

# Un nuovo progetto di biblioteca digitale con mappatura semantica dei commenti alla «Commedia»: L'«Hypermedia Dante Network»

Carlo Meghini, Michelangelo Zaccarello

Pubblicato: 15 dicembre 2021

## *Abstract*

Dante is the most studied Italian author in the world, and it is not really a coincidence that some of the most pioneering initiatives in Digital Humanities are due to Dante's philology. The *Hypermedia Dante Network* project aims to build a digital library that contains comments to the *Comedy* formally represented according to the standards of the Semantic Web, with two main purposes: to put at the service of humanists a body of knowledge to be explored also with the help of agents digital to formulate and verify scientific hypotheses; and to build new applications for the visualization of this knowledge for non-experts. The digital library will account for the phenomena of intertextuality and interdiscursiveness in the *Comedy*, continuing the work undertaken by a similar previous project focused on Dante's minor works. To articulate knowledge in a logically transparent format, the standards that have been chosen are the Resource Description Framework (RDF) for the encoding of factual knowledge and OWL (Web Ontology Language) 2 for the encoding of the ontology. In addition to developing the ontology, the project will provide humanists with a tool for the collection of knowledge and a portal to provide access to the knowledge collected.

Dante è l'autore italiano più studiato al mondo, e non è davvero un caso che proprio alla filologia dantesca si debbano alcune fra le iniziative più pionieristiche nelle Digital Humanities. Il progetto *Hypermedia Dante Network* si propone di costruire una biblioteca digitale che contenga i commenti alla *Commedia* formalmente rappresentati secondo gli standard del web semantico, con due scopi principali: mettere al servizio degli umanisti un corpus di conoscenze da esplorare anche con l'ausilio di agenti digitali per formulare e verificare ipotesi scientifiche; e costruire nuove applicazioni di fruizione di tale conoscenza anche per i non esperti. La biblioteca digitale darà conto dei fenomeni dell'intertestualità e interdiscorsività nella *Commedia*, proseguendo il lavoro intrapreso da un analogo progetto precedente focalizzato sulle opere minori dantesche. Per articolare la conoscenza in un formato logicamente trasparente, gli standard che si sono scelti sono il Resource Description Framework (RDF) per la codifica della conoscenza fattuale e OWL (Web Ontology Language) 2 per la codifica dell'ontologia impiegata. Oltre a sviluppare l'ontologia, il progetto metterà a disposizione degli umanisti un *tool* per la raccolta della conoscenza e un portale per dare la conoscenza raccolta in accesso.

**Parole chiave:** «Commedia»; Dante Alighieri; intertestualità; ontologie; umanistica digitale.

**Carlo Meghini:** ISTI-CNR

✉ [carlo.meghini@isti.cnr.it](mailto:carlo.meghini@isti.cnr.it)

**Michelangelo Zaccarello:** Università degli Studi di Pisa

✉ [michelangelo.zaccarello@unipi.it](mailto:michelangelo.zaccarello@unipi.it)

Copyright © 2021 Carlo Meghini, Michelangelo Zaccarello

The text in this work is licensed under Creative Commons BY-SA License.

<https://creativecommons.org/licenses/by-sa/4.0/>

### 1. *Dante e il web semantico: nuove prospettive nell'organizzazione delle conoscenze sui commenti alla Commedia*

Dante è l'autore italiano più studiato al mondo, e la sua bibliografia è davvero sterminata, con commentatori ed esegeti antichi e moderni che divengono a loro volta oggetto di ricerche ed edizioni. Nel contesto di una tale dispersione, il ricorso al mezzo digitale per indicizzare e valorizzare le risorse è obbligatorio, e non è davvero un caso che proprio alla filologia dantesca si debbano alcune fra le iniziative più pionieristiche nelle Digital Humanities. Strumento ormai indispensabile per analizzare e confrontare i commenti al poema dantesco, il [Dartmouth Dante Project](#), ad esempio, è stato avviato fra 1982 e 1988, quando ne è stata rilasciata una prima versione pubblica. Al giorno d'oggi, disponiamo di una grande quantità di risorse e interfacce disponibili, il tipo di ricerche che possiamo svolgere sul materiale è abbastanza limitata, ma dipende in gran parte dalla dimensione testuale: la via d'accesso a questa conoscenza è cioè delimitata da stringhe di testo in varie combinazioni (le *parole chiave*). Semplici o multicampo, le nostre abituali ricerche usano in buona parte gli 'operatori booleani' (AND, OR, NOT ecc.), sviluppati da George Boole nell'Irlanda di metà Ottocento (essi fanno parte della cosiddetta 'algebra booleana', un reticolo logico che ammette solo la coppia minima di presenza/assenza, 0/1 ecc.).<sup>1</sup>

In altre parole, si tratta di un sistema che ammette solo coppie binarie: siamo sicuri che sia lo strumento più idoneo per la ricerca umanistica, dove naturalmente un certo grado di ambiguità e incertezza interpretativa è tratto costitutivo della conoscenza? Questo intervento è dedicato a un progetto di concettualizzazione o mappatura semantica della conoscenza relativa ai commenti danteschi, in base alla quale le risorse siano reperibili in base ai loro contenuti, anche senza alcuna coincidenza verbale, e le ricerche non siano soggette al consueto "rumore di sottofondo" determinato dal grado di ambiguità delle parole chiave.

Le informazioni presenti sul Web tradizionale sono talmente vaste che, per realizzare quest'obiettivo, occorre operare in due direzioni opposte e complementari: da un lato codificare i dati in formati standard o compatibili (ad es. l'eXtensible Markup Language o XML); dall'altro progettare metodi di estrazione automatica delle informazioni, mediante protocolli d'interrogazione mirati. Nel caso di Dante, la necessità di un approccio interdisciplinare e ipermediale (nel senso di McGann 2014) da un lato, e la sterminata bibliografia dall'altro rendono imperativa la razionalizzazione e classificazione della conoscenza per mezzo di analisi automatica dei contenuti web già esistenti, da individuare mediante standard largamente condivisi, gli Universal Resource Identifiers (URI, sequenze di caratteri che identificano, senza margini di ambiguità, una risorsa web).<sup>2</sup> È questo il contesto che (secondo uno dei padri del

<sup>1</sup> S. Givant, P. Halmos, *Introduction to Boolean Algebras, Undergraduate Texts in Mathematics*, New York, Springer, 2009.

<sup>2</sup> Per allargare tali combinazioni a caratteri non presenti nei set ASCII, si parla oggi di [Internationalized Resource Identifiers](#) (IRI). Ambedue sono standard raccomandati dal consorzio W3C.

World Wide Web, Tim Berners-Lee) può favorire l'implementazione e diffusione del Semantic Web, proposto fin dai primi anni Duemila:<sup>3</sup> «Con il termine web semantico, [...] si intende la trasformazione del World Wide Web in un ambiente dove i documenti pubblicati (pagine web, file, immagini ecc.) siano associati ad informazioni (metadati) che ne specifichino contesto e relazioni in un formato adatto all'interrogazione (ad es. tramite motori di ricerca) e, più in generale, all'elaborazione automatica» (*Wikipedia*, s. v. *Web Semantico*).

Per articolare la conoscenza in un formato logicamente trasparente, lo standard che si è affermato è la *semantic triple*, o anche *triple*, dato elementare nello schema di descrizione della Resource Description Framework (RDF).<sup>4</sup> Si tratta di «A set of three entities that codifies a statement about semantic data in the form of subject–predicate–object expressions ... explicitly represent inference, from premises to conclusions» (*Wikipedia*, s. v. *Semantic Triple*). Un esempio di rapporto complesso che è stato codificato in questo modo è quello dei fenomeni dell'intertestualità e interdiscorsività nelle opere minori dantesche, oggetto del *tool Dante Sources*. Quest'ultimo può essere descritto come una biblioteca digitale che consente agli studiosi di:

- visualizzare i dati relativi alle fonti primarie citate da Dante nelle sue opere;
- visualizzare i dati riguardanti gli autori citati da Dante;
- visualizzare le aree tematiche a cui le opere citate appartengono;
- contestualizzare la distribuzione dei dati relativi ad autori ed opere citate da Dante all'interno del suo background culturale e nel tempo.

Per implementare il *tool Dante Sources* è stato utilizzato OWL ([Web Ontology Language](#)) 2, un linguaggio della famiglia RDF anch'esso raccomandato dal consorzio W3C: lo scopo di tali linguaggi RDF è descrivere delle basi di conoscenze, effettuare delle deduzioni su di esse e integrarle con i contenuti delle pagine web. OWL intende rendere possibile, ad esempio:

- ricerche nel web che superino i problemi di omonimia e ambiguità presenti nelle normali ricerche testuali;
- applicazioni che effettuino delle deduzioni sui dati disponibili.

La procedura che «legge» in grande quantità i dati e li 'traduce' in forme più complesse di conoscenza è svolta dall'*ontology / semantic reasoner*, il cui compito è applicare su larga scala gli assiomi definiti dal gruppo di lavoro secondo il linguaggio di codifica prescelto: il motore di ricerca semantico può dunque essere definito come «a software able to infer logical consequences from a set of asserted axioms [...]. The inference rules are commonly specified by means of an ontology language» (*Wikipedia*, s. v. *Semantic Reasoner*). Di conseguenza, il lavoro deve procedere tanto nella direzione della raccolta e annotazione dei dati di partenza, quanto nella progettazione di ontologie dedicate e di un idoneo *search engine* semantico che

<sup>3</sup> T. Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web*, «Scientific American Magazine», CCLXXXIV, 2001, 5.

<sup>4</sup> R. Cyganiak, D. Wood, M. Lanthaler, [RDF 1.1 Concepts and Abstract Syntax](#), W3C Recommendation, WWW Consortium, February 2014.

riorganizzi le conoscenze già maturate. Nel nostro caso, le fonti da utilizzare sono ovviamente i commenti e la bibliografia dantesca. Vale la pena di specificare che, nel nostro caso, il *search engine* analizza la combinazione di unità concettuali in un documento, in un sito web e nella query digitata dall'utente, senza più la necessità di un'esatta coincidenza tra le parole chiave, marcando dunque un superamento della ricerca tradizionale *text-based* di cui si parlava all'inizio.<sup>5</sup> In tal modo, il progetto *Hypermedia Dante Network* (HDN) ha lo scopo di costituire – sul fondamento degli standard del web semantico per le biblioteche digitali – un ambiente collaborativo per il commento delle opere dantesche, con attenzione alle questioni di lingua, stile e intertestualità.

## 2. Il progetto

In questa seconda parte daremo conto del progetto HDN, discutendo le sue fasi di lavoro principali e gli aspetti più rilevanti di ciascuna fase.

Ogni progetto software inizia con la raccolta e l'analisi dei requisiti. In questa fase, i futuri utenti del sistema specificano le loro esigenze funzionali e collaborano fattivamente alla costruzione delle interfacce grafiche che daranno accesso alle operazioni che implementano tali esigenze. Nel caso del progetto HDN, la fase di specifica delle esigenze funzionali è stata svolta in sede di scrittura della proposta di progetto, anche a partire dai risultati del progetto che di HDN è stato il precursore, il PRIN "Per una enciclopedia dantesca digitale". Per questa ragione, il progetto vero e proprio è partito dalla seconda fase, quella di progettazione del sistema, in cui si definiscono le componenti principali dell'architettura. Di questa fase interessa qui mettere in rilievo la metodologia seguita per la realizzazione del sistema, perché tale metodologia recepisce un metodo di lavoro innovativo e di grande interesse per le Digital Humanities.

Come già il progetto precursore, HDN adotta un approccio basato su due cardini fondamentali:

- la rappresentazione della conoscenza, e
- l'adozione di standard.

Riguardo al primo punto, il nucleo del sistema informatico è una base di conoscenza che raccoglie gli enunciati degli umanisti riguardo alle citazioni della *Commedia*, espressi in un linguaggio logico in cui sono definiti anche i termini del discorso umanistico. Tale linguaggio logico prende il nome di *ontologia*. Da notare come l'ontologia, essendo del tutto estranea a qualsiasi tecnologia implementativa, è il luogo ideale per l'incontro tra umanisti e informatici perché permette ai primi di specificare il significato dei termini del discorso umanistico in forma puramente logica, lasciando poi ai secondi la libertà di scegliere le tecnologie più adatte per implementare la base di conoscenza. In altre parole, il progetto adotta un approccio *dichiarativo* alla rappresentazione della conoscenza, che pone l'enfasi su *che cosa* debba essere rappre-

<sup>5</sup> Ho affrontato di recente questi temi negli ultimi due capitoli del volume M. Zaccarello, *Leggere senza libri. Conoscere gli e-book di letteratura italiana*, Firenze, Franco Cesati Editore, 2020, pp. 75-87.

sentato e non su *come*. Tale approccio, affermatosi agli inizi degli anni 80 nell'Intelligenza Artificiale per ragioni di natura ingegneristica, acquista oggi un valore aggiunto nelle Digital Humanities, perché favorisce la comunicazione tra soggetti (umanisti e informatici) portatori di saperi molto diversi.

Riguardo al secondo punto, l'adozione di standard ha l'obiettivo di salvaguardare l'investimento che il progetto fa in termini di capitale umano, rendendo il risultato dell'annotazione manuale dei testi danteschi il più possibile [FAIR](#) (Findable, Accessible, Interoperable e Reusable). Gli standard adottati sono:

- ontologia basata sugli standard del settore, in particolare il CIDOC CRM<sup>6</sup> (standard ISO e standard de facto in ambito Beni Culturali).<sup>7</sup>
- forma logica dell'ontologia, basata sul linguaggio OWL 2 DL,<sup>8</sup> raccomandato dal Comitato del Web.
- notazione, basata su RDF, RDFS<sup>9</sup> e SPARQL,<sup>10</sup> anch'essi raccomandati dal Comitato del Web.
- infrastruttura, basata sul Web e sul protocollo HTTP.

### 2.1. Concettualizzazione

Una volta fissati i criteri di realizzazione del sistema, si passa allo sviluppo dell'ontologia. Questa altro non è che una specifica logica di una concettualizzazione, ossia di un insieme di categorie su cui si fonda il discorso umanistico. In HDN, si parte dagli enunciati che costituiscono la base di conoscenza del sistema, che sono del tipo: 'Dice A che la conoscenza su B si può arricchire grazie a C' dove

- A, la sorgente, è una frase (tipicamente proveniente da una nota in un commento) che asserisce il riferimento e riflette una credenza del commentatore
- B, il soggetto, è un frammento di testo da un'opera (dantesca, nel nostro caso) che viene chiarito dal riferimento
- C è l'oggetto del riferimento, cioè l'entità cui il riferimento rimanda.

Nella maggioranza dei casi, l'oggetto è un frammento di testo. Ci sono tuttavia delle eccezioni:

<sup>6</sup> M. Doerr, *The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata*, «AI Magazine», XXIV, 2003, 3, pp. 75-92; DOI 10.1609/aimag.v24i3.1720.

<sup>7</sup> M. Doerr, *Ontologies for Cultural Heritage*, in S. Staab, R. Studer (eds.), *Handbook on Ontologies*, Switzerland, Springer, 2009, pp. 463-480.

<sup>8</sup> B. Motik, P. F. Patel-Schneider, B. Parsia, *OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition)*, W3C recommendation, W3C, December 2012.

<sup>9</sup> D. Brickley, R.V. Guha, *RDF Schema 1.1*, W3C Recommendation, WWW Consortium, February 2014.

<sup>10</sup> The W3C SPARQL Working Group, *SPARQL 1.1 Overview*, W3C Recommendation, 21 March 2013.

- l'oggetto può essere un insieme di opere caratterizzato estensionalmente, nel qual caso è una collezione (e.g., 'le opere dei maestri parigini'), o intensionalmente, nel qual caso è un concetto (e.g., 'le opere del tardo romanticismo').
- l'oggetto può essere un'entità, (e.g., un personaggio, un'opera, una teoria) non necessariamente riferibile a un frammento di testo specifico; in questo caso l'oggetto non è specificato e l'entità riferita è data come contenuto.

In aggiunta alle tre caratteristiche fondamentali viste sopra, un riferimento ha:

- un tipo: riferimento stringente, esplicito o generico;
- contenuto: a che cosa ci si riferisce. La concettualizzazione individua tre macro categorie, all'interno delle quali si individuano sotto-categorie. Le macro categorie sono:
  1. Corrispondenze testuali, ulteriormente suddivise in:
    - immagine (il commentatore individua un'immagine presente in una fonte come precedente per una descrizione o concettualizzazione di Dante)
    - stilema (il commentatore identifica un precedente linguistico e/o retorico per un'espressione dantesca)
  2. Corrispondenze tematiche, ulteriormente suddivise in:
    - personaggio (il commentatore ricollega un personaggio menzionato da Dante a una fonte specifica)
    - episodio (il commentatore identifica un episodio letterario alla base di un riferimento dantesco)
    - topografia (il commentatore ritiene che dietro la menzione dantesca di un luogo agisca una memoria letteraria)
  3. Corrispondenze concettuali, ulteriormente suddivise in:
    - motivo (il commentatore identifica in una fonte precisa il precedente per lo sviluppo di un tema dantesco)
    - teoria (il commentatore identifica un brano teorico alla base di un passo del discorso dantesco)
- funzione: lo scopo del riferimento: identificativo, esplicativo (il riferimento chiarisce concetti espressi nell'opera dantesca), intratestuale (il riferimento offre corrispondenze testuali interne alle opere di Dante), Intertestuale (il riferimento offre corrispondenze testuali esterne all'opera dantesca), Interdiscorsivo (il riferimento offre corrispondenze concettuali esterne all'opera dantesca).
- rapporto retorico tra soggetto e oggetto: palinodia/contraddizione, ampliamento/rielaborazione, conferma/ omologia
- altri riferimenti rilevanti al riferimento presente.

Infine, si considerano tre categorie di riferimento:

- citazione, incluse le auto-citazioni
- supporto esterno
- luogo parallelo

## 2.2. Ontologia

L'ontologia specifica la concettualizzazione, fornendo nomi alle categorie (classi) e alle loro caratteristiche (proprietà), e usando tali nomi per esprimere assiomi che catturano la struttura individuata dalla concettualizzazione e legano classi e proprietà a ontologie standard.

La Figura 1 mostra (in celeste) le classi che catturano le categorie del discorso riguardante i riferimenti. Le classi sono tra loro collegate da archi direzionati che esprimono la relazione di sottoclasse, propria delle tassonomie. La direzione di ciascun arco va dalla classe più specifica a quella più generica. Quindi, per esempio, la classe dei Riferimenti (*Reference*, i nomi delle classi sono dati in inglese per maggior generalità) ha due sottoclassi principali: Supporto Esterno (*External Support*) e Luogo Parallelo (*Loci Paralleli*), la quale ha a sua volta la classe Citazione (*Citation*) come caso speciale.

Le classi in bianco nel diagramma sono classi dell'ontologia standard di riferimento, il CIDOC CRM. Tali classi occupano i vertici della tassonomia a significare il fatto che le classi che catturano i concetti HDN si ottengono per raffinamento dalle classi CRM [vd. Fig. 1].

Il discorso umanistico non si esaurisce però con i riferimenti e le citazioni: la contestualizzazione di questi richiede l'espressione di enunciati riguardanti:

- persone, per esempio, quali autori dei commenti, o quali oggetto di riferimenti
- evento, come oggetto di riferimenti
- concetto, ancora come oggetto di riferimenti o topica di un'opera letteraria
- periodo temporale,

e altro ancora. Tali classi sono adottate direttamente dal CIDOC CRM che, come ontologia top, ha una copertura concettuale pressoché universale.

In aggiunta alle classi, l'ontologia definisce proprietà, necessarie per esprimere le caratteristiche degli individui del discorso, collegando fra loro le risorse che li rappresentano. Nel nostro caso, le proprietà dell'ontologia HDN esprimono:

- la sorgente, il soggetto, l'oggetto, il tipo, il contenuto di un riferimento
- l'opera di appartenenza e il contenuto testuale di una sorgente, di un soggetto o di un oggetto
- l'autore, il titolo, l'edizione, e tutti gli altri dati bibliografici di un'opera
- e altro ancora.

Come le classi, anche le proprietà dell'ontologia HDN sono mutate dall'ontologia standard di riferimento CIDOC CRM o derivate per raffinamento dalle proprietà di questa. A titolo esemplificativo, mostriamo due diagrammi riguardanti le proprietà di HDN.

Il primo diagramma [vd. fig. 2] mostra le proprietà che servono per collegare fra loro le parti di un'opera, e queste ai frammenti che sono sorgenti, soggetti o oggetti di riferimenti. Ogni opera o parte di opera è un'istanza della classe F2 Expression, che è una classe

dell'ontologia LRMoo,<sup>11</sup> un'estensione del CRM che tratta risorse bibliografiche. Un'opera è collegata alle sue parti, e queste alle sottoparti relative, fino al livello di granularità più fine, tramite la proprietà P148 has component del CIDOC CRM. Quindi la *Commedia* e le tre cantiche che la compongono sono istanze di F2 Expression, e l'opera è collegata a ciascuna cantica dalla proprietà P148. Analogamente, i canti di ciascuna cantica sono istanze di F2 Expression e la cantica relativa li connette ancora tramite P148. La stessa struttura logica permette di collegare un commento alle note che lo compongono. I frammenti testuali sono istanze di Text Fragment e la proprietà che lega la più piccola parte di un'opera al frammento che essa contiene è la proprietà P106 is composed of, di nuovo una proprietà del CIDOC CRM. Osserviamo quindi che l'espressione di questa conoscenza sulla struttura delle opere e sull'appartenenza dei frammenti non richiede l'introduzione di proprietà specifiche, ma può essere fatta semplicemente riusando le proprietà del CIDOC CRM e della sua estensione bibliografica, LRMoo.

Il secondo diagramma [vd. fig. 3] mostra le caratteristiche del riferimento e le proprietà dell'ontologia che le collegano al riferimento stesso. Come abbiamo visto nella sezione precedente, un riferimento ha una sorgente (una frase) e un soggetto (un frammento di testo). Come mostra il diagramma:

- La proprietà che collega un riferimento alla sua sorgente è hasSource, che è una sottoproprietà della proprietà P129 is the subject of del CIDOC CRM.
- Similmente, la proprietà che collega un riferimento al suo soggetto è hasSubject, che è una sottoproprietà della proprietà P106 is composed of del CIDOC CRM.

Il diagramma esemplifica il ruolo dell'ontologia di riferimento anche nelle proprietà.

Concludiamo questa sezione sull'ontologia HDN mettendo in evidenza la funzione che svolge l'ontologia in un sistema come quello che sta realizzando HDN. Queste sono due:

1. documentare la concettualizzazione, specificando in un linguaggio preciso e rigoroso le classi, le proprietà e gli assiomi che la costituiscono, con ciò fornendo ai fruitori della conoscenza un quadro completo del vocabolario usato in essa.
2. calcolare la conoscenza implicita grazie al motore di inferenza dell'ontologia è possibile derivare conoscenza implicita, ossia conoscenza non esplicitamente immessa dall'utente, ma derivata dal sistema sulla base di questa. Questo calcolo emerge in due circostanze:
  - quando si interroga il sistema; per esempio, ogni citazione essendo un riferimento, quando un utente chiede la lista dei riferimenti conosciuti al sistema questo include nel computo anche le citazioni, sebbene ciò non sia stato esplicitamente indicato dall'utente. Questa inferenza è analoga alla ben nota infe-

<sup>11</sup> B. Chryssoula, M. Doerr, P. Le Boeuf, P. Riva (eds.), *LRMoo (Formerly FRBRoo) Object-Oriented Definition and Mapping from IFLA LRM (Version 0.6)*, International Working Group on LRM, FRBR and CIDOC CRM Harmonisation, February 2020.



renza che porta a concludere per via sillogistica che Socrate è mortale dalle premesse che Socrate è un uomo e che ogni uomo è mortale.

- quando si testa la consistenza della base di conoscenza; per esempio, se l'ontologia indica che un canto può appartenere a una sola cantica mentre nel sistema è stata erroneamente inserita la conoscenza che un canto appartiene a due cantiche, il sistema dedurrà che le due cantiche sono la stessa cantica, salvo rilevare l'inconsistenza una volta che sia stato segnalato che le tre cantiche della *Commedia* sono tutte diverse tra loro.

Il calcolo della conoscenza implicita è un valore aggiunto di enorme importanza che le ontologie permettono mentre altri schemi di rappresentazione, per esempio quello basato su XML o su basi di dati relazionali, permettono in misura minore o non permettono affatto.

### 2.3. Implementazione

In questa sezione descriviamo brevemente l'implementazione dell'ontologia e della base di conoscenza che raccoglie gli enunciati degli umanisti riguardo ai riferimenti della *Commedia*.

L'ontologia viene creata usando un editore che conosce la sintassi del linguaggio OWL 2 DL. L'editore impiegato nel progetto è [WebProtégé](#)<sup>12</sup> che permette di esprimere, modificare e visualizzare gli assiomi OWL 2 DL in modo amichevole, supportando inoltre la collaborazione tra più soggetti allo sviluppo della stessa ontologia.

Da WebProtégé l'ontologia viene esportata come un insieme di triple RDF (detto *grafo*) in un file secondo una fra un insieme di notazioni più comuni. Nel nostro progetto abbiamo scelto la notazione Turtle in cui le triple di un grafo RDF sono semplicemente listate una dopo l'altra, con l'ausilio di abbreviazioni che rendono la lettura particolarmente agevole per un umano.

Il file contenente il grafo viene inserito in una speciale componente software, detta triple store, che ne supporta la persistenza e l'accesso secondo un'interfaccia programmatica standard, quella del linguaggio SPARQL. La componente scelta per il progetto è [BlazeGraph](#).

Il popolamento della base di conoscenza con gli enunciati degli umanisti viene effettuato tramite un *tool* di inserimento della conoscenza, sviluppato ad hoc per il progetto HDN, come versione evoluta dell'analogo *tool* usato per il progetto predecessore di HDN, che ha realizzato l'applicazione [DanteSources](#). Il *tool* è un'applicazione web, basata sulla concettualizzazione discussa nella sezione precedente, che visualizza una per una le note dei commenti e supporta gli umanisti nell'estrazione dei riferimenti da ciascuna nota, guidandoli nella specifica delle caratteristiche di ogni riferimento.

La conoscenza inserita dagli umanisti viene processata da un altro *tool*, detto triplificatore, che la riduce in triple RDF, che vengono a loro volta aggiunte al triple store che contiene l'ontologia. Le triple e l'ontologia vanno a formare la base di conoscenza su cui insistono le funzionalità di accesso messe a disposizione degli utenti.

<sup>12</sup> T. Tudorache, C. Nyulas, N. F. Noy, M. A. Musen, *WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web*, «Semantic Web», IV, 2013, 1, pp. 89-99.

L'accesso alla base di conoscenza così costruita avverrà attraverso una web app molto simile a *DanteSources*, che permette all'utente di formulare la sua richiesta in modo amichevole con il supporto della grafica. Per esempio, la richiesta può riguardare la visualizzazione delle citazioni all'*Etica* di Aristotele contenute nel X canto dell'*Inferno*.

La richiesta utente intercettata dall'interfaccia viene trasformata dalla web app in una interrogazione (query) SPARQL<sup>13</sup> e inviata al triple store per essere risolta. A questo scopo, il triple store mette a disposizione uno SPARQL end-point, che è un canale sul quale si pone in ascolto per ricevere le query alla base di conoscenza. Ricevuta una query sull'end-point, il triple store ne calcola il risultato, previa il calcolo della conoscenza implicita, come già segnalato. Il risultato della query viene ricevuto dalla web app che lo formatta graficamente per essere mostrato all'utente. L'utente può esportare il risultato in un formato digitale di sua scelta tra quelli proposti.

Il pattern di accesso appena descritto è pensato per l'utente umano, che viene facilitato nell'interazione con il triple store dalla web app. Ma esso si presta ad essere eseguito anche da un agente digitale che attinge alla stessa conoscenza per conto di un'applicazione diversa da quella di HDN. Tale agente dovrà inviare una query SPARQL al triple store sullo SPARQL end-point messo a disposizione da quest'ultimo, bypassando così la web app, ricevendo in risposta il risultato per elaborarlo per conto dell'applicazione che tale agente serve.

Una ulteriore modalità di accesso si ha grazie al fatto che la base di conoscenza HDN è un dataset Linked Data: ogni URI (IRI) del dominio HDN viene risolto dal web server di HDN e restituisce una rappresentazione (dello stato attuale) della risorsa riferita: pagina web per gli umani e grafo RDF per tutti gli altri.

Tale rappresentazione è calcolata previa una query SPARQL (per esempio di tipo DESCRIBE) al triple store e una formattazione del risultato nel formato appropriato. Questo ultimo pattern di accesso permette di riusare gli IRIs creati dal progetto HDN nel contesto di altre basi di conoscenza, creando una porzione della rete semantica che viene normalmente indicata con il nome di *web semantico*. L'attenzione posta dal progetto nel ricondurre i termini della propria ontologia a un'ontologia standard, il CIDOC CRM, massimizza la "comprensibilità" della conoscenza prodotta da HDN da parte di comunità estranee al progetto ed è un *best effort* all'interoperabilità di tale conoscenza.

### 3. Conclusioni

L'articolo ha presentato i tratti essenziali del progetto *Hypermedia Dante Network*, finalizzato alla costruzione di una base di conoscenza sui riferimenti presenti nella *Commedia* e alla realizzazione di un portale per l'accesso a tale conoscenza in maniera guidata. La presentazione si è focalizzata sull'approccio, basato sulla rappresentazione della conoscenza e sull'adesione agli standard del web Semantico. Punto nodale di entrambe è l'ontologia, che fissa il significato dei termini usati nella base di conoscenza in un linguaggio standard, favorendo il calcolo della conoscenza implicita e della sua consistenza, ma anche l'interpretazione dei dati e la loro interoperabilità.

<sup>13</sup> S. Harris, A. Seaborne, [SPARQL 1.1 Query Language](#), W3C Recommendation, 21 March 2013.

DOSSIER ICONOGRAFICO

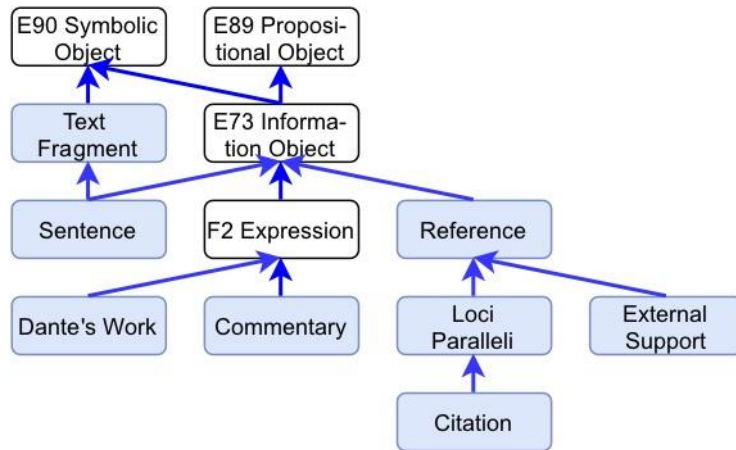


Fig. 1

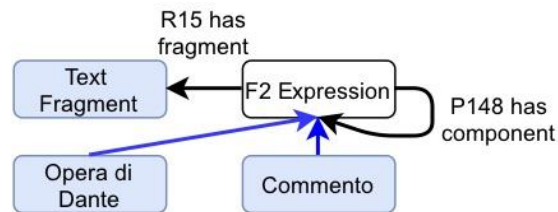


Fig. 2

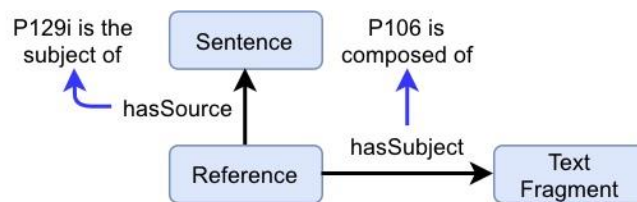


Fig. 3