

# **Dal Web of Documents al Web of Things: le tecnologie semantiche come opportunità di crescita e sviluppo**

**W3C ITALIA**  
**Oreste Signore**

**CROSS KM**

## **Semantic Web – Internet of Things**

### **Sommario**

Open Data, Smart Cities, IoT, Industry 4.0 sono linee progettuali che promettono di modificare radicalmente la nostra esistenza, e utilizzano in maniera efficace le tecnologie del (Semantic) Web nel contesto della Open Web Platform. Partecipare alla definizione degli standard web e restare aggiornati sulle sue tecnologie è un'opportunità di crescita tecnologica, occupazionale ed economica a livello di singolo e di impresa.

### **1 Introduzione**

Il Web è una realtà pervasiva, ormai così inserita nella nostra vita quotidiana, che spesso non si fa caso alla tecnologia sottostante. Molte iniziative progettuali, che promettono di avere un impatto significativo sul nostro modo di vivere, utilizzano in maniera sostanziale tecnologie sviluppate nel contesto del Web. Tra queste, quelle del Semantic Web costituiscono uno degli elementi fondamentali, e spesso ignorato da chi ritiene che il Semantic Web sia solo un argomento puramente accademico.

In questo lavoro si richiamano per prima cosa i principi fondamentali del Web, facendo riferimento alla sua genesi e mettendo in evidenza i suoi principi architetturali. Successivamente vengono richiamati alcuni concetti di base del Semantic Web. Infine, partendo dalla descrizione dell'Internet of Things, argomento che richiama al momento l'interesse dei ricercatori e delle imprese, viene descritta la linea di azione del W3C per la realizzazione del Web of Things, inteso come la piattaforma e il complesso di tecnologie che promette di superare gli ostacoli che si frappongono ad un pieno sviluppo dell'Internet of Things.

### **2 Il Web: questo (s)conosciuto**

Per comprendere appieno il Web è bene far riferimento alla sua storia, per comprenderne i principi fondanti e le linee di sviluppo.

Il Web è nato al CERN, ambiente di punta nella ricerca fondamentale, come strumento per condividere la conoscenza e favorire l'interazione interpersonale. Caratteristiche del Web sono l'interoperabilità e l'accessibilità. L'interoperabilità è forse da ritenere la causa prima che portò alla sua invenzione, a seguito della frustrazione determinata dal dover condividere dati dispersi nelle varie istituzioni scientifiche e gestiti su piattaforme diverse, spesso non in grado di comunicare tra di

loro. L'accessibilità è poi un fattore essenziale, perché il Web è nato anche come strumento sociale di interazione, e come tale non è pensabile che il suo utilizzo sia precluso a persone che hanno funzionalità ridotte, in misura temporanea o definitiva.

## 2.1 Architettura del Web

L'architettura del Web è basata su tre elementi fondamentali, individuati fin dal primo momento: URI, HTTP e HTML.

Gli **URI** (*Uniform Resource Identifier*) costituiscono l'innovazione fondamentale del Web, e possono identificare qualsiasi cosa (risorse, concetti). Un URI è una sequenza di caratteri utilizzata per identificare il nome di una risorsa, che consente l'interazione con la rappresentazione della risorsa sulla rete (tipicamente il World Wide Web) utilizzando uno specifico protocollo. Un URL (Uniform Resource Locator) è un URI che, oltre a identificare una risorsa, fornisce mezzi per agire su di essa o per ottenere una rappresentazione della risorsa descrivendo il suo meccanismo di accesso primario o la sua "ubicazione" ("location") in una rete.

**HTTP** (*HyperText Transfer Protocol*) è il protocollo per recuperare le risorse sul Web (fetch resources), e ha, tra le sue caratteristiche, la Format Negotiation. L'HTTP funziona su un meccanismo richiesta/risposta (client/server): il client esegue una richiesta e il server restituisce la risposta. Nell'uso comune il client corrisponde al browser ed il server al sito web. HTTP differisce da altri protocolli di livello 7 come FTP, per il fatto che le connessioni vengono generalmente chiuse una volta che una particolare richiesta (o una serie di richieste correlate) è stata soddisfatta.

**HTML** (*HyperText Markup Language*) è il linguaggio utilizzato per la strutturazione dei documenti. Si ricordi che gli aspetti di formattazione dei documenti vengono invece gestiti a livello di CSS (Cascading Style Sheets).

## 2.2 Verso il Semantic Web

Per superare i limiti del Web attuale, dalla fine degli anni 1990 i ricercatori hanno lavorato intensamente per la realizzazione del Semantic Web, che può essere definito come un' infrastruttura basata su metadati per svolgere ragionamenti sul Web. Nel Semantic Web la conoscenza è rappresentata in maniera elaborabile dalla macchina, e può essere utilizzata da componenti automatizzati, denominati agenti software.

I metadati sono informazioni, elaborabili in modo automatico, relative alle "risorse" Web, identificate univocamente da un URI. La tecnologia di riferimento per la codifica, lo scambio e il riutilizzo di metadati strutturati è Resource Description Framework (RDF), basata su un modello molto semplice di "statement", rappresentabili come triple (in termini più formali, una tripla forma un "grafo orientato etichettato").

Per esprimere le restrizioni sulle associazioni, quindi per evitare che possano essere codificati degli "statement" sintatticamente corretti, ma privi di senso, è necessario un meccanismo per rappresentare "classi di oggetti". Da questa esigenza nasce "RDF Vocabulary Description Language", che mantiene anche, per ragioni storiche, il nome di "RDF Schema" (RDFS).

Per poter effettuare dei ragionamenti, per definire le classi, e per varie altre esigenze, però, RDFS non è sufficiente, e occorre un modo per rappresentare la conoscenza e le regole che permettono di dedurre ulteriore conoscenza: l' ontologia.

Ma il Web è intrinsecamente distribuito, e quindi occorre un linguaggio che non solo permetta di esprimere dati e regole sui dati, ma che consenta anche di esportare questa conoscenza (ontologia) per renderla disponibile a qualunque applicazione. Il W3C ha definito, per questa esigenza, il Web Ontology Language (OWL).

### 2.2.1 Resource Description Framework (RDF)

RDF è lo strumento base per la codifica, lo scambio e il riutilizzo di metadati strutturati, e consente l'interoperabilità tra applicazioni che si scambiano sul Web informazioni machine-understandable. RDF fornisce un modello per descrivere le risorse, che hanno delle proprietà (o anche attributi o caratteristiche). RDF definisce una risorsa come un qualsiasi oggetto che sia identificabile univocamente mediante un URI. Il data model RDF, che consente di rappresentare statement RDF in modo sintatticamente neutro, è molto semplice ed è basato su tre tipi di oggetti:

**Resources** Qualunque cosa descritta da una espressione RDF viene detta risorsa (*resource*). Una risorsa può essere anche un oggetto non direttamente accessibile via Web (per es. un libro, un dipinto, etc.). Qualunque cosa può essere identificata da un URI.

**Properties** Una *property* (proprietà) è un aspetto specifico, una caratteristica, un attributo, o una relazione utilizzata per descrivere una risorsa. Ogni proprietà ha un significato specifico, definisce i valori ammissibili, i tipi di risorse che può descrivere, e le sue relazioni con altre proprietà. Le proprietà associate alle risorse sono identificate da un *nome*, e assumono dei *valori*.

**Statements** Una risorsa, con una proprietà distinta da un nome, e un valore della proprietà per la specifica risorsa, costituisce un RDF *statement* (asserzione). Uno statement è quindi una *tripla* composta da un *soggetto* (risorsa), un *predicato* (proprietà) e un *oggetto* (valore). L'oggetto di uno statement (cioè il *property value*) può essere un'espressione oppure un'altra risorsa.

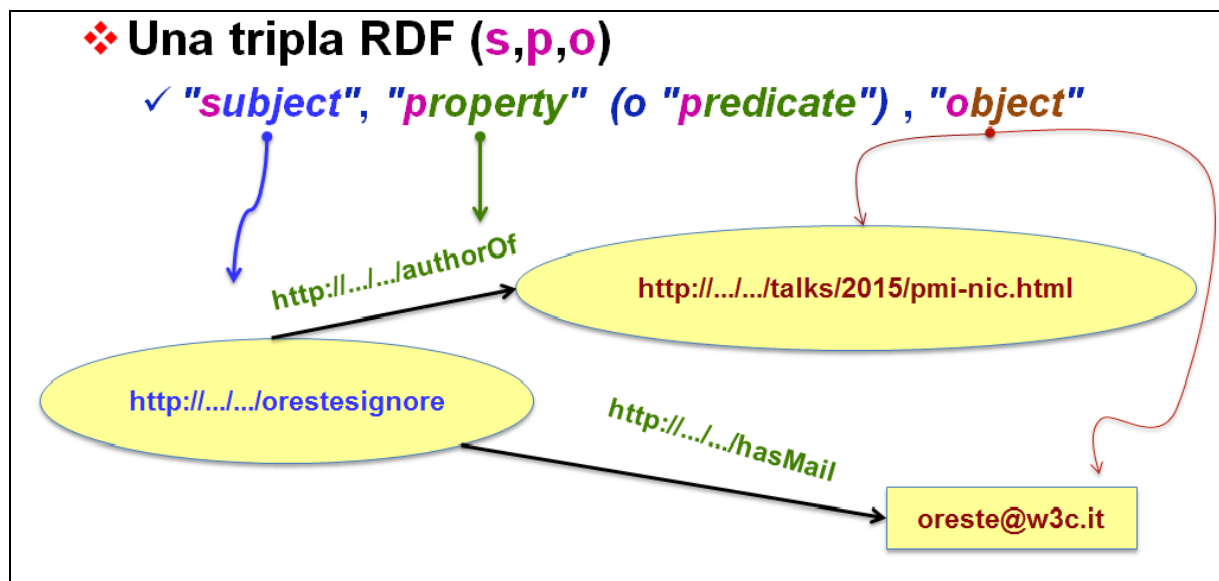


Figura 1. La rappresentazione grafica di una tripla RDF

Graficamente, le relazioni tra Resource, Property e Value vengono rappresentate mediante grafi orientati etichettati, in cui le risorse vengono identificate come nodi (graficamente delle ellissi), le proprietà come archi orientati etichettati, e i valori corrispondenti a sequenze di caratteri come rettangoli (vedi Figura 1). Lo statement può poi essere codificato in vari modi, a seconda della sintassi adottata.

Si noti che Subject, Property e Object sono individuati univocamente da URI (a parte il caso in cui un Object è costituito da una sequenza di caratteri), per cui un grafo RDF può *includere qualunque risorsa identificabile sul Web*.

RDF è per il Semantic Web ciò che HTML è stato per il Web.

### 2.3 La Open Web Platform

La Open Web Platform è un insieme di tecnologie aperte (quindi che non prevedono il pagamento di diritti) che rendono operativo il Web. Tutti hanno diritto di implementare componenti software del Web, senza dover chiedere autorizzazioni né pagare alcun diritto. Il W3C e i suoi partner hanno ideato e continuano a sviluppare la Open Web Platform, che ha già attirato l'interesse di molte imprese, che hanno compreso come la sua ricchezza, universalità e interoperabilità su diversi dispositivi offra delle enormi potenzialità di business. Una delle tecnologie centrali della Open Web Platform è HTML5, che, secondo molti analisti, promette di offrire un enorme numero di posti di lavoro nel prossimo futuro.

### 2.4 Linked Open Data

In termini molto semplici, il paradigma dei *Linked Open Data* (LOD) è il modo per evolvere dal "*Web of Documents*" al "*Web of Data*". Il *Web of Documents* è il Web al quale siamo abituati, quindi un Web che può essere considerato un *filesystem globale*, in cui gli oggetti principali sono i documenti (in genere poco o per nulla strutturati), connessi tra di loro da link non tipati (quindi senza una particolare semantica), e la semantica dei link e dei documenti è implicita. Evidentemente il Web of Documents, che è intrinsecamente semplice, è stato progettato essenzialmente per essere utilizzato da esseri umani: i dati non sono connessi tra di loro, e solo la mente umana, con un processo di interpretazione e ragionamento, può estrarre nuova conoscenza dai documenti.

Il *Web of Data*, invece, può essere visto come un *database globale*, in cui gli oggetti primari sono *Things* (cioè "Cose" o una loro descrizione). Tra queste "Cose" (compresi i documenti) vi sono dei *link tipati*. Grazie all'elevato livello di strutturazione delle "Cose" (o delle loro descrizioni) e alla presenza di link tipati, la semantica dei link e dei contenuti è esplicita, per cui per le macchine possono effettuare dei ragionamenti e derivare nuova conoscenza. Ne deriva che il Web of Data è progettato essenzialmente per le macchine, e in seconda istanza per gli esseri umani. Le regole dei LOD sono abbastanza semplici e sono in sostanza:

1. Usare URI per dare un nome alle "Cose".
2. Usare HTTP URI, in modo che le persone possano cercare e trovare quei nomi.
3. Quando qualcuno esamina un URI, fornire informazioni utili, usando gli standard (tecnologie RDF, SPARQL).
4. Includere link ad altri URI, in modo che sia possibile scoprire ulteriori "Cose".

Un approccio LOD realmente operativo ed efficace richiede il supporto di *ontologie*. In maniera molto semplice si può definire un'ontologia come: "*insieme di termini di*

*conoscenza, che includono il vocabolario, le interconnessioni semantiche e alcune semplici regole di inferenza e di logica relativamente a qualche specifico dominio di conoscenza*". Esistono molte altre definizioni di ontologia, senz'altro più sofisticate, ma va sottolineato come, nonostante le tante definizioni, ci sia un accordo generale sul significato del termine e soprattutto sul fatto che un'ontologia contiene una *conoscenza condivisa*, che costituisce il collante che consente alle macchine di comprendere i dati, rendendoli effettivamente interoperabili.

### 3 Dov'è il Semantic Web?

Le tecnologie del Semantic Web sono presenti, spesso in modo non evidente, in moltissime applicazioni moderne, o possono offrire una via per risolvere problemi implementativi migliorando la qualità delle applicazioni e contribuendo a ridurre i costi di sviluppo e manutenzione. Ci sono alcuni settori che stanno attirando l'interesse del mondo produttivo e offrono scenari interessantissimi di sviluppo, con prospettive rilevanti di crescita e sviluppo. Tra i vari settori, basti qui ricordare Open Data, Smart City, Industria 4.0, ambient intelligence, e automobili.

Lo scenario di base è costituito dall' *Internet of Things* o Internet delle cose. Nel seguito si darà una breve descrizione di cosa si intende per Internet of Things e di come iniziative portate avanti dal World Wide Web Consortium mirino a superare alcune difficoltà intrinseche proponendo il *Web of Things* nel quale le tecnologie semantiche giocano un ruolo assai importante.

#### 3.1 Internet of Things (IoT)

Per *Internet of Things* si intende una rete in continua espansione di oggetti fisici, che dispongono di un indirizzo IP per la connettività Internet, e la comunicazione che avviene tra questi oggetti e altri dispositivi Internet-enabled e sistemi. L'Internet delle cose estende la connettività internet oltre ai dispositivi tradizionali come computer desktop e portatili, smartphone e tablet, anche a una vasta gamma di dispositivi e oggetti di uso quotidiano che utilizzano la tecnologia "*embedded*" (incorporata) per comunicare e interagire con l'ambiente esterno. Il tutto via Internet.

Moltissimi sono gli oggetti che ricadono nel campo dell'IoT. Possiamo pensare per esempio a sistemi di sicurezza, termostati, automobili e loro componenti, apparecchiature elettriche ed elettroniche, luci in ambienti casalinghi o commerciali, sveglie, sistemi di altoparlanti, distributori automatici, etc. Le aziende possono sfruttare le applicazioni dell'IoT per automatizzare le attività di sicurezza (ad esempio, informare le autorità quando un estintore o un ascensore in un palazzo sono bloccati) per eseguire l'analisi del mondo reale con telecamere di rete e sensori per rilevare come i clienti interagiscono con i prodotti. Gli analisti prevedono che nei prossimi anni saranno distribuiti cento miliardi di dispositivi IoT, e che quindi il potenziale di crescita sarà enorme.

Tuttavia, non si possono sottovalutare alcuni problemi, che bloccano i vantaggi offerti dalla rete. Come è facile immaginare, tali problemi derivano essenzialmente dalla scarsa interoperabilità delle soluzioni che si affacciano sul mercato, per cui si hanno catene di prodotti che non sono in grado di interoperare e una grande varietà di approcci e piattaforme incompatibili. Questi problemi determinano delle difficoltà per gli sviluppatori, che sono costretti ad apprendere piattaforme diverse e incontrano ostacoli nel porting da una piattaforma all'altra, con conseguenti e ovvi

maggiori costi. In un ambiente così eterogeneo, senza standard di riferimento, diventa una vera sfida creare servizi che operino su piattaforme e domini diversi. È quindi interesse degli sviluppatori sbloccare il potenziale commerciale dell'IoT, riducendo i costi di sviluppo per applicazioni e servizi IoT. In tal modo sarebbe possibile soddisfare le richieste dei clienti per servizi che richiedono l'integrazione con altre piattaforme e aumentare la dimensione del mercato.

