



Università degli Studi di Udine

Dipartimento di studi umanistici

*Dottorato in Scienze bibliografiche, del testo e del documento
XXVII ciclo*

Interoperabilità e semantica
nei domini specialistici: l'approccio Linked Data
nel settore delle costruzioni

Dottoranda
Giovanna Aracri

Tutor
Prof. Mauro Guerrini

Co-tutor
Prof. Federico Valacchi

Anno Accademico 2014/2015

SOMMARIO

SOMMARIO.....	1
INDICE DELLE FIGURE.....	2
INTRODUZIONE.....	3
CAPITOLO 1 - IL CONCETTO DI INTEROPERABILITÀ: LA DIMENSIONE TECNICA E LA DIMENSIONE SEMANTICA	7
1. <i>Analisi del concetto di interoperabilità.....</i>	<i>9</i>
2. <i>L'interoperabilità nelle scienze dell'informazione</i>	<i>13</i>
3. <i>I sistemi di organizzazione della conoscenza come strumenti di supporto dell'interoperabilità semantica</i>	<i>19</i>
4. <i>Il concetto di interoperabilità nel settore delle costruzioni</i>	<i>26</i>
CAPITOLO 2 - IL WEB SEMANTICO E L'APPROCCIO LINKED DATA	32
1. <i>Dal web tradizionale al web semantico.....</i>	<i>33</i>
2. <i>L'approccio Linked Data come strumento di supporto al web semantico</i>	<i>44</i>
2.1. <i>Caratteristiche e principi dell'approccio linked data</i>	<i>46</i>
2.2. <i>I linguaggi dell'approccio Linked Data.....</i>	<i>49</i>
2.2.1. <i>Resource Description Framework (RDF).....</i>	<i>50</i>
2.2.2. <i>SPARQL Protocol and RDF Query Language (SPARQL)</i>	<i>54</i>
3. <i>La diffusione dell'approccio Linked Data</i>	<i>57</i>
CAPITOLO 3 - ANALISI DELLO STATO DELL'ARTE	62
1. <i>I KOS disponibili in modalità Linked Data</i>	<i>63</i>
2. <i>L'organizzazione della conoscenza di dominio</i>	<i>83</i>
CAPITOLO 4 - METODOLOGIA PER LA REALIZZAZIONE DELLO SCHEMA DI CORRISPONDENZE.....	96
1. <i>Il trasferimento del Lessico INNOVance in Protégé</i>	<i>96</i>
2. <i>Identificazione delle risorse terminologiche e semantiche da mettere in comunicazione</i> <i>104</i>	
3. <i>La scelta del modello da impiegare per il mapping tra le risorse.....</i>	<i>106</i>
4. <i>Definizione dei mapping tra il lessico INNOVance, il Nuovo Soggettario e DBpediaItalia</i> <i>111</i>	
CONCLUSIONI	115
BIBLIOGRAFIA	118
SITOGRAFIA.....	124
APPENDICE	127

INDICE DELLE FIGURE

Figura 1. Tipologie di Knowledge Organization System.	23
Figura 2. La visione originaria del web.....	34
Figura 3. Elementi essenziali del web: http, URL, <i>tag</i> HTML.....	35
Figura 4. Pagina di disambiguazione di Wikipedia per il termine <i>quadro</i>	36
Figura 5. Visione a strati del web semantico.	39
Figura 6. Due visioni a confronto: il web ipertestuale e il web semantico.....	43
Figura 7. Esempificazione della modalità di arricchimento semantico del Nuovo Soggettario.	45
Figura 8. Il criterio delle 5 stelle.	47
Figura 9. Rappresentazione del modello di descrizione RDF.....	50
Figura 10. Grafo RDF dell'asserzione <i>Umberto Eco è autore di "Il nome della rosa"</i>	51
Figura 11. Query SPARQL - <i>Tutti i film di produzione italiana</i>	55
Figura 12. Query SPARQL - <i>Elenco dei senatori attualmente in carica</i>	55
Figura 13. Rappresentazione tabulare di una query SPARQL.	56
Figura 14. <i>LOD data cloud</i> aggiornata a Aprile 2014.....	57
Figura 15. RDF/XML e collegamenti semantici esterni nel Nuovo Soggettario.....	66
Figura 16. Esempificazione della struttura di un concetto in AGROVOC ed esplicitazione di link esterni.....	69
Figura 17. I <i>mapping</i> generati.	70
Figura 18. Il modello ontologico SKOS di EuroVoc.	73
Figura 19. Struttura organizzativa di GEMET e evidenza delle correlazioni esterne.	75
Figura 20. Strutturazione in EARTH e specificazione del collegamento esterno a GEMET.....	77
Figura 21. Intestazione di soggetto della LBSH e specificazione dei collegamenti esterni.	79
Figura 22. Corrispondenza tra una pagina Wikipedia e una risorsa DBpedia.....	81
Figura 23. criteri di indirizzo della UNI 11337.	90
Figura 24. Scheda terminologica nel lessico INNOVance.	97
Figura 25. Definizione della classe Lessico_INNOVance.....	99
Figura 26. Confronto tra la struttura del lessico INNOVance e SKOS.....	102
Figura 27. <i>Object properties</i> definite.	103
Figura 28. Specificazione delle relazioni per il termine <i>cemento</i>	103
Figura 29. Modello prescelto.	108
Figura 30. Esempificazione della modalità di <i>matching</i> adottata.	112
Figura 31. Le tre classi rappresentative dei tre vocabolari.	113
Figura 32. I tag per esprimere i <i>mapping</i> esterni	113
Figura 33. Mapping esterni e relazioni interne.....	114

INTRODUZIONE

La presente attività di ricerca si propone innanzitutto, di analizzare il concetto di interoperabilità, in relazione alla dimensione tecnica e a quella semantica, come prerequisito per lo scambio e la condivisione di dati,¹ informazioni² e conoscenza³ sul web, e successivamente di indagare più a fondo il concetto di interoperabilità semantica nel settore delle costruzioni, attraverso la realizzazione di uno schema di corrispondenze tra concetti appartenenti a risorse terminologiche e semantiche differenti.

L'obiettivo consiste, quindi, nel definire collegamenti tra concetti, tenendo conto delle buone pratiche raccomandate dal web semantico, ed in particolar modo, dall'approccio linked data.⁴ L'attenzione è rivolta all'insieme dei termini e dei significati accettati e convenzionalmente impiegati per la designazione e la caratterizzazione dei prodotti da costruzione,⁵ inteso come sotto-dominio del settore delle costruzioni, che costituisce il campo di applicazione prescelto.

Nello specifico, è stata indagata l'interoperabilità semantica tra i termini del lessico specialistico strutturato INNOVance⁶ e i termini ospitati dal thesaurus multidisciplinare del Nuovo Soggettario⁷ e dalla base di conoscenza DBpediaItalia,⁸ con lo scopo di

¹ "A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing." «cfr.» ISO/IEC 2382-1:1993(EN) Information technology – Vocabulary – Part 1: Fundamental terms, 1993.

² "Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning". Ibidem.

³ "La conoscenza è nota come l'abilità di trasformare ed interpretare dati, informazioni e idee al fine di eseguire diverse attività, prendere decisioni e risolvere problemi". «cfr.» MARIA TERESA GUAGLIANONE, MNEMO. Una metodologia combinata per l'acquisizione e la modellizzazione della conoscenza delle organizzazioni, Tesi di Dottorato Università della Calabria, 2012.

⁴ "Is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods [...]" <<http://linkeddata.org/>>.

⁵ "Prodotto da costruzione: "Qualsiasi prodotto o kit fabbricato e immesso sul mercato per essere incorporato in modo permanente in opere di costruzione o in parti di esse e la cui prestazione incide sulla prestazione delle opere di costruzione rispetto ai requisiti di base delle opere stesse". «cfr.» Regolamento UE N. 305/2011 del Parlamento Europeo e del Consiglio, 9 marzo 2011, Condizioni armonizzate per la commercializzazione dei prodotti da costruzione e che abroga la direttiva 89/106/CEE del Consiglio.

⁶ Il lessico INNOVance è stato realizzato dall'Istituto per le Tecnologie della Costruzione del Consiglio Nazionale delle Ricerche (ITC-CNR) nell'ambito dell'omonimo progetto finanziato dal Ministero dello Sviluppo Economico (MISE), il cui obiettivo consiste nella realizzazione della prima banca dati nazionale per il settore delle costruzioni, all'interno del quale organizzare, strutturare e codificare in maniera univoca i soggetti, gli oggetti e le azioni coinvolti nel processo edilizio. Consultabile al seguente indirizzo <<http://www.innovance.it/lessico>> (ultima consultazione 16/12/2014).

⁷ <<http://thes.bncf.firenze.sbn.it/ricerca.php>> (ultima consultazione 16/12/2014).

⁸ <<http://it.dbpedia.org/>> (ultima consultazione 16/12/2014).

delineare uno schema di corrispondenze tra i concetti e la conoscenza del dominio delle costruzioni, che questi sistemi esprimono.⁹

La necessità di procedere alla definizione di siffatte correlazioni, è dovuta all'assenza in ambito nazionale di un sistema di organizzazione e gestione della conoscenza¹⁰ di riferimento per questo specifico settore, che consideri contemporaneamente i diversi contesti in cui la terminologia viene impiegata e l'uso che ne viene fatto da parte dei soggetti che quotidianamente sono coinvolti negli scambi comunicativi e nella filiera edilizia.

Il linguaggio del settore delle costruzioni, infatti, è caratterizzato dal susseguirsi da frequenti aggiornamenti tecnici e normativi, dalle sempre nuove esigenze avvertite dagli operatori e da vari livelli di specializzazione e di conoscenza posseduti da parte dei numerosi attori coinvolti nel processo edilizio. La commistione tra questi fattori, può determinare incoerenza terminologica e ambiguità semantica a discapito della diffusione e condivisione dell'informazione.

Dati questi presupposti, è evidente la necessità di procedere alla predisposizione di adeguati strumenti di gestione, controllo e organizzazione dell'informazione, che tengano conto delle peculiarità e delle specificità del linguaggio e che forniscano contestualmente, una visione quanto più esaustiva, congruente, condivisa ed interoperabile della conoscenza di dominio, di modo da supportare gli scambi comunicativi, favorire l'interazione tra più utenti, ridurre le incomprensioni ed evitare la perdita di informazioni utili.

Indipendentemente dall'ambito di applicazione, tali sistemi svolgono un ruolo chiave soprattutto in ambiente digitale, caratterizzato dalla proliferazione di dati e informazioni non sempre di qualità. In questi contesti l'interoperabilità intesa nella sua duplice accezione tecnica e semantica, assume un valore preminente, tant'è vero che per ottenere il suo pieno conseguimento è importante disporre, non soltanto di strumenti e soluzioni di natura prettamente tecnologica, ma anche di adeguati strumenti di organizzazione e gestione della conoscenza, chiamati in causa dalle attuali prospettive del web.

Si sta infatti, via via affermando una nuova visione basata sul principio dell'interconnessione, in cui i dati e le informazioni prodotti dai vari provider

⁹ La scelta è ricaduta su questi sistemi in ragione della loro diffusione e del loro largo utilizzo a supporto del processo di recupero dell'informazione.

¹⁰ A tali sistemi è dedicata un'ampia trattazione nel Cap. 1, par.3.

necessitano di essere messi in comunicazione attraverso la specificazione di collegamenti tipizzati che rendono evidente la natura delle possibili relazioni.

Questi principi sono alla base del web semantico e dell'approccio linked data, riconosciuto come strumento principale attraverso cui realizzare la visione di interconnessione avanzata da Tim Berners Lee (ideatore del web attuale e precursore del web semantico). Secondo questa prospettiva è importante che i contenuti delle pagine web siano resi interoperabili, che siano strutturati e resi accessibili mediante la predisposizione di collegamenti concettuali e funzionali tra vocabolari controllati e condivisi.

In questo contesto denso di opportunità e di possibilità di sviluppo, si colloca la presente attività di ricerca, il cui obiettivo è indagare, esaminare e mettere in luce le correlazioni semantiche tra sistemi di organizzazione e gestione della conoscenza differenti, al fine di valorizzare la conoscenza del dominio delle costruzioni e mantenere il livello di dettaglio e di specificità che la contraddistingue, tenendo in ampia considerazione i diversi contesti d'uso del linguaggio.

L'attività di *mapping* consente di fatto, di far emergere le relazioni semantiche che intercorrono tra i termini, e i concetti da essi veicolati, provenienti da repertori terminologici e semantici eterogenei, e di rendere esplicita la struttura concettuale derivante dalla loro connessione. Si tratta quindi di un processo mediante il quale è possibile connettere potenzialmente tutto ciò che è espresso dai sistemi di partenza, creando una vera e propria rete di relazioni tra oggetti (*things*).

L'attività condotta, ha tenuto conto delle raccomandazioni predisposte dalla recente normativa in materia di thesauri e interoperabilità con altri vocabolari¹¹: la ISO 25964-2:2013 *Information and Documentation – Thesauri and interoperability with other vocabularies, Interoperability with other vocabularies* la quale predispone le raccomandazioni da seguire per lo sviluppo di vocabolari finalizzati al recupero delle informazioni. Inoltre, le correlazioni predisposte ricorrendo al *Simple Knowledge Organization System* (SKOS),¹² hanno beneficiato dell'analisi di alcune tra le più note esperienze maturate principalmente in ambito internazionale e propriamente rivolte all'interoperabilità come prerequisito per la condivisione e lo scambio dell'informazione.

¹¹ “[...] include classification schemes, file plans, taxonomies, subject headings schemes, ontologies, terminologies, name authority list and synonym rings. «cfr.» ALAN GILCHRIST, JUDI VERNAU, *Facets of Knowledge Organization: Proceedings of the ISKO UK Second Biennial Conference, 4th - 5th July, London, 2011*, p. 131, Emerald Group Publishing Limited, 2012. ISBN-10: 1780526148. ISBN-13: 978-1780526140.

¹² Il linguaggio SKOS pensato per definire correlazioni semantiche tra sistemi di organizzazione della conoscenza eterogenei, sarà ampiamente discusso nel Capitolo 4 inerente alla metodologia impiegata per la definizione dei collegamenti semantici.

Infatti, l'obiettivo ultimo consiste nell'indirizzare gli esperti di dominio e i non esperti verso un uso efficace e condiviso della terminologia di dominio di modo da ridurre possibili casi di *misunderstanding* e di incoerenza concettuale, per incoraggiare e agevolare il trasferimento preciso e non ambiguo delle informazioni, durante tutte le fasi del processo edilizio.

Nel primo capitolo sarà analizzato il concetto di interoperabilità e il valore che assume in relazione ai sistemi di organizzazione e gestione della conoscenza e più specificatamente nel settore delle costruzioni, in qualità di ambito di applicazione prescelto; nel secondo capitolo saranno presentate le principali differenze che sussistono tra il web tradizionale e il web semantico ed, inoltre, illustrate le potenzialità derivanti dall'uso di convenzioni e linguaggi condivisi per favorire l'interoperabilità sia tecnica che semantica sul web; nel terzo capitolo saranno passate in rassegna alcune tra le più note iniziative riguardanti la pubblicazione di sistemi di organizzazione e gestione della conoscenza in SKOS inerenti a diversi ambiti del sapere e in aggiunta sarà delineato lo stato dell'arte dei sistemi predisposti per il settore delle costruzioni; nel quarto capitolo saranno descritti i modi e le fasi che hanno portato al raggiungimento degli obiettivi proposti ossia al colloquio tra fonti terminologiche e semantiche eterogenee.

CAPITOLO 1

IL CONCETTO DI INTEROPERABILITÀ:

LA DIMENSIONE TECNICA E LA DIMENSIONE SEMANTICA

Il progresso tecnologico accompagnato da una crescente diffusione di sistemi in grado di interconnettere, rendere accessibili e riutilizzabili grandi quantità di dati e informazioni, sono fattori che attengono all'interoperabilità ed in particolar modo, all'insieme delle potenzialità che possono e potrebbero scaturire da una sua progressiva e corretta applicazione.

L'interoperabilità in termini estensivi, si propone di mettere in comunicazione tra loro due o più sistemi, altrimenti incompatibili, poiché rappresentativi di specifiche esigenze, non sempre riconosciute e accettate al di fuori del contesto in cui sono state percepite. Anche nei campi di studio delle scienze dell'informazione e delle discipline bibliografiche e biblioteconomiche, l'interoperabilità è un argomento di discussione molto sentito, che richiede un approccio interdisciplinare, affinché si sviluppino sinergie tra realtà che fino a questo momento hanno operato in maniera indipendente.

Dati questi presupposti e considerato il ruolo di rilevanza che l'interoperabilità è chiamata a svolgere, si ritiene importante procedere all'analisi del suo valore concettuale sia in relazione alla dimensione tecnologica, derivata dall'Informatica; che a quella semantico-concettuale, maggiormente connessa agli ambiti umanistici, con l'obiettivo di mettere in rilievo opportunità che possono derivare in termini di efficienza, correttezza, affidabilità e consistenza di dati e informazioni trasferiti, ma anche, di economicità del processo di generazione e mantenimento degli stessi.

Talune iniziative portate avanti in ambito specialistico e finalizzate allo studio dell'interoperabilità semantica delle informazioni medico sanitarie ospitate nell'*Electronic Health Record* (EHR)¹³ o dei dati nelle Infrastrutture di Dati Territoriali

¹³ A livello internazionale è stata predisposta la ISO 13606:2008 *Health informatics - Electronic health record communication*, il cui obiettivo è fornire un'architettura informatica interoperabile per la condivisione, la comunicazione e il trasferimento dei dati clinici. Nello specifico il corpo normativo della ISO si compone di cinque parti: la Parte 1:2008 - *Reference model* specifica le caratteristiche e i componenti generali della cartella clinica elettronica; la Parte 2:2008 - *Archetype interchange specification* specifica l'architettura delle informazioni per la comunicazione e il trasferimento; Parte 3:2009 - *Reference archetypes and term lists* riguarda la comunicazione di parti o dell'intera cartella clinica elettronica e

(SDI),¹⁴ danno evidenza del suo impiego a supporto dei processi di gestione, organizzazione e mantenimento di dati, informazioni e conoscenza in maniera condivisa e partecipata in diversi ambiti del sapere.

La descrizione e la rappresentazione di contenuti (dati e informazioni) secondo principi e standard condivisi, sono i fondamenti di due tra le più interessanti azioni avviate a livello europeo. La prima, promossa nell'ambito dell'Agenda Digitale Europea¹⁵ è rivolta alla definizione di strategie e strumenti comuni in grado di favorire l'accesso, la consultazione e lo scambio di informazioni in un ambiente condiviso ed interoperabile a supporto della crescita economica e sociale. La seconda, invece, che coinvolge diverse istituzioni europee (biblioteche, musei e archivi), è il progetto Europeana¹⁶ grazie al quale gli oggetti digitali (libri, film, archivi, archivi sonori, ecc.) provenienti da diversi paesi europei, sono raccolti in un unico portale, resi disponibili e liberamente accessibili. Questi due progetti, nonostante siano destinati a contesti differenti, perseguono il medesimo fine. Entrambi infatti, mirano a garantire l'accesso e il recupero dell'informazione in maniera partecipata, sfruttando le potenzialità degli strumenti informatici e l'espressività dei linguaggi impiegati per la rappresentazione dell'informazione da essi veicolata.

Le problematiche connesse alla mancata predisposizione di comuni strategie di organizzazione, comunicazione e diffusione delle informazioni (come ad esempio l'assenza di risorse terminologiche e semantiche di riferimento), che possono causare incomprensioni o perdita di informazione, sono essenzialmente dovute all'insufficiente attenzione rivolta all'interoperabilità semantica.

In questo capitolo sarà fornita un'analisi del concetto di interoperabilità, dei vantaggi e delle difficoltà che una sua applicazione può comportare, attraverso: la disamina e il raffronto di talune definizioni proposte dall'Organizzazione Internazionale di

fornisce le liste di termini da impiegare per rappresentare l'insieme dei valori che un certo attributo può assumere; Parte 4:2009 – *Security* descrive l'approccio metodologico da usare per la definizione e la specificazione delle modalità di accesso ai dati conservati nella cartella clinica elettronica; Parte 5:2010 - *Interface specification* fornisce l'interfaccia per il popolamento della cartella clinica elettronica e per l'interrogazione e la fruizione dei dati in essa contenuti. <<http://www.iso.org/iso/home.htm>> (ultima consultazione 16/12/2014).

¹⁴In Europa il punto di riferimento è rappresentato dalla direttiva 2007/2/CE del Parlamento e del Consiglio Europeo del 14 marzo 2007 che ha istituito un'infrastruttura tecnologica comunitaria per la gestione, l'organizzazione e il monitoraggio dei dati territoriali, ovvero l'*INfrastructure for Spatial InfoRmation in Europe* (INSPIRE). INSPIRE si propone quindi di superare le problematiche inerenti alla disponibilità, alla condivisione e al riuso dei dati di carattere geo-spaziale fornendo contestualmente un valido supporto per la predisposizione di politiche ambientali. In Italia tale direttiva è stata recepita con decreto legislativo del 27 gennaio 2010, n. 32. <<http://www.minambiente.it/pagina/inspire>>, (ultima consultazione 16/12/2014).

¹⁵L'Agenda Digitale si compone di 101 azioni e 7 pilastri e si propone di supportare la crescita economica dell'UE e consentire ai cittadini e alle imprese europei di utilizzare in maniera efficace ed efficiente l'insieme delle tecnologie digitali. Le esigenze legate all'interoperabilità rientrano nel pilastro 2, *Interoperability & Standards*. <<http://ec.europa.eu/digital-agenda/en>> (ultima consultazione 16/12/2014).

¹⁶<<http://www.europeana-libraries.eu/web/guest/home>> (ultima consultazione 16/12/2014).

Standardizzazione (ISO)¹⁷ o provenienti da altrettanti fonti autorevoli; l'approfondimento del suo valore in relazione ai sistemi di organizzazione e gestione della conoscenza; lo studio dell'interoperabilità nel settore delle costruzioni, campo di indagine prescelto.

1. Analisi del concetto di interoperabilità

L'interoperabilità è un presupposto fondamentale per lo sviluppo di specifiche e sistemi in grado di favorire lo scambio e la condivisione di dati e informazioni. Il suo raggiungimento pieno, prevede la predisposizione di standard e prassi che superino la visione circoscritta ai soli aspetti tecnici.

Per meglio comprendere il valore e le peculiarità proprie di questo concetto, apparentemente semplice ma che in realtà è intriso di potenzialità, si analizzano di seguito talune definizioni differenti per intensione¹⁸ ed estensione.¹⁹

La ISO/IEC 2382-1:1993 - *Information technology - Vocabulary - Part 1: Fundamental terms*²⁰, definisce l'interoperabilità come:

“the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units”,

ovvero la capacità di programmi o più generalmente di sistemi (*functional units*) di comunicare, eseguire o trasferire dati attraverso l'uso di modalità, tali da non richiedere all'utente l'obbligatoria conoscenza di queste unità funzionali. Tale concettualizzazione accentua il carattere bivalente dell'interoperabilità, data cioè dalla combinazione di più aspetti (dimensione tecnica e dimensione semantica) che di volta in volta devono essere valutati. L'enfasi è sull'utente e sulla sua capacità di saper interpretare e decodificare dati prodotti da altri, grazie alla descrizione del contesto e delle relazioni che li caratterizzano. Ne consegue quindi, che l'interoperabilità non possa esaurirsi nella mera trasmissione di dati e/o informazioni ma si concretizza più ampiamente nella loro condivisa interpretazione.

¹⁷ <<http://www.iso.org/iso/home.html>> (ultima consultazione 16/12/2014).

¹⁸ Una definizione si dice intensionale quando partendo dal concetto immediatamente superiore si specializza attraverso l'indicazione dell'insieme delle caratteristiche che lo denotano e lo distinguono da altri concetti ad esso correlati. «cfr.» HELMUT RIEDIGER, *Cos'è la terminologia e come si fa un glossario*, 2010, <http://www.term-minator.it/corso/doc/mod3_termino_glossa.pdf> (ultima consultazione 16/12/2014).

¹⁹ Una definizione si dice estensionale quando il concetto viene designato e descritto attraverso la presentazione degli elementi singoli che lo caratterizzano. *Ibidem*.

²⁰ <<https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-1:ed-3:v1:en>> (ultima consultazione 16/12/2014).

Questa consapevolezza è ripresa da (Gnoli, 2014),²¹ il quale si fa promotore della visione secondo la quale è importante che i collegamenti siano definiti non soltanto tra dati (oggetti concreti e/o astratti) ma anche tra i contenuti, al fine di non vanificare gli sforzi che i provider di informazioni (biblioteche, musei, centri di documentazione, editori, ecc.) hanno compiuto fino a questo momento.

Il connubio tra la dimensione tecnologica e semantica dell'interoperabilità secondo lo stesso Gnoli è ormai divenuto "un paradigma delle scienze dell'informazione" che si manifesta attraverso la specificazione di correlazioni di natura semantica. Per cui non sarà più sufficiente dire che tra un oggetto A e un oggetto B esiste semplicemente un legame, ma la natura del legame che determina l'implicazione tra i due oggetti, dovrà essere esplicitamente rappresentata. Questa concezione non fa altro che tenere conto dell'evoluzione tecnologica e culturale che sta investendo la rete e i suoi utenti.

La seconda definizione, ripresa dalla ISO 25964-2:2013 – *Information and documentation – Thesauri and interoperability with other vocabularies – Part 2 – Interoperability with other vocabularies*, alla voce interoperabilità riporta:

"ability of two or more systems or components to exchange information and to use information that has been exchanged".

La sinteticità di questa definizione presenta in maniera trasparente le finalità e gli elementi essenziali dell'interoperabilità: garantire lo scambio e il riuso di informazioni contestualizzate mediante la messa a punto di soluzioni tecnologiche in grado di assicurare il corretto trasferimento di dati e informazioni. Inoltre, considerate le finalità della norma orientate cioè alla riduzione dell'ambiguità terminologica e semantica e alla valorizzazione di collegamenti tra risorse differenti è implicitamente desumibile il diretto riferimento alla necessità di definire linguaggi di rappresentazione delle informazioni precisi e non ambigui.

Similitudini con le due visioni precedenti si possono riscontrare in quanto proposto dall'*Institute of Electrical and Electronics Engineers* (IEEE),²² il quale si riferisce all'interoperabilità come:

²¹ «cfr.» CLAUDIO GNOLI, *Connettere la conoscenza: quattro livelli di interoperabilità*, «Biblioteche oggi», vol. 32, n. 5, 2014, p. 9-16, <http://biblioteche.provincia.bergamo.it/portali/c/document_library/get_file?p_l_id=22502&folderId=271699&name=DLFE-9361.pdf> (ultima consultazione 16/12/2014).

²² <http://www.ieee.org/index.html?WT.mc_id=hpf_logo> (ultima consultazione 16/12/2014).

*“ability of a system or a product to work with other systems or products without special effort on the part of the customer. Interoperability is made possible by the implementation of standards”.*²³

Questa definizione apparentemente circoscritta ai soli aspetti tecnici, fa esplicito riferimento al ruolo dell'utente e alla necessità di garantire un accesso semplificato all'informazione. Essa infatti, nell'ultima parte, promuovendo la predisposizione di standard di riferimento comuni e condivisi, concernenti l'interoperabilità e l'insieme degli aspetti ad essa correlati, introduce quello che può essere inteso come il suo elemento discriminante,

La visione proposta dell'IEEE, emerge dalle linee guida presentate nell'ambito dell'*European Interoperability Framework for pan-European egovernment services*,²⁴ di supporto all'interazione di amministrazioni pubbliche, imprese e cittadini. Nello specifico, per interoperabilità si intende:

*“the ability of information and communication technology (ICT) systems and of the business processes they support to exchange data and to enable the sharing of information and knowledge.”*²⁵

Quest'ultima, così come quanto esplicitato dall'IEEE, intende l'interoperabilità a livello di sistema (dimensione tecnica) tralasciando il livello di astrazione dato dalla semantica.

Il raffronto tra le definizioni summenzionate, mostra come in prima battuta, la connotazione tecnologica dell'interoperabilità sia privilegiata rispetto alla sua controparte semantica, pur essendo quest'ultima divenuta strategica nel campo dell'Information and Communication Technology (ICT) per la realizzazione di sistemi volti all'organizzazione e al trattamento dell'informazione.

Tale propensione è probabilmente dovuta ad una consuetudine ormai diffusa, secondo la quale le potenzialità e le criticità scaturite dall'interoperabilità siano prima di tutto, una questione tecnica di sistema, e solo successivamente, una questione semantica, legata cioè alla condivisione e alla corretta interpretazione delle informazioni.

L'esigenza di armonizzare il più possibile i linguaggi e creare le condizioni per il corretto trasferimento delle informazioni, è tanto più stringente in ambienti digitali

²³ <http://www.ieee.org/education_careers/education/standards/standards_glossary.html> (ultima consultazione 16/12/2014).

²⁴ <<http://ec.europa.eu/idabc/en/document/2319/5644.html>> (ultima consultazione 16/12/2014).

²⁵ «cfr.» EUROPEAN COMMISSION, *European Interoperability Framework for pan-European egovernment services*, Belgium, 2004. ISBN 92-894-8389-X.

chiamati a gestire e a rendere disponibili una quantità e un'eterogeneità sempre più elevata di informazioni. In tali ambiti, il concetto di interoperabilità è strategico poiché le interazioni avvengono per mezzo di calcolatori elettronici ed è pertanto importante tenere conto delle varie sfaccettature del concetto, equilibrandone le dimensioni in funzione del contesto di applicazione e degli obiettivi da perseguire.

In un saggio di (Agosti e Ferro, 2010)²⁶ comparso sulla rivista *Digitalia*, l'interoperabilità è analizzata in relazione alle biblioteche digitali precisando come lo sviluppo di sistemi di gestione e organizzazione dell'informazione veramente interoperabili, richiedano una valutazione del concetto nella sua interezza.

Gli autori per sottolineare questa prospettiva riprendono l'approccio suggerito dal gruppo di lavoro *European Commission Working Group on Digital Library Interoperability*, in base al quale è possibile scomporre l'interoperabilità in sei distinte dimensioni, ciascuna delle quali mette in evidenza un elemento distintivo del concetto:

- *interoperating entities* (istituzioni cooperanti): istituzioni e/o enti che decidono di cooperare sinergicamente affinché il sistema biblioteca digitale sia gestito e mantenuto nel tempo;
- *information objects* (oggetti informativi): insieme delle risorse rappresentate per mezzo di metadati,²⁷ rese disponibili ad un'utenza sempre più vasta, variegata ed esigente e inquadrato in uno specifico contesto;
- *functional perspective* (prospettiva funzionale): complesso delle modalità e delle funzionalità da implementare grazie alle quali l'interoperabilità si concretizza;
- *multilinguality* (multilinguismo): dimensione che intende indirizzare i sistemi bibliotecari verso un'apertura tale da favorire l'incontro tra lingue diverse e valorizzare le lingue minori che convivono nella medesima cultura;
- *user perspective* (prospettiva dell'utente): dimensione che tiene conto della frammentarietà della tipologia di utenza a cui ci si riferisce analizzandone e valutandone l'orientamento di modo da incontrare l'esigenza informativa di volta in volta espressa;

²⁶ «cfr.» MARISTELLA AGOSTI, NICOLA FERRARI, , *Interoperabilità tra sistemi di biblioteche digitali*, «Digitalia», vol. 1, n. 1, 2010. ISSN: 1972-621X <<http://digitalia.sbn.it/article/view/258/169>> (ultima consultazione 16/12/2014).

²⁷ «Data that identify attributes of document typically used to support functions such a location, discovery [...]». ISO 25964-1:2011 *Information and documentation – Thesauri and interoperability with other vocabularies - Part1: Thesauri for information retrieval*, p.6.

- *interoperability technology* (interoperabilità tecnologica): dimensione che si esplica attraverso l'uso e l'implementazione di soluzioni tecnologiche in grado di garantire lo scambio e la condivisione di dati e informazioni.

Il *European Commission Working Group on Digital Library Interoperability* mostra una prospettiva pragmatica dell'interoperabilità proprio attraverso la specificazione dei diversi elementi che è importante valutare affinché i sistemi di biblioteche digitali siano realmente interoperabili. Significativo è il livello in cui la dimensione tecnologica è collocata, ciò è dovuto non tanto al mancato riconoscimento del suo ruolo, quanto piuttosto alla volontà di voler sottolineare il valore rappresentato dalle dimensioni che la precedono.

Sulla base delle definizioni associate al concetto di interoperabilità, esposte in questo paragrafo, è possibile trarre talune considerazioni:

- l'interoperabilità assume valori diversi in funzione del contesto in cui viene calato;
- l'interoperabilità può avere una valenza sociale, in quanto promotrice del trasferimento di dati e informazioni secondo modalità aperte e condivise;
- l'interoperabilità necessita della predisposizione di strumenti e metodi che siano di supporto alla connessione dei sistemi informatici (interoperabilità tecnica) e all'armonizzazione e al riconoscimento di dati e informazioni attraverso l'esplicitazione delle mutue corrispondenze (interoperabilità semantica).

Quest'ultima osservazione, come sarà illustrato nel capitolo 2, costituisce il punto di partenza affinché l'idea di web semantico (*semantic web*) proposto per la prima volta da Tim Berners-Lee nel 2001²⁸ diventi tangibile.

2. L'interoperabilità nelle scienze dell'informazione

L'interoperabilità come anticipato nel paragrafo precedente, è ormai riconosciuta come il nuovo "paradigma delle scienze dell'informazione", essa deve garantire che i sistemi informatici e i sistemi concettuali per loro natura eterogenei siano messi in comunicazione di modo da facilitare l'accesso e il recupero delle informazioni.

²⁸ «cfr.» TIM BERNERS-LEE, et alii, *The Semantic Web*, «Scientific American», May 2001, p. 29-37 <<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>> (ultima consultazione 16/12/2014).

Come è noto, l'evoluzione tecnologica e la disponibilità di sistemi sempre più potenti ha dato un forte impulso alla produzione di dati e informazioni non sempre di qualità. Come si legge in (Guerrini, Possemato 2012)²⁹ qualità e quantità non sempre si equivalgono ma molto più verosimilmente si tratta di variabili inversamente proporzionali; ne scaturisce che non sempre disporre di una grande collezione di dati e/o informazioni voglia dire avere la capacità di rispondere in maniera efficace ed efficiente ad una richiesta informativa. La rete ne è un esempio tangibile, tanti sono i dati e le informazioni che essa veicola ma è importante valutare quanti di questi siano realmente utili, disponibili e riutilizzabili.

I provider di dati e informazioni, infatti, se fino a poco tempo fa erano poco avvezzi a mettere a disposizione della collettività il loro patrimonio informativo, oggi avvertono non soltanto la necessità di renderlo disponibile ma di farlo in modo intelligente, ovvero realizzando archivi informativi strutturati, il cui contenuto sia codificato per mezzo di linguaggi di marcatura sempre più formali ed espressivi.

L'interconnessione di repository informativi distinti rappresenta una delle grandi sfide lanciate dalla rete e a cui la comunità internazionale riconosciuta sotto il nome di *World Wide Web Consortium (W3C)*³⁰ sta rispondendo sviluppando standard aperti e tecnologie sempre più innovative. Il W3C infatti, è un consorzio che si propone la realizzazione di un *Web of all*³¹ in cui tutti gli utenti possano comunicare e condividere informazioni al fine di generare conoscenza, e di un *Web on Everything*³² dove di fianco a calcolatori elettronici vi sono altri dispositivi (*devices*) quali cellulari e altri sistemi interattivi.

Il principio della connessione che è alla base dell'attuale visione del web non è tuttavia da imputare interamente alla rete. Si tratta piuttosto, della rivisitazione in chiave moderna di alcuni dei principi fondamentali alla base dei più noti sistemi di classificazione: la Classificazione Decimale Dewey (CDD) e la Classificazione Decimale Universale (CDU). Tali sistemi di classificazione infatti, si sono preoccupati di procedere alla definizione di un ordinamento del sapere umano mediante l'identificazione di criteri di classificazione basati sulla somiglianza tra oggetti (concreti e astratti) e sulla loro disuguaglianza. La sistematizzazione che ne deriva non fa altro che dare evidenza dei legami intrinseci che sussistono tra concetti.

²⁹ «cfr.» MAURO GUERRINI, TIZIANA POSSEMATO, *Linked data: un nuovo alfabeto del web semantico*, «Biblioteche oggi», vol. 30, n.3, 2012, p. 7-15 <<http://www.bibliotecheoggi.it/content/201200300701.pdf>> (ultima consultazione 16/12/2014).

³⁰ <<http://www.w3.org/>> (ultima consultazione 16/12/2014).

³¹ <<http://www.w3.org/Consortium/mission.html>> (ultima consultazione 16/12/2014).

³² *Ibidem*.

La connessione quindi come riportato in (Gnoli, 2014),³³ diviene la *metafora della rete*, all'interno della quale ad essere connessi non sono unicamente dati bensì concetti. La rete si configura come un grafo costituito da nodi (i concetti) e da archi (le relazioni), non si connota per la sua strutturazione a priori ma per la scelta di criteri di classificazione effettuata sulla base delle motivazioni che portano alla sistematizzazione della conoscenza. La rete infatti, rispetto ai tradizionali modelli arborescenti di tipo gerarchico presenta il grande vantaggio di rendere combinabili un numero potenzialmente infinito di correlazioni semantiche.

Il riconoscimento dell'interconnessione come pratica per supportare la diffusione del patrimonio informativo ha in un certo senso indotto le biblioteche, gli archivi e i musei, istituzioni che tradizionalmente si sono fatte carico di garantire la conservazione, l'organizzazione e la valorizzazione di questa immensa proprietà a rivedere consuetudini ormai consolidate nel tempo e ad allineare modelli e strumenti in un'ottica maggiormente interoperabile.

Queste istituzioni più di altre, stanno vivendo un ulteriore rinnovamento come d'altro canto è avvenuto con il passaggio dal catalogo cartaceo al catalogo elettronico. Così come il superamento del primo in favore del secondo, non è stato un mero cambiamento di supporto ma piuttosto un ampliamento delle funzionalità del catalogo,³⁴ così anche il superamento di alcuni paradigmi a favore dell'accrescimento della condivisione di risorse informative, attraverso l'implementazione di nuove applicazioni informatiche e l'attivazione di nuove procedure, porterà numerosi vantaggi. Naturalmente i rischi in cui è possibile incorrere sono molteplici e vanno dalla diffusione di dati e informazioni inconsistenti, alla tutela della proprietà intellettuale. Tali incognite tuttavia, non devono destare eccessiva preoccupazione né tanto meno procurare inutili allarmismi dovuti alla creazione di una possibile situazione di disordine.³⁵ Piuttosto, è importante che le stesse istituzioni si interrogino sull'adeguatezza dei criteri e delle procedure di descrizione utilizzati e sulla natura degli standard bibliografici adottati.

Una comunità sempre più densa di ricercatori, studiosi e professionisti del settore sta lavorando a pieno ritmo, per far sì che la mole di informazioni conservata nelle biblioteche, negli archivi e nei musei sia resa disponibile ad una utenza sempre più variegata e ampia e sia riconosciuto loro il ruolo di promotori del sapere. L'obiettivo

³³ «cfr.» CLAUDIO GNOLI, *Op. cit.*

³⁴ «cfr.» PAUL GABRIELE WESTON, *Il catalogo elettronico. Dalla biblioteca cartacea alla biblioteca digitale*, (Collana: Beni culturali, 26), Roma, Carocci editore, 2002. ISBN 88-430-2174-5.

³⁵ «cfr.» DAVID WEIMBARGER, *Elogio del disordine: le nuove regole del mondo digitale*, Milano, BUR Rizzoli, 2010. ISBN 978-88-17-03911-6.

consiste pertanto, nella messa in pratica delle cinque leggi della biblioteconomia³⁶ che Shiyali Ramamrita Ranganathan, matematico e bibliotecario indiano ha lasciato in eredità, ovvero:

1. i libri sono per l'uso: il libro è la manifestazione del pensiero intellettuale di chi lo concepisce e lo crea. Questa legge sottolinea il rapporto tra libro e lettore e puntualizza che il valore di un libro aumenta in funzione dell'uso che ne viene fatto;
2. a ogni lettore il suo libro: ciascun lettore deve essere considerato nella sua individualità e la scelta dei libri da leggere è una questione che attiene al singolo. La biblioteca quindi deve essere in grado di tener conto di questa varietà di utenza guidando ciascuno verso il riconoscimento del libro di cui ha bisogno;
3. a ogni libro il suo lettore: i libri sono oggetti inerti, da soli non sono in grado di incontrare il lettore, necessitano pertanto della presenza di qualcuno che faccia da tramite e favorisca questo incontro ovvero il bibliotecario che funge da referente tra il libro e il lettore;
4. risparmia il tempo del lettore: questa legge inserisce rispetto a quelle precedenti la variabile tempo, sottolineando che la riduzione del tempo dipende dalla familiarità del bibliotecario con la collezione di libri e dalla capacità di comprendere l'interesse informativo dell'utente;³⁷
5. la biblioteca è un organismo in crescita: questa legge riporta l'attenzione sulla biblioteca come luogo che ospita un numero sempre maggiore di libri (crescita quantitativa) e che necessita dell'aggiornamento del personale, delle pratiche adottate e degli strumenti utilizzati (crescita qualitativa).

Questi principi che sintetizzano il pensiero di Ranganathan conferiscono centralità all'utente, poiché come sottolineato da (Gnoli, 2000),³⁸ riprendendo le parole dello stesso bibliotecario: *“qualsiasi operazione che avviene in biblioteca deve essere condotta tenendo sempre presente l'utilità che in ultima analisi avrà per le persone.”*

³⁶ «cfr.» SHIYALI RAMAMRITA RANGANATHAN, *Le cinque leggi della biblioteconomia*, a cura di Laura Toti, Firenze, Le Lettere, 2010 (Collana: Pinakes, 6). ISBN 9788860873309.
Edizione originale: SHIYALI RAMAMRITA RANGANATHAN, *The Five Laws of Library Science*, Sarada Ranganathan Endowment for Library Science, 1957.

³⁷ «cfr.» CARLO BIANCHINI, *Dalla professione all'etica. Il servizio di reference come esperienza esistenziale*, in *Il servizio di reference*, a cura di Carlo Bianchini, Firenze, Le Lettere, p. XIX-XLII. ISBN 9788860873095
<<http://eprints.rclis.org/16294/1/Dalla%20professione%20all%20etica%20Prefaz%20a%20Il%20servizio%20di%20reference.pdf>> (ultima consultazione 16/12/2014).

³⁸ «cfr.» CLAUDIO GNOLI, *Il tavolino di Ranganathan*, «Bibliotime», a. III, n. 3, novembre 2000, <<http://www.aib.it/aib/sezioni/emr/bibtime/num-iii-3/gnoli.htm#5leggi>> (ultima consultazione 16/12/2014).

In quest'ottica di rinnovato cambiamento, le istituzioni della memoria registrata sono consapevoli del ruolo che sono chiamate a svolgere e pertanto, stanno procedendo all'aggiornamento delle proprie strutture di metadati descrittive e all'allineamento degli strumenti impiegati per l'indicizzazione dei documenti.³⁹ L'obiettivo è di semplificare l'accesso alle informazioni e migliorare la qualità del servizio reso agli utenti.⁴⁰

In sostanza, le biblioteche, gli archivi e i musei devono fare in modo che il loro patrimonio informativo diventi parte integrante del web e sia reso navigabile grazie all'adozione dei principi del web semantico, il quale spostando il baricentro dal documento al dato, si propone l'adozione di standard per la rappresentazione precisa e non ambigua dei dati in un modo tale da poter essere direttamente interpretabili e processabili da una macchina.

Il dato costituisce quindi, un elemento atomico il cui valore è veicolato dalla natura delle relazioni esplicitate tra dati. Questa dimensione che rappresenta una delle peculiarità dell'approccio linked data, che sarà ampiamente trattato nel capitolo 2 è ormai largamente diffusa e condivisa ragion per cui, è importante che i metadati, e soprattutto il loro significato siano espressi in maniera precisa e non ambigua secondo i formalismi raccomandati.

L'adesione a questo approccio permetterebbe alle biblioteche di:

- collegare e integrare i propri dati con quelli prodotti da altre istituzioni anche se espressi con standard differenti, assicurando inoltre, la collaborazione con provider diversi attraverso la promozione di una politica di sistema;
- collaborare sinergicamente con la Pubblica Amministrazione in considerazione delle caratteristiche che contraddistinguono entrambe le istituzioni: le biblioteche in quanto generatori di dati qualitativamente validi e garantiti; e la Pubblica Amministrazione in quanto produttore di dati che riguardano la collettività;
- favorire l'incontro tra un bisogno informativo e la collezione documentale che si sta interrogando, attraverso l'uso di un linguaggio che sia il più possibile vicino a quello conosciuto dall'utente, ovvero garantire l'interoperabilità sul

³⁹ *“L'operazione mediante la quale si creano gli accessi al contenuto semantico del documento. Consta delle fasi di analisi concettuale e di traduzione dei concetti individuati e delle relazioni logiche individuate nei termini e nelle forme proprie del linguaggio di indicizzazione prescelto.”* «cfr.» ALBERTO CHETI, *Manuale ipertestuale di analisi concettuale*, 1996, <http://biblioteche.unibo.it/manuals/html_1/HOME.HTML> (ultima consultazione 16/12/2014).

⁴⁰ CARLO BIANCHINI, MAURO GUERRINI, *Introduzione a RDA*, Milano, Editrice Bibliografica, 2014, (Collana: Biblioteconomia e scienza dell'informazione, 3). ISBN/EAN 9788870757804.

piano semantico tra il linguaggio comune e linguaggio controllato e standardizzato dei sistemi impiegati per l'organizzazione delle informazioni;⁴¹

- identificare univocamente i dati bibliografici rendendoli tecnicamente compatibili con le tecnologie del web semantico.

In ambito bibliotecario l'*International Federation of Library Associations and Institutions* (IFLA)⁴² e altri gruppi di lavoro quali il già menzionato *European Commission Working Group on Digital Library Interoperability*, l'*ISBD Review Group*⁴³ e altri, hanno avviato il processo di revisione e pubblicazione del proprio set di metadati in modalità linked data, al fine di favorire l'interoperabilità e la portabilità dei dati bibliografici rendendoli compatibili con le nuove tecnologie proposte dal web semantico.

Nello specifico, queste comunità si sono impegnate affinché i loro set di dati siano trasferiti in formato *Resource Description Framework* (RDF)⁴⁴ che, data la sua elevata espressività, rappresenta lo standard *de facto* del web semantico in quanto facilita lo scambio e il riutilizzo dei dati strutturati in un formato *machine-understandable* garantendo contemporaneamente l'interoperabilità tra applicazioni diverse.

Gli elementi coinvolti in questo processo di allineamento rispetto alle priorità previste dal web semantico sono i metadati, ai quali è riconosciuto non solo un valore funzionale ma anche un valore semantico. I metadati letteralmente dati sui dati agevolano e semplificano il processo di recupero dell'informazione, ottimizzano i tempi computazionali e migliorano il grado di precisione della ricerca. I metadati veicolano informazioni inerenti sia alla struttura di un documento, ad esempio titolo, autore, editore, data di pubblicazione, ecc., sia al suo contenuto ad esempio descrizione, parole chiave o *key words*,,, sono dotati di un alto livello di espressività e capaci di esplicitare e rendere accessibile l'informazione implicitamente contenuta nei documenti.

Il processo di conversione ha interessato i più importanti standard bibliografici tra cui si ricordano a titolo esemplificativo: l'*International Standard Bibliographic Description* (ISBD),⁴⁵ il *Resource Description and Access* (RDA),⁴⁶ il *Dublin Core Metadata Element*

⁴¹ «cfr.» ELENA ESCOLANO RODRIGUEZ, *L'adattamento di ISBD al web semantico dei dati bibliografici espressi in linked data*, «JLIS.it 3.2», DOI: 10.4403/jlis.it-5484, 2012.

⁴² <<http://www.ifla.org/>> (ultima consultazione 16/12/2014).

⁴³ <<http://www.ifla.org/isbd-rg>> (ultima consultazione 16/12/2014).

⁴⁴ <<http://www.w3.org/RDF/>> (ultima consultazione 16/12/2014).

⁴⁵ ISBD è uno standard per la descrizione di risorse documentali. Esso prevede l'uso dei segni di punteggiatura da adottare convenzionalmente per agevolare l'accesso alle risorse e fornendo dei criteri di descrizione univoci e condivisi. «cfr.» IFLA, *International Standard Bibliographic Description*, Edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, 2012,

Set (DCMI).⁴⁷ Inoltre tale procedura ha riguardato anche la parte più propriamente semantica relativa ai sistemi di organizzazione e gestione della conoscenza, che ha riguardato la definizione di un *framework* di modo da renderli compatibili con il web semantico e favorire la realizzazione di una rete semantica virtuale grazie all'approccio linked data, presentato nel Capitolo 2, paragrafo 2.

3. I sistemi di organizzazione della conoscenza come strumenti di supporto dell'interoperabilità semantica

I sistemi di organizzazione della conoscenza, dall'inglese *Knowledge Organization Systems* (KOS) sono strumenti semantici estremamente efficaci nella rappresentazione e gestione delle informazioni. Il termine KOS largamente diffuso anche in italiano, è stato coniato dal *Networked Knowledge Organization Systems Working Group* (NKOS) e designa l'insieme degli strumenti (terminologie, sistemi di classificazione, thesauri, reti semantiche, ontologie, ecc.), che si preoccupano di sistematizzare e esplicitare la struttura semantica di un dato dominio (o più genericamente del sapere) e supportano l'organizzazione, la gestione e il recupero delle informazioni attraverso la rete.⁴⁸

I KOS, nati in contesto bibliotecario e impiegati tradizionalmente nella catalogazione e collocazione sugli scaffali delle risorse documentali, hanno dimostrato di poter essere impiegati anche dai motori di ricerca come ad esempio stanno facendo Google, Bing, Yahoo!, per ottimizzare l'efficacia dei loro sistemi di *searching*, aumentare il grado di precisione (o *Precision*)⁴⁹ e di richiamo (o *Recall*)⁵⁰ e ridurre conseguentemente il rumore dovuto al recupero di documenti non rilevanti. I KOS infatti, insieme alle

<http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2012/ISBD_NOV2012_online.pdf> (ultima consultazione 16/12/2014).

⁴⁶ RDA è un insieme di linee guida per la catalogazione di risorse documentali particolarmente incentrato sull'accesso e il recupero delle stesse da parte degli utenti. Esse sono state elaborate da *Joint Steering Committee Joint Steering Committee (JSC)* e subentrano alle *Anglo-American Cataloguing Rules, 2nd (ACCR2)*, <<http://www.rda-jsc.org/docs/rdabrochureJanuary2010.pdf>>, (ultima consultazione 16/12/2014).

⁴⁷ DC è un sistema di metadati impiegato per la descrizione di risorse documentali e prevede 15 elementi essenziali e volutamente generici di modo da poter includere le varieghe tipologie di risorse documentali. <<http://dublincore.org/documents/dces/>> (ultima consultazione 16/12/2014).

⁴⁸ <<http://nkos.slis.kent.edu/>> (ultima consultazione 16/12/2014).

⁴⁹ "Rapporto fra il numero di documenti trovati con una ricerca giudicati rilevanti e il numero complessivo dei documenti trovati". «cfr.» CLAUDIO GNOLI, *Glossario di organizzazione della conoscenza*, 2011, <<http://www.iskoi.org/doc/glossario.htm>> (ultima consultazione 16/12/2014).

⁵⁰ "La proporzione di documenti rilevanti noti nella collezione che sono stati reperiti dalla configurazione di sistema in risposta ad un'interrogazione". MASSIMO MELUCCI, *Information retrieval. Metodi e modelli per i motori di ricerca*, Franco Angeli Editore, 2013 (Collana: Informatica), p.123. ISBN: 8820421127. ISBN-13: 9788820421120.

strategie di *Information Retrieval* (IR)⁵¹ garantiscono il colloquio tra l'esigenza informativa espressa da un utente e la collezione che si sta interrogando.

La loro riscoperta in contesti digitali, avvenuta in concomitanza dell'affermarsi delle tecnologie del web semantico, li rende strumenti molto efficaci in *termini di ordinamento, organizzazione e recupero dell'informazione*.⁵² I KOS supportano le seguenti operazioni:

- Indicizzazione dei documenti

L'indicizzazione è il processo mediante il quale il contenuto concettuale di un documento viene espresso per mezzo di una o più voci di soggetto, che rappresentano il tema di base del documento, a sua volta specificato tramite l'ausilio di un linguaggio di indicizzazione.⁵³ I linguaggi di indicizzazione svolgono una funzione di controllo terminologico molto forte al fine di limitare l'ambiguità semantica propria del linguaggio naturale. Questo controllo è esercitato attraverso: 1) la gestione della polisemia (lo stesso termine può assumere significati diversi); 2) la gestione della sinonimia (più termini possono designare il medesimo significato); 3) la specificazione del contesto d'uso del termine.

Taluni KOS maggiormente strutturati - ad esempio i thesauri - delimitano i rischi di ambiguità organizzando i termini orizzontalmente (relazione sinonimica) e verticalmente (relazione gerarchica) all'interno di una rete semantica e esplicitando la natura delle relazioni che li caratterizza, costruendo quindi un vero e proprio percorso che accompagna l'utente verso le scelte terminologiche preferite dall'indicizzatore.⁵⁴

- Classificazione dei documenti

In taluni KOS maggiormente diffusi in ambito bibliotecario come i sistemi di classificazione, il soggetto di una risorsa documentale non viene espresso per mezzo di un enunciato esplicativo (un termine) ma da una notazione numerica

⁵¹ Le tecniche di IR prevedono che a fronte di una richiesta formulata (query) dall'utente in linguaggio naturale sia attivata una procedura di conversione in cui la forma della domanda (sintassi) e il suo significato (semantica) siano trasferiti in un linguaggio formale, preciso e non ambiguo e quindi pronta per essere sottoposta al processo automatico di recupero dell'informazione.

⁵² ANTONIETTA FOLINO, *Tassonomie e thesauri*, in *Documenti Digitali*, a cura di Roberto Guarasci, Antonietta Folino, Milano, ITER, 2013, p. 390, p. 387-444. ISBN 9788890341939.

⁵³ «cfr.» ALBERTO CHETI, *Manuale ipertestuale di analisi concettuale*, 1996, <http://biblioteche.unibo.it/manuals/html_1/HOME.HTML> (ultima consultazione 16/12/2014).

⁵⁴ DOUGLAS TUDHOPE, MARIANNE LYKKE NIELSEN, *Introduction to knowledge organization systems and services*, «New Review of Hypermedia and multimedia», vol. 12, n. 1, June 2006, p.4. DOI: 10.1080/13614560600856433.

o alfanumerica che designa in maniera univoca e precisa una certa classe di apparenza. Questa tipologia di KOS concepita per le classificazioni bibliografiche, si fonda su una logica di scomposizione che va dal generale al particolare, è basata sul criterio della *garanzia letteraria* (literary warrant)⁵⁵ e muta in funzione dell'evoluzione della conoscenza. In aggiunta tali sistemi trovano un ampio utilizzo a supporto dei processi di classificazione semi-automatica dei documenti e di clustering.⁵⁶

- Metadatazione

I sistemi di organizzazione della conoscenza possono essere impiegati a supporto della descrizione delle risorse documentali. Essi infatti, come già detto sono costituiti da un set di dati che definiscono una serie di proprietà formali ascrivibili ad una certa risorsa: il titolo, l'autore, l'editore, ecc.. Accanto a questi ve ne possono essere degli altri che forniscono informazioni inerenti al contenuto di una risorsa; a titolo esemplificativo si ricorda il *tag Subject* previsto dal DCMI, *Version 1.1*, che, come raccomandato dallo stesso standard, dovrebbe essere espresso da una parola chiave o da una notazione estratta - se possibile - da un vocabolario controllato. Questa prospettiva orientata alla semantica rende l'attività di metadatazione particolarmente significativa in fase di *tagging* delle pagine web. Il processo di *tagging* o annotazione è un concetto proveniente dalla Linguistica dei corpora (LC) e consiste nell'associare a un dato testuale un *tag* per l'appunto, che esprime la funzione svolta dal dato all'interno del testo.

- Recupero dell'informazione (*browsing* e *searching*)

Sul web gli utenti hanno la possibilità di recuperare informazioni ricorrendo all'uso di due strategie: il *browsing* e il *searching*. Il primo prevede che il recupero dell'informazione di interesse avvenga tramite un processo di navigazione sequenziale delle pagine messo in atto dalla presenza di collegamenti ipertestuali (*hyperlink*) in cui l'utente viene guidato verso l'uso delle parole chiave più appropriate; il secondo invece prevede la specificazione di una parola chiave all'interno di un motore di ricerca, che restituisce un

⁵⁵ "Justification for the development of a class or the explicit inclusion of a topic in the schedules, tables, or Relative Index, based on the existence of a body of literature on the topic.".
<<http://www.oclc.org/content/dam/oclc/webdewey/help/glossary.pdf>> (ultima consultazione 16/12/2014).

⁵⁶ "Clustering is the most common form of unsupervised learning. No super vision means that there is no human expert who has assigned documents to classes. In clustering, it is the distribution and makeup of the data that will determine cluster membership". CHRISTOPHER D. MANNING, et alii, *Introduction to Information Retrieval*, Cambridge, Cambridge University Press, 2007, p.349. ISBN-10: 0521865719. ISBN-13: 978-0521865715.

insieme di documenti organizzati secondo criteri di *page ranking*.⁵⁷ Nonostante le due metodologie siano ampiamente utilizzate, esse presentano alcuni limiti dovuti alla difficoltà di reperire informazioni efficaci e pertinenti rispetto alla richiesta specificata, ragion per cui i KOS possono ricoprire un importante ruolo se integrati in sistemi informatici più ampi ottimizzando l'accuratezza dei risultati, aggiungendo semantica all'insieme delle regole di tipo sintattico attualmente impiegate.⁵⁸

- Organizzazione della conoscenza di dominio

I KOS presentano un diverso livello di granularità che dipende essenzialmente dagli scopi per cui sono realizzati. Infatti, a fronte di KOS generali che si propongono la sistematizzazione del sapere, ve ne sono altri maggiormente specialistici che si riferiscono ad un determinato dominio della conoscenza. In quest'ottica l'uso dei sistemi di organizzazione della conoscenza viene esteso, in quanto non sono più considerati esclusivamente ai fini dell'indicizzazione e della classificazione di risorse documentali, ma come supporto all'organizzazione e alla strutturazione di un dato ambito del sapere. In tal senso i KOS si propongono di: 1) ridurre l'ambiguità semantica alimentata dai fenomeni di sinonimia e polisemia, 2) limitare i casi di *misunderstanding*; 3) agevolare gli scambi comunicativi all'interno di una comunità di pratiche; 4) orientare gli utenti verso un uso maggiormente condiviso della terminologia di settore mettendo a disposizione adeguati strumenti di riferimento; 5) favorire l'accesso all'informazione di dominio; 6) supportare l'interoperabilità semantica e il riutilizzo delle informazioni.

È evidente quindi come la motivazione alla base dell'utilizzo dei KOS influenzi la struttura e il livello di granularità con cui si procede alla rappresentazione della conoscenza.

L'immagine che segue riprende il sistema ad assi cartesiani e mette in evidenza come il livello di complessità e sofisticatezza dei KOS aumenti progressivamente via via che la variabile tende verso il linguaggio controllato. Ed è proprio in funzione di tali peculiarità che i KOS sono generalmente raggruppati in tre categorie: 1) terminologie

⁵⁷ Il *page ranking* o semplicemente *rank*, è un l'ordine dei documenti restituiti a seguito di un'interrogazione che riflette il grado di rilevanza dei documenti restituiti rispetto all'interrogazione.

⁵⁸ PAOLO FERRAGGINA, *Sui motori di ricerca*, in *Documenti Digitali*, a cura di Roberto Guarasci, Antonietta Folino, Milano, ITER, 2013, . p. 362, p. 331-370. ISBN 9788890341939.

intese come liste di termini talvolta accompagnati da una definizione; 2) sistemi di classificazione e categorizzazione; 3) liste strutturate di concetti.⁵⁹

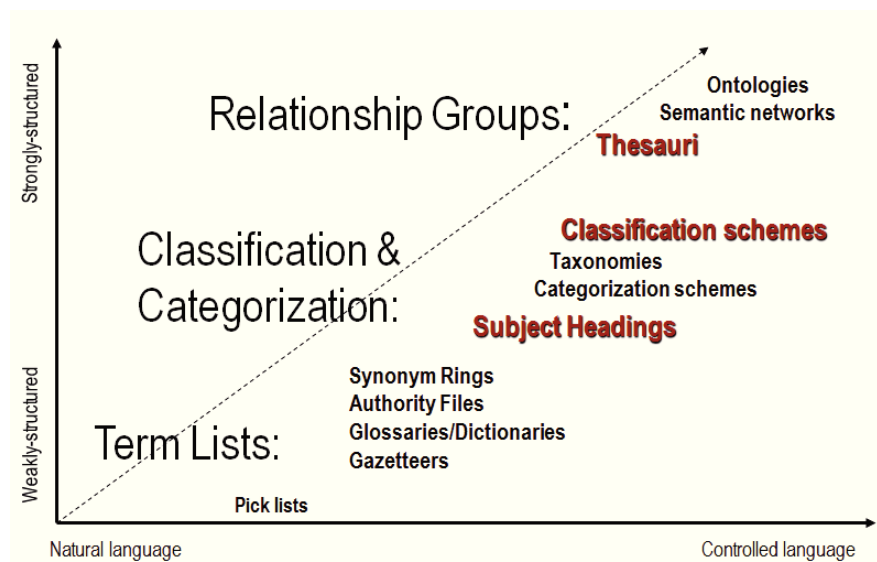


Figura 1. Tipologie di Knowledge Organization System.⁶⁰

Per meglio comprendere le similitudini e le differenze esistenti tra i diversi KOS si riporta una breve descrizione delle tre tipologie summenzionate:

- terminologie: rientrano in questa categoria liste di termini normalizzati - solitamente non strutturate - e accompagnate da una definizione. Nello specifico ci si riferisce a:
 - lessici specialistici: liste di termini inerenti a un ben definito dominio e *adottati in maniera convenzionale e condivisa*⁶¹ dagli individui che vi operano;
 - glossari: liste di termini, circoscritti a un preciso settore della conoscenza, accompagnati da definizioni;
 - authority file (archivio di autorità): liste di termini finalizzate al controllo terminologico di varianti di nomi di persone, organizzazioni, ecc.,

⁵⁹ GAIL HODGE, *Systems of knowledge organization for digital libraries: Beyond traditional authority files*, Washington, DC, The Digital Library Federation Council on Library and Information Resources, 2000 p. 5. ISBN 1-887334-76-9, <<http://www.clir.org/pubs/reports/pub91/pub91.pdf>> (ultima consultazione 16/12/2014).

⁶⁰ «cfr.» MARCIA LEI ZENG, ATHENA SALABA, *Toward an International Sharing and Use of Subject Authority Data*, FRBR Workshop, OCLC, 2005.

⁶¹ HELMUT REDIGUER, *Op. cit.*, p. 6.

impiegate principalmente per garantire l'accesso univoco e condiviso alle informazioni presenti in un catalogo,⁶²

- dizionari: lista ordinata di termini e definizioni, si differenzia dal glossario per gli scopi maggiormente generali che persegue e perché fornisce informazioni circa il significato che un dato termine assume in diversi ambiti.
- Sistemi di classificazione e categorizzazione: i sistemi di classificazione si propongono la razionalizzazione della realtà secondo un principio di scomposizione, in cui la complessità disgiunta in entità più piccole (più maneggevoli) è collocata all'interno di uno schema concettuale preciso. Essi sono impiegati da biblioteche, archivi, musei, centri di documentazione, ecc. la cui filosofia e mission ruota intorno all'ordinamento e all'organizzazione del sapere, mediante l'analisi di contenuti e dati informativi propri di libri e oggetti documentali. I sistemi di classificazione maggiormente noti sono:
 - la Classificazione Decimale Dewey (CDD) elaborata da Mevil Dewey per la biblioteca dell'Amherst College nel 1876, prevede la suddivisione dello scibile in dieci classi, organizzate in dieci divisioni a loro volta suddivise in altrettante dieci sezioni;
 - la Classificazione Decimale Universale (CDU) concepita da Paul Otlet e Henry La Fontaine, pubblicata nel 1905 riprende i principi della CDD e si propone a differenza di questa, l'organizzazione piuttosto che la collocazione del materiale documentale.

La CDD e la CDU sono sistemi enumerativi, vale a dire elencano a priori l'insieme delle voci che è possibile impiegare in sede di indicizzazione. Nonostante, i due sistemi siano largamente impiegati nelle biblioteche a livello internazionale risultano essere rigidi e difficili da aggiornare e hanno una scarsa espressività.

- Liste strutturate di concetti: si tratta di sistemi di organizzazione della conoscenza in cui oltre ai concetti sono segnalate le relazioni semantiche che intercorrono tra di essi. Questa categoria include:

⁶² «cfr.» LAURA BONANNI, *L'Authority file*, ICCU, 2011, <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/sbn_notizie_1998_2001/pagina_185.html;jsessionid=1809496D6B6E95FE5B32ABF089034FDD> (ultima consultazione 16/12/2014).

- thesaurus: strumento per l'indicizzazione e per il recupero di risorse documentali che identifica e organizza sistematicamente i concetti e l'insieme delle relazioni che li governano.⁶³ Si differenzia dalle tassonomie⁶⁴ propriamente dette, poiché prevede in aggiunta alla relazione gerarchica anche quella di equivalenza e quella associativa. Il thesaurus deve essere sviluppato e mantenuto in conformità allo standard ISO 25964-1:2011 *Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval* e tenere quindi conto delle raccomandazioni formulate al suo interno. Il thesaurus è un valido strumento di supporto all'organizzazione e al recupero dell'informazione in quanto consente di tenere traccia, attraverso rinvii reciproci, dei sinonimi, dei quasi-sinonimi e delle varianti lessicali di un termine, e di specificare oltre a questa tipologia di associazioni anche relazioni gerarchiche e legami associativi. Questa peculiarità lo rende particolarmente adatto e utilizzabile in contesti digitali.
- Ontologia: strumento del web semantico impiegato per rappresentare, tramite l'ausilio di linguaggi altamente formali ed espressivi, un certo dominio della conoscenza. L'ontologia consente di specificare i concetti principali di un dominio e tipizzare le relazioni semantiche che esistono tra gli stessi tramite l'ausilio di linguaggi formali quali il succitato RDF e l'Ontology Web Language (OWL) di modo da rendere tale struttura comprensibile sia all'utente umano che alla macchina. Le ontologie sono largamente applicate nel campo dell'Intelligenza Artificiale (IA) e nella realizzazione di Decision Support System (DSS) in cui grazie alla caratterizzazione delle relazioni semantiche e alla specificazione di assiomi è possibile dedurre conoscenza a partire da elementi noti. Esempificazioni significative si trovano in ambito biomedico in cui combinando ad esempio la conoscenza relativa a sintomi, malattie e trattamenti farmacologici è possibile sviluppare sistemi di monitoraggio sull'efficacia delle cure.⁶⁵

⁶³ VANDA BROUGHTON, *Costruire thesauri: Strumenti per indicizzazione e metadati semantici*, edizione italiana a cura di Piero Cavalieri, traduzione di Laura Balestra e Luisa Venuti, Milano, Editrice Bibliografica, 2006, p.280. ISBN 978-88-7075-674-6.

⁶⁴ "Sistema di organizzazione della conoscenza formato da termini disposti in un albero gerarchico, generalmente privi di notazione; anche, disciplina che si occupa dei criteri secondo i quali conviene classificare gli oggetti di studio" <<http://www.iskoi.org/doc/glossario.htm>> (ultima consultazione 16/12/2014).

⁶⁵ <<http://www.w3.org/standards/semanticweb/ontology>> (ultima consultazione 16/12/2014).

Questa panoramica sui sistemi di organizzazione e gestione della conoscenza, oltre a rappresentare le peculiarità di ognuno e a mostrare le differenze strutturali esistenti tra di essi, intende mettere in evidenza il ruolo strategico che i KOS sono chiamati a ricoprire in contesti digitali a supporto della condivisione delle informazioni e dell'interoperabilità semantica.

4. Il concetto di interoperabilità nel settore delle costruzioni

Nei paragrafi precedenti l'interoperabilità è stata analizzata nella sua dimensione concettuale generale, sono state presentate talune definizioni e avanzate riflessioni in merito alla necessità di sviluppare applicazioni tecnologiche e condividere strategie linguistiche tali da favorire lo scambio e il riutilizzo di dati e informazioni, per evitarne la perdita o la mancata comprensione.

Anche in questo paragrafo dedicato al valore che l'interoperabilità assume in uno specifico dominio della conoscenza, ovvero quello delle costruzioni, in qualità di ambito di applicazione prescelto, si farà ricorso a definizioni proposte dalla comunità scientifica di settore, tenendo presente come all'interoperabilità di dati e informazioni e alla loro conseguente condivisione corrisponda un valore quantificato in termini economici stimato in base al fattore di *inadequate inetoperability*.⁶⁶

In questo settore infatti, l'interoperabilità è un'esigenza particolarmente avvertita poiché sottintende alla messa a punto di metodologie e infrastrutture tecnologiche⁶⁷ per gestire e supportare la mole di informazioni prodotte e utilizzate contestualmente da più operatori che hanno competenze diversificate. Si tratta nello specifico di ingegneri, architetti, geometri, imprenditori, capi cantiere, ecc. che quotidianamente prestano il loro lavoro e la loro esperienza alla realizzazione di un progetto edilizio, ma anche dei beneficiari finali.

La stratificazione delle competenze e delle conoscenze possedute da ciascuno di essi, influenza il livello di specializzazione degli scambi comunicativi e fa emergere come anche all'interno di un medesimo dominio lo stesso concetto possa essere espresso da una pluralità di termini o locuzioni. Ad esempio i due termini *finestra all'inglese* e

⁶⁶ «cfr.» GUILLERMO ARACNDA-MENA, RON WAKEFIELD, *Interoperability of building information: myth or reality?*, in *eWork and eBusiness, Engineering and Construction*, a cura di Manuel Martinez, Raimar Scherer, London, Taylor & Francis Group, 2006, p. 127-133, <<http://mams.rmit.edu.au/lrss7jid7nd4.pdf>> (ultima consultazione 16/12/2014).

⁶⁷ Nel settore delle costruzioni l'interoperabilità è legata alla messa a punto dell'approccio BIM (Building Information Modeling) che prevede di associare ad un disegno tecnico un apparato informativo che riporta solitamente: informazioni tipologiche; informazioni di tipo geometrico e dimensionale; informazioni prestazionali, ecc..

finestra a battente con apertura verso l'esterno non sono altro che varianti sinonimiche dello stesso concetto, con la sola differenza che la prima forma viene privilegiata in contesti d'uso comune e da individui che non presentano precise competenze nel settore; la seconda invece, di carattere normativo e attestata in fonti autorevoli, risulta essere preferita soprattutto nella redazione di documenti tecnici.

Il linguaggio delle costruzioni non è esente da fenomeni linguistici dovuti al mutamento semantico, al prestito linguistico, all'introduzione di neologismi in quanto si arricchisce e si evolve in funzione delle esigenze comunicative e dell'evoluzione tecnologica e normativa. Per cui anche se la maggior parte dei termini che vi afferiscono veicolano significati precisi e fortemente legati al contesto, trattandosi di un linguaggio specialistico, casi di ambiguità semantica e lessicale sono sempre possibili.

Questa constatazione è tanto più vera quando la terminologia di un certo settore entra a far parte o viene richiamata in altri contesti, come avviene per esempio per la redazione di linee guida, testi normativi nazionali e comunitari,⁶⁸ ecc., in cui più linguaggi settoriali si amalgamano rispondendo ai requisiti linguistici e concettuali di due o più settori.

È importante quindi, dotare le comunità di dominio di strumenti terminologici e semantici in grado di fornire una visione alquanto consolidata del linguaggio in uso e che contestualmente siano integrate e rese interoperabili rispetto a quelle risorse caratterizzate da una copertura semantica più generale, proprio per rispondere alle esigenze di un'utenza sempre più variegata.

Questo orientamento (specialistico vs generale) è strategico nel settore delle costruzioni, in ragione della frammentarietà del *workflow* dell'informazione relativo al ciclo di vita di un'opera, essendo quest'ultimo un percorso segmentato che si svolge lungo un arco temporale medio-lungo e che richiede che le informazioni inerenti ai prodotti e alle tecnologie costruttive impiegati siano espresse in un linguaggio controllato e unificato.

Tali presupposti sono alla base della missione dell'*International Alliance for Interoperability* (IAI)⁶⁹ che è nata con lo scopo di garantire lo scambio di dati tra

⁶⁸ MARIA TERESA ZANOLA, *La terminologia dei sistemi fotovoltaici tra comunicazione istituzionale e saperi specialistici. Un'introduzione*, in *Costruire un glossario: la terminologia dei sistemi fotovoltaici*, a cura di Maria Teresa Zanola, Milano, Vita&Pensiero, 2012, p. 7, p. 7-14. ISBN 9788834322970.

⁶⁹ "The IAI was started by twelve companies involved in the AEC/FM Industry. They were all interested in being able to work together 'interoperably'; that is, they wanted to be able to work with each other's information without being concerned about the software they were using or which anyone else was using.".
«Cfr.» INTERNATIONAL ALLIANCE FOR INTEROPERABILITY, *An Introduction to the International Alliance for Interoperability and the Industry Foundation Classes*, 1999,

applicazioni diverse mediante l'uso di un formato standard (interoperabilità tecnica) e la definizione di un linguaggio di riferimento (interoperabilità semantica). L'IAI predilige una visione multidisciplinare del dominio poiché tante ed eterogenee sono le competenze che interagiscono e si propone la definizione, la promozione e la pubblicazione di un framework che supporti la condivisione dei dati per tutto il ciclo di vita di un progetto edilizio tra i diversi attori e tra le diverse applicazioni.

Il punto di vista dell'IAI viene ripreso dal *National Institute of Standards and Technology (NIST)*⁷⁰ che definisce l'interoperabilità come:

“the ability to manage and communicate electronic product and project data between collaborating firms' and within individual companies' design, construction, maintenance and business process systems”,

ovvero la capacità di gestire e trasmettere in maniera elettronica dati relativi a prodotti e progetti tra imprese e tra i singoli che intervengono nel processo economico, di progettazione, di costruzione e di mantenimento.

La definizione privilegia una visione integrata dell'interoperabilità. Essa non intende l'interoperabilità solamente come una capacità di sistemi e/o software di comunicare e trasmettere dati, ma piuttosto interpreta il concetto come la capacità di un'organizzazione di condividere e scambiarsi dati in maniera precisa, sottolineando come i dati debbano essere condivisi a vari livelli (principio della multidisciplinarietà dell'IAI). Una definizione siffatta ha un forte valore simbolico, poiché non privilegia una dimensione del concetto piuttosto che un'altra, ma presenta il giusto equilibrio tra la componente tecnologica e quella semantica.

In (Yum, et alii, 2000)⁷¹ invece, l'interoperabilità è:

“the ability to exchange/share information between separate computer programs without any loss of content or meaning”,

cioè la capacità di scambiare e condividere informazioni tra programmi distinti senza perdita di contenuto o di informazione. La componente tecnologica è espressamente richiamata ma è altrettanto evidente come l'attenzione sia rivolta all'informazione piuttosto che al dato in sé ovvero al valore che il dato assume in funzione del contesto in cui si colloca.

<http://iaiweb.lbl.gov/Resources/IFC_Releases/IFC_Release_2.0/BETA_Docs_for_Review/IFC_R2_IntroTOIAI_Beta_d2.PDF> (ultima consultazione 16/12/2014).

⁷⁰ <<http://www.nist.gov/>> (ultima consultazione 16/12/2014).

⁷¹ KI HWAN YUM, ROBIN DROGEMULLER, *How Much Interoperability Can Be Achieved For the Construction Industry Today*, INCITE 2000 Conference, Implementing IT To Obtain a Competitive Advantage in the 21st Century, The Hong Kong Polytechnic University, Hong Kong, January 17.

Malgrado le due prospettive presentino delle differenze, esse non si escludono a vicenda ma più verosimilmