": "Georgia"
  }
}
```

`</script>`

JSON-LD: modello dei dati

JSON-LD document



Snippet

Se osserviamo la **SERP** (Search Engine Result Page) di Google, noteremo che i blocchi di risultati, che rappresentano le pagine che soddisfano la ricerca dell'utente, sono composti da:

- **Titolo:** l'elemento più visibile e che corrisponde al tag “*title*” della pagina;
- **URL:** in verde, rappresenta l'URL della pagina;
- **Descrizione:** poche righe di testo che hanno il compito di introdurre il contenuto della pagina.

Pasta e fagioli - Wikipedia

it.wikipedia.org/wiki/Pasta_e_fagioli ▼

La **pasta e fagioli** è un piatto tipico italiano di cui esistono diverse varianti regionali. Gli ingredienti possono tipicamente comprendere pasta corta, fagioli, aglio, ...

Lo **snippet** è proprio la descrizione della pagina web. Tale elemento gioca un ruolo importantissimo all'interno della SERP perché, oltre a fornire un “*assaggio*” della pagina, ci comunica il motivo per il quale il risultato è pertinente alla nostra ricerca.

Rich snippet

Ristorante Al Girone dei Golosi

4,4 ★★★★★ 37 recensioni - Ristorante

Tunnel Pasquale Rossi - Frittura di mare e carni in posto sobrio

Aperto alle 20:00



Pasta e fagioli alla veneta | Sale&Pepe



www.salepepe.it > Ricette > Primi > Pasta fresca > Maltagliati ▾

2 h 5 min - Calorie: 430

Una minestra deliziosa dal sapore antico: prova la **pasta e fagioli** alla veneta secondo la semplice ricetta di Sale&Pepe.

Angry Birds - App Android su Google Play

<https://play.google.com/store/apps/details?id=com.rovio.angrybirds...> ▾

★★★★★ Valutazione: 4,5 - 4.092.739 voti - Gratis

Usa i poteri unici degli **Angry Birds** per distruggere le difese degli avidi maiali! È in gioco la sopravvivenza stessa degli **Angry Birds**. Servi la tua vendetta agli avi.

Fornendo informazioni aggiuntive nelle pagine Web possiamo arricchire gli snippets:

Rich Snippet

Google supporta Rich Snippets per i seguenti tipi di dati:

- *Prodotto* - informazioni su un prodotto, compreso il prezzo, la disponibilità e voti delle recensioni.
- *Ricetta* - ricette che possono essere visualizzati nelle ricerche web e *Recipe View*.
- *Recensione* - Una revisione di un oggetto, come un ristorante, cinema, o negozio.
- *Evento* - Un evento organizzato, come concerti musicali o festival d'arte, ai quali persone possono partecipare in un determinato momento e luogo.
- *SoftwareApplication* - Informazioni su una applicazione software, compresa la URL, voti delle recensioni e il prezzo.

Abilitare Rich Snippet per Applicazioni Software: markup in *JSON-LD*

```
<script type="application/ld+json">
{
  "@context": "http://schema.org/",
  "@type": "SoftwareApplication",
  "name": "Angry Birds",
  "operatingSystem": "ANDROID",
  "applicationCategory": "http://schema.org/GameApplication",
  "aggregateRating": {
    "@type": "AggregateRating",
    "ratingValue": "4.6",
    "ratingCount": "8864"
  },
  "offers": {
    "@type": "Offer",
    "price": "1.00",
    "priceCurrency": "USD"
  }
}
</script>
```

Abilitare Rich Snippet per Applicazioni Software: markup in *microdata*

```
<div itemscope itemtype="http://schema.org/SoftwareApplication">  
  <span itemprop="name">Angry Birds</span> -
```

```
  REQUIRES <span itemprop="operatingSystem">ANDROID</span>  
  <link itemprop="applicationCategory" href="http://schema.org/GameApplication"/>
```

RATING:

```
  <div itemprop="aggregateRating" itemscope  
    itemtype="http://schema.org/AggregateRating">  
    <span itemprop="ratingValue">4.6</span> (  
    <span itemprop="ratingCount">8864</span> ratings )  
  </div>
```

```
  <div itemprop="offers" itemscope itemtype="http://schema.org/Offer">  
    Price: $<span itemprop="price">1.00</span>  
    <meta itemprop="priceCurrency" content="USD" />  
  </div>  
</div>
```

Abilitare Rich Snippet per Applicazioni Software: markup in *RDFa*

```
<div vocab="http://schema.org/" typeof="SoftwareApplication">  
  <span property="name">Angry Birds</span> -
```

```
  REQUIRES <span property="operatingSystem">ANDROID</span>  
  <link property="applicationCategory" href="http://schema.org/GameApplication"/>
```

RATING:

```
<div property="aggregateRating" typeof="AggregateRating">  
  <span property="ratingValue">4.6</span> (  
    <span property="ratingCount">8864</span> ratings )  
</div>
```

```
<div property="offers" typeof="Offer">  
  Price: $<span property="price">1.00</span>  
  <meta property="priceCurrency" content="USD" />  
</div>  
</div>
```

Knowledge Graph

- Il Knowledge Graph è il sistema di Google per organizzare le informazioni su milioni di "entità" ben noti: le persone, i luoghi e le organizzazioni nel mondo reale.
 - Algoritmi di Google si fondono informazioni su soggetti provenienti da molte fonti di dati. Per alcuni tipi di informazioni, però, la migliore fonte di dati è l'entità stessa.
- Attualmente è possibile personalizzare i propri dati nel Knowledge Graph. In particolare, possono essere personalizzati i seguenti tipi di dati :
 - Logos
 - Company Contact Numbers
 - Social Profile Links

Personalizzazione di Logo

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "Organization",
  "url": "http://www.example.com",
  "logo": "http://www.example.com/images/logo.png"
}
</script>
```

Personalizzazione di Contatti

```
<script type="application/ld+json">
{ "@context" : "http://schema.org",
  "@type" : "Organization",
  "url" : "http://www.your-company-site.com",
  "contactPoint" : [
    { "@type" : "ContactPoint",
      "telephone" : "+1-401-555-1212",
      "contactType" : "customer service"
    }
  ]
}
</script>
```

Personalizzazione di Profilo

JSON_LD template per un'azienda:

```
<script type="application/ld+json">
{ "@context" : "http://schema.org",
  "@type" : "Organization",
  "name" : "Your Organization Name",
  "url" : "http://www.your-site.com",
  "sameAs" : [ "http://www.facebook.com/your-profile",
    "http://www.twitter.com/yourProfile",
    "http://plus.google.com/your_profile" ]
}
</script>
```

- Facebook
- Twitter
- Google+
- Instagram
- YouTube
- LinkedIn
- Myspace

Personalizzazione di Profilo

JSON_LD template per una persona:

```
<script type="application/ld+json">
{ "@context" : "http://schema.org",
  "@type" : "Person",
  "name" : "your name",
  "url" : "http://www.your-site.com",
  "sameAs" : [ "http://www.facebook.com/your-profile",
               "http://instagram.com/yourProfile",
               "http://www.linkedin.com/in/yourprofile",
               "http://plus.google.com/your_profile" ]
}
</script>
```

- Facebook
- Twitter
- Google+
- Instagram
- YouTube
- LinkedIn
- Myspace

Structured Data Testing Tool

The screenshot shows the Google Developers Structured Data Testing Tool interface. The browser address bar displays the URL `https://developers.google.com/structured-data/testing-tool/`. The page header includes the Google Developers logo, a search bar with the text "Testing ..X", and a user profile icon for "i.trubitsyna@gmail.com". The breadcrumb navigation shows "Prodotti > Search > Structured Data > Testing Tool". The main heading is "Testing Tool" with a Google+ badge showing 3,809 likes. Below the heading, there are tabs for "Recupera URL", "Esempi", and a red "CONVALIDA" button. The "Esempi" tab is selected, showing a JSON-LD snippet for an Organization. The right panel, titled "Risultati", shows a single result for "Organization (1)" with the status "Nessun problema" and a green checkmark. Below this, a table lists the properties and values for the Organization.

Prodotti > Search > Structured Data > Testing Tool

Testing Tool 3.809

Recupera URL Esempi **CONVALIDA**

Risultati - Filtra per caso di utilizzo

```
1 <script type="application/ld+json">
2 { "@context" : "http://schema.org",
3   "@type" : "Organization",
4   "name" : "Your Organization Name",
5   "url" : "http://www.your-site.com",
6   "sameAs" : [ "http://www.facebook.com
7                 /your-profile",
8                 "http://www.twitter.com/yourProfile",
9                 "http://plus.google.com
10                /your_profile"]
11 }
```

Organization (1) **Nessun problema** ✓

Organization	
name:	Your Organization Name
url:	http://www.your-site.com
sameAs:	http://www.facebook.com/your-profile
sameAs:	http://www.twitter.com/yourProfile
sameAs:	http://plus.google.com/your_profile

Conclusione

Due tendenze all'utilizzo delle tecnologie del Web Semantico:

- Applicare tali tecnologie all'intelligenza artificiale
 - Fornire migliori risultati di ricerca
 - Inferire nuove informazioni
- Flessibilità del modello dei dati
 - Particolarmente adatto per i dati non strutturati es. documenti (Word, Pdf, Excel, etc.), le pagine Web, pubblicazioni scientifiche, notizie, etc.
 - Quando il modello dei dati o del loro utilizzo non può essere prevedibile apriori.