### 3.2 Web of Things (WoT)

Per sbloccare il potenziale commerciale vi è la necessità di *ecosistemi aperti* basati su *standard aperti*. Questo include standard per l'identificazione, la scoperta e l'interoperabilità dei servizi attraverso piattaforme di diversi fornitori, e richiederà descrizioni ricche e modelli di dati condivisi, così come particolare attenzione alla sicurezza, privacy, scalabilità e accessibilità. Gli ecosistemi aperti stimoleranno la crescita attraverso la creazione di mercati più ampi per gli sviluppatori, liberandoli dal peso di dover adattare i prodotti a piattaforme specifiche del singolo fornitore.

L'identità è importante per dispositivi, utenti, applicazioni e servizi, ad esempio, come parte di end to end per la gestione della sicurezza e della fiducia (*trust*). A differenza delle normali applicazioni web, non si può supporre che l'utente sia presente e in grado di autenticarsi. La gestione della fiducia comporterà mezzi per verificare i metadati, ad esempio la provenienza dei dati, la posizione di un dato sensore, e così via.

Applicazioni e servizi spesso necessitano di dati ad un livello superiore a quello dei dati grezzi forniti dai sensori. Inoltre, i dati devono essere interpretati nel contesto di altre fonti di informazione. Lo stesso vale per sistemi di controllo le cui azioni devono essere tradotte in un contesto in azioni su entità di livello inferiore. Il WoT deve essere in grado di modellare il mondo reale a diversi livelli di astrazione, e consentire l'apertura dei mercati in libera concorrenza per i servizi attraverso questi livelli. Le cose nel WoT possono essere considerate come rappresentazioni virtuali di oggetti.

Una conseguenza di questo è che le "cose" nel WoT non si limitano ai dispositivi collegati, ma possono anche includere cose che non sono e non possono essere collegate, come persone e luoghi, e idee astratte, come eventi (ad esempio, un concerto), organizzazioni, periodi di tempo (ad esempio, anni '60). Ogni cosa può avere una o più rappresentazioni virtuali (avatar). Le cose possono anche avere una storia, per esempio, per un'automobile, registrare la sequenza dei precedenti proprietari. Gli avatar hanno identità, descrizioni ricche, servizi, controllo degli accessi e politiche per la gestione dei dati. Gli avatar hanno URI e sono accessibili tramite tecnologie web. Gli avatar rendono più facile la costruzione di applicazioni e servizi che combinano le informazioni provenienti da fonti diverse e diversi livelli di astrazione.

Internet fornisce una base per connettere i sistemi, ma, esattamente come il sistema telefonico, non è utile se le persone non parlano nella stessa lingua. Il World Wide Web Consortium (W3C) sta cercando di cambiare la situazione attraverso il lavoro su standard globali che consentano l'utilizzo di tecnologie web per interconnettere piattaforme IoT attraverso il Web, basandosi su una nuova classe di server Web, implementati a diversi livelli. Il W3C propone, come base per l'interoperabilità, un quadro concettuale in cui siano condivisi semantica e formati di dati.

Nel framework proposto per il WoT si espongono piattaforme e dispositivi IoT tramite il World Wide Web, grazie ad uno strato di astrazione dei dispositivi. Le “Cose” vengono modellate in termini di eventi, proprietà e azioni (vedi una esemplificazione nella tabella seguente).

Evento	Proprietà	Azioni
<ul style="list-style-type: none"> <li>• Il campanello suona</li> <li>• Chiave nella toppa</li> </ul>	La porta può essere: <ul style="list-style-type: none"> <li>• aperta</li> <li>• chiusa</li> </ul>	Apri la porta

Si noti infine il parallelismo architetturale tra il Web of Documents e il Web of Things:

- Il *Web delle pagine* è basato su IRI per l'indirizzamento, HTTP come protocollo di accesso, HTML per struttura e descrizione delle pagine;
- Il *Web of Things* è basato su IRI per l'indirizzamento, HTTP e altri protocolli accesso, TDL (Thing Description Language) per descrivere la semantica e il formato dei dati (interoperabilità) e le relazioni con altre “Cose” (per consentire la scoperta di altre “Cose”).

#### 4 Conclusioni

Il Web è nato per *condividere conoscenza* e consentire interazioni sociali, superando le difficoltà derivanti dall'esistenza di molteplici protocolli proprietari. L'*interoperabilità* è quindi uno dei principi cardine del Web. L'architettura del Web è molto semplice, e il Web si è evoluto restando fedele ai suoi principi e alla sua architettura.

Un settore in cui sono state investite molte energie da parte dei ricercatori di tutto il mondo è il *Semantic Web*, che costituisce un'evoluzione del Web attuale. Le tecnologie del Semantic Web costituiscono la base di molte iniziative interessanti e di svariate applicazioni.

L'interoperabilità e gli standard aperti sono le chiavi di lettura per comprendere il significato della *Open Web Platform* e dei *Linked Open Data*.

Nella stessa linea si pone l'attività volta a definire il framework per il *Web of Things*, che si presenta come il modo per superare alcune difficoltà intrinseche dell'Internet of Things, il cui pieno sviluppo è frenato dalle incompatibilità tra i vari protocolli proprietari e l'assenza di informazioni semantiche che consentano il pieno sviluppo delle potenzialità di un settore che si annuncia vitale e di grande impatto sulla nostra esistenza quotidiana, in quanto fattore abilitante di progetti straordinari quali Smart City, Ambient Intelligence, Industria 4.0, automobili intelligenti e tanti altri.

Una volta di più, le idee di base che hanno portato alla nascita del Web, e il modo in cui il W3C, grazie all'apporto dei suoi membri, ne guida lo sviluppo, dimostrano come sarebbe essenziale, per la competitività delle imprese, *partecipare in prima persona al W3C*. Investire sulle tecnologie W3C costituisce un'opportunità di crescita professionale per i singoli, e un'opportunità di sviluppo competitivo per le imprese.

## Bibliografia/Sitografia

URI: <http://www.w3.org/Addressing/>

URI clarification: <http://www.w3.org/TR/uri-clarification/>

URI spec: <http://www.w3.org/Addressing/URL/uri-spec.ps>

URI wikipedia: [https://en.wikipedia.org/wiki/Uniform\\_Resource\\_Identifier](https://en.wikipedia.org/wiki/Uniform_Resource_Identifier)

URI/IRI IETF: <https://tools.ietf.org/html/rfc7595>

Hansen, T.; Hardie, T. (June 2015). Thaler, D., ed. "[\*Guidelines and Registration Procedures for URI Schemes\*](#)". *Internet Engineering Task Force*. ISSN 2070-1721

*Hypertext Markup Language (HTML) - A Representation of Textual Information and MetaInformation for Retrieval and Interchange*. Tim Berners-Lee e Daniel Connolly. 1993. <http://www.w3.org/MarkUp/draft-ietf-iiir-html-01.txt>

History da HTML5 - A vocabulary and associated APIs for HTML and XHTML - W3C Recommendation, §1.4 . World Wide Web Consortium. 28 ottobre 2014. <http://www.w3.org/TR/html/introduction.html#history-0>

Tim Berners-Lee: <http://www.w3.org/DesignIssues/LinkedData.html> (2007)

Tom Heath, Christian Bizer, *Linked Data: Evolving the Web into a Global Data Space* (1st edition). *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1:1, 1-136. Morgan & Claypool. (2011) <http://linkeddatabook.com/book>

Rudi Studer, V. Richard Benjamins, Dieter Fensel, *Knowledge Engineering: Principles and Methods*, *Data, Knowl. Eng.* 25(1-2): 161-197 (1998)

Asunción Gómez-Pérez, Mariano Fernández-López, Oscar Corcho, *Ontological Engineering*, Springer-Verlag (2004), ISBN 1-85233-551-3

<http://www.w3.org/WoT/>

W3C Workshop on the Web of Things - Enablers and services for an open Web of Devices (25–26 June 2014, Berlin, Germany) - <http://www.w3.org/2014/02/wot/>

Oreste Signore: *RDF per la rappresentazione della conoscenza* (<http://www.w3c.it/papers/rdf.pdf>)