

# Il web semantico: il contesto di lavoro odierno

di Mauro Guerrini

Il *Manifesto sul web semantico* dell'AIB emanato il 12 ottobre 2020 inizia con questa frase: «Il Web è diventato il più grande insieme di significato condiviso nella storia umana, eppure questo significato risulta frammentato a causa di differenze linguistiche, concettuali e normative»<sup>1</sup>. C'è tutto: il contesto digitale e il paradigma del lavoro intellettuale, il *digital divide* tra chi può e chi non può approfittare della grande opportunità della ricerca e del lavoro in rete, la responsabilità e la condivisione.

Subito dopo il *Manifesto* parla del passaggio dal web dei documenti al web dei dati, di dati auto-esplicativi: è un passaggio sostanziale che introduce la tecnologia dei linked data, il cui uso, come recita il punto finale del primo paragrafo, non deve ingannare: «Dietro alla semplicità dei Linked Data è presente una descrizione complessa dei dati con i quali si formano le *triple*, e tale complessità aumenta se l'obiettivo è quello di collegare i dati fra domini e set di dati differenti»<sup>2</sup>. Prosegue il documento: «La maggior parte delle iniziative di gestione dell'informazione stanno adottando strategie aperte, decentralizzate e globali, incentrate sui dati»<sup>3</sup>, ovvero il *Manifesto* introduce la dimensione dell'interoperabilità e dell'accessibilità e apertura dei dati, ovvero dell'open access, filosofia che si inserisce nella strategia più ampia dell'open science. Si tratta di concetti qualificanti e decisivi.

MAURO GUERRINI, Università degli studi di Firenze, Dipartimento di storia, archeologia, geografia, arte e spettacolo, Firenze, e-mail mauro.guerrini@unifi.it.

Ultima consultazione dei siti web: 26 gennaio 2022.

Viene qui pubblicata la trascrizione dell'intervento tenuto dall'autore nell'ambito del seminario "Luci e ombre del semantic web" (21 gennaio 2022), durante il quale è stato presentato il *Manifesto sul web semantico* a cura del Gruppo di studio Catalogazione, indicizzazione, linked open data e web semantico dell'AIB. I dettagli e la videoregistrazione dell'evento sono disponibili alla pagina <<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-studio-catalogazione-ed-indicizzazione/2022/96428-luci-ombre-del-semantic-web/>>.

1 Associazione italiana biblioteche. Gruppo di studio catalogazione ed indicizzazione, *Manifesto per il web semantico: la comunità dei dati*. 12 ottobre 2020, <<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-studio-catalogazione-ed-indicizzazione/2020/86520-manifesto-per-il-web-semantico-la-comunita-dei-dati/>>.

2 *Ibidem*.

3 *Ibidem*.



Prosegue parlando di ontologie, che costituiscono il cuore del web semantico; senza questi strumenti esso non esisterebbe; le ontologie sono il risultato di studi e di schematizzazioni elaborate da esperti di dominio; ciò per ricordare che l'intervento umano è ancora fondamentale in un complesso tecnologico concepito per le macchine.

Nel mondo delle biblioteche BIBFRAME è un'ontologia che sta affermandosi come un buon candidato (anche se non l'unico) per l'implementazione del web semantico e dei linked data. Tuttavia – come ci ha ricordato Richard Wallis all'“International Conference on Bibliographic Control in the Digital Ecosystem (BC2021)”:  
«We mustn't expect the rest of the world to use our vocabularies. Having a fully Linked Data catalogue is not going to do a lot, for people finding your resources across the web»<sup>4</sup>. Dobbiamo sviluppare bene gli standard del nostro specifico dominio senza rinunciare agli strumenti che garantiscono la nostra precisione e la nostra autorevolezza, ma nello stesso tempo dobbiamo facilitare l'accesso a chi proviene da mondi generalisti. Wallis ricorda che è strategico per le biblioteche prendere in considerazione 'anche' l'ontologia usata dai motori di ricerca – Schema.org. Come ci ricorda la Tesi 3 del *Nuovo Manifesto per le biblioteche digitali* dell'AIB:

La visione del Web semantico si sta sempre più affermando: le biblioteche digitali hanno tutti gli strumenti per giocare un ruolo fondamentale nel web dei dati: grazie alla intermediazione di applicazioni (o agenti software) molte altre conversazioni in differenti contesti possono essere attivate o influenzate<sup>5</sup>.

In un web aperto, pertanto, chiunque può contribuire a migliorare la qualità dei dati bibliografici (per esempio, Wikidata). L'esclusività delle agenzie bibliografiche nazionali viene meno; sono, infatti, numerosi gli attori che possono partecipare alla costruzione dinamica e condivisa di nuova informazione bibliografica che prendono parte a un processo di riutilizzo e arricchimento dei dati, collegando singoli *dataset*.

L'obiettivo è rendere i dati strutturati interconnessi e riusabili tramite l'apparente semplicità di una tripla RDF (*resource description framework*). La tecnologia dei linked data, infatti, garantisce l'interoperabilità dei dati in contesti diversi rispetto a quello originario in cui sono stati prodotti, mantenendo la loro valenza semantica. È come un gioco di entità atomiche che, unendosi, formano le molecole, che a loro volta strutturano un dominio specifico: le entità atomiche relazionate. L'attenzione è riposta sulle varie entità, che sono in relazione tra loro, anziché sulle relazioni in sé. RDF va inteso nel senso di un'infrastruttura, di un'architettura di sostegno e di organizzazione della conoscenza disponibile in dati singoli, granulari, che esplicita i valori semantici delle relazioni tra i dati che concernono una risorsa. RDF, in sostanza, è la grammatica e la sintassi del web semantico; la tripla costituisce la struttura normativa del linguaggio del web semantico. RDF offre strumenti che facilitano la collaborazione, ma la collaborazione o il superamento dei confini dipendono da scelte di realizzazione e non sono il risultato automatico dell'adozione di una determinata tecnologia. Il rischio di una mera adesione tecnologica – senza alcuna scelta politica

<sup>4</sup> Wallis Richard, *Follow me to the library! Bibliographic data in a discovery driven world*, «JLIS.it», 13 (2022), n. 1, p. 37-44: p. 44, DOI: 10.4403/jlis.it-12762.

<sup>5</sup> Associazione italiana biblioteche. Gruppo di lavoro sulle biblioteche digitali, *Nuovo manifesto per le biblioteche digitali*. 5 maggio 2020, <<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-lavoro-biblioteche-digitali/2020/82764-nuovo-manifesto-per-le-biblioteche-digitali/>>.

e senza abbracciare e condividere lo spirito di apertura e condivisione proposto dai linked data – è di creare nuovi silos informativi chiusi per definizione.

Nel web semantico i collegamenti non avvengono più tra documenti ma tra dati. Il record tradizionale monolitico e rigido in ogni contesto si trasforma in un insieme di elementi. In altre parole, i dati, prodotti dallo spezzettamento dell'originario record, diventano usabili e riusabili in contesti diversi, come i mattoncini del Lego.

Il *Manifesto* ricorda anche i rischi quando afferma che attuare il web semantico non «significa sostituire i documenti primari con dati granulari»<sup>6</sup>. Il web semantico facilita l'accesso, non sostituisce le risorse documentarie, librerie, digitali o tradizionali. Per usare le parole di Tim Berners-Lee: «Affinché il web semantico possa funzionare, i computer devono avere accesso a raccolte strutturate di informazioni e a una serie di regole di inferenza utilizzabili per condurre ragionamenti automatici»<sup>7</sup>.

Si pone il problema della qualità dei dati; chi e come li controlla? Le macchine sono in grado di farlo? Viene ancora naturale il riferimento all'authority control e al ruolo, che può essere nuovamente centrale, delle agenzie bibliografiche nazionali. Il riferimento è al principio AAA, *anyone can say anything about anything*, interpretato dal W3C come: *anyone can make statements about any resource*<sup>8</sup>, ben consapevoli del rischio che questo principio porta con sé e a come, in ottemperanza al principio di *proof and trust*, le agenzie bibliografiche nazionali possano riacquisire un ruolo determinante nel garantire autorevolezza nel controllo bibliografico. Qui il modello può essere Wikidata, che prevede che ogni tripla possa essere, idealmente, collegata ai dati che contestualizzano l'informazione veicolata dalla tripla (*qualifier*, qualificatori) e alla fonte dell'affermazione esposta nella tripla. 'Chi ha detto cosa?' e 'in che contesto?' è importante anche nel web semantico.

Elaine Svenonius, parlando di risorse digitali, sostiene che «ciò che è difficile da identificare è difficile da descrivere e quindi difficile da organizzare»<sup>9</sup>.

RDA è componente fondamentale e presupposto per questo nuovo contesto del web semantico. Dato che 'identificare' e 'collegare' le entità sono i due obiettivi fondamentali dello standard, il catalogo continuerà la sua trasformazione in un portale per la navigazione, riprendendo il concetto di 'navigare' legato al nome di Svenonius e ripreso da ICP, *International Cataloguing Principles*. RDA punta all'identificazione delle risorse per poi consentirne l'accesso.

Il nuovo concetto di authority control sta sempre più evolvendo verso una forma di attività universale e collaborativa d'identificazione: l'identificazione di entità, intese come agenti, concetti, risorse o meglio, come *real world object*, è l'attività sulla quale convergono grandi progetti locali e internazionali, di agenzie autorevoli, come ISNI e VIAF, e collaborative, come Wikidata. L'identificazione ha assunto un ruolo centrale nel concetto di authority control e l'assegnazione di identificatori diventa preponderante, o forse più funzionale allo scopo, anche rispetto all'assegnazione di attributi più tradizionalmente 'qualificanti' l'entità.

<sup>6</sup> AIB. CILW, *Manifesto per il web semantico* cit.

<sup>7</sup> Dunsire Gordon, *RDA and the semantic web: lectio magistralis in library science: Firenze, Università degli studi di Firenze, 4 marzo 2014*. Fiesole (FI): Casalini libri, 2014, p. 3.

<sup>8</sup> Cfr. <<https://www.w3.org/TR/rdf-concepts/#section-anyone>>.

<sup>9</sup> Elaine Svenonius, *Il fondamento intellettuale dell'organizzazione dell'informazione*, traduzione di Maria Letizia Fabbri, introduzione di Mauro Guerrini. Firenze: Le lettere, 2008, p. 13.

Secondo IFLA LRM, l'identificazione di una risorsa può avvenire tramite i suoi attributi e tramite il *nomen* associato a essa. Il *nomen* rappresenta la denominazione di un'entità e può essere un nome, un titolo, un identificatore o un punto d'accesso. Gli identificatori acquisiscono particolare rilevanza nel web semantico. Essi costituiscono delle stringhe numeriche o alfanumeriche di caratteri, che, secondo protocolli convenzionali e specifiche modalità di applicazione vengono associate in modo permanente a una risorsa per garantire un accesso stabile ai suoi metadati assicurando l'unicità. Le biblioteche, insieme ad altre istituzioni della memoria registrata,

hanno sempre mostrato grande attenzione verso la pratica dell'identificazione degli oggetti, creando meccanismi e strumenti d'indicizzazione e controllo, per evitare e risolvere, per esempio, omonimie e sinonimie: la creazione di accessi univoci alle informazioni presenti nel catalogo, la formulazione di voci d'autorità per entità quali nomi, titoli, voci di soggetto, la registrazione di identificatori univoci quali ISBN/ISSN. È questa sensibilità e consuetudine all'identificazione delle risorse che indirizza il web semantico al mondo delle biblioteche, degli archivi e dei musei<sup>10</sup>.

Tra le diverse tipologie di identificatori, rivestono un'importanza particolare gli URI (*uniform resource identifier*). Essi costituiscono le fondamenta del web e rappresentano qualsiasi entità che ne faccia parte. Un URI è costituito da una stringa di caratteri interpretabili da una macchina ed è costruito secondo criteri di semplicità, stabilità, manutenibilità e persistenza a lungo termine; inoltre, esso dovrebbe essere dereferenzabile, ossia capace di garantire il reperimento, da parte di una macchina, degli attributi utili e necessari all'identificazione di una risorsa. La stabilità e la persistenza, in particolare, costituiscono delle proprietà fondamentali per gli identificatori all'interno del web: l'assenza di queste caratteristiche impoverisce il sistema d'identificazione e fruizione delle risorse, rendendolo inadatto a usi da parte di altre comunità.

Il *Manifesto* parla, quindi, di accordo intellettuale, responsabilità umana, credibilità e autorevolezza e chiama in causa la responsabilità che in questo processo di completa revisione e rifondazione di certi riferimenti che credevamo stabili e consolidati rivestono le istituzioni e i loro rappresentanti: in una rivoluzione globale dell'informazione, provare a rimanerne fuori, illesi, espone solo al rischio dell'esclusione. Rimanere su posizioni tradizionali, seppur confortevoli per aver garantito per decenni un sapere consolidato e sicuro, è una posizione poco costruttiva, perché significa ignorare la realtà e cedere il passo a istituzioni meno autorevoli ma capaci di stare 'sul pezzo' e seguire il movimento. Certe posizioni conservatrici che proteggono dalla fatica dell'aggiornamento (che richiede un enorme investimento di energie intellettuali e di risorse economiche) espongono, in realtà, un'intera comunità al rischio di perdere autorevolezza e non avere più alcuna voce in capitolo nelle importanti discussioni in corso nella comunità GLAM internazionale. Questo rischio non possiamo permettercelo: essere seduti a certi tavoli internazionali, anche a costo di rimanere inizialmente silenziosi, consente di poter, al momento opportuno, esporre la propria idea e ribadire le proprie posizioni. Essere del tutto assenti da certi ambiti, conduce alla marginalizzazione.

<sup>10</sup> Mauro Guerrini; Tiziana Possemato, *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica, 2015, p. 35.

Le parole di Tim Berners-Lee possono chiudere questa breve riflessione:

Le attività dell'uomo sono caratterizzate da un'eterna tensione tra gruppi che agiscono indipendentemente e la necessità di inserirsi in una comunità più vasta. Un piccolo gruppo può sviluppare innovazioni velocemente e con efficienza, ma così facendo genera una sottocultura i cui concetti non sono capiti dagli altri. Coordinare le azioni di un vasto gruppo, tuttavia, è un processo lento, e richiede un'enorme mole di comunicazione. Il mondo funziona a tutti i livelli tra questi estremi, con una tendenza a iniziare dal piccolo – dalle idee personali – per muoversi col tempo verso una comprensione più generale. [...] Il Web semantico, definendo ogni concetto semplicemente con un URI, permette a chiunque di inventarne di nuovi, a suo piacimento. Il suo linguaggio logico di unificazione permetterà a questi concetti di essere progressivamente collegati in un Web universale. Ciò renderà disponibili conoscenza e lavoro umano all'analisi da parte degli agenti software, offrendoci una nuova classe di strumenti con i quali potremo vivere, lavorare e imparare insieme<sup>11</sup>.

Articolo proposto il 28 gennaio 2022 e accettato il 7 febbraio 2022.

---

**ABSTRACT**

AIB studi, 62 n. 1 (gennaio/aprile 2022), p. 113-117. DOI 10.2426/aibstudi-13395  
ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2022 Mauro Guerrini

---

MAURO GUERRINI, Università degli studi di Firenze, Dipartimento di storia, archeologia, geografia, arte e spettacolo, Firenze, e-mail mauro.guerrini@unifi.it.

**Il web semantico: il contesto di lavoro odierno**

L'articolo riporta l'intervento tenuto dall'autore nell'ambito del seminario "Luci e ombre del semantic web" (21 gennaio 2022), durante il quale è stato presentato il *Manifesto sul web semantico* a cura del Gruppo di studio Catalogazione, indicizzazione, linked open data e web semantico dell'AIB.

Interconnessione, interoperabilità, accesso, qualità del dato sono i pilastri su cui deve poggiare il web semantico, senza temere di mettere in discussione principi e riferimenti ritenuti oramai stabili e consolidati.

**The semantic web: the context of today's work**

The article reports the speech given by the author as part of the seminar "Lights and shadows of the semantic web" (January 21, 2022), during which the *Manifesto sul web semantico* (Semantic Web Manifesto) was presented by the Cataloguing, Indexing, Linked Open Data and Semantic Web Study Group of the Italian Libraries Association (AIB).

Interconnection, interoperability, access and data quality are the pillars on which the semantic web must rest, without fearing to question principles and references that, at this point, are considered stable and consolidated.

<sup>11</sup> Tim Berners-Lee; James Hendler; Ora Lassila, *Il web semantico*, «Le Scienze», (2001), n. 393, p. 77-84: p. 84, <[http://download.kataweb.it/mediaweb/pdf/espresso/scienze/2001\\_393\\_6.pdf](http://download.kataweb.it/mediaweb/pdf/espresso/scienze/2001_393_6.pdf)> (ed. or. Tim Berners-Lee; James Hendler; Ora Lassila, *The semantic web*, «Scientific American», 284 (2001), n. 5, p. 34-43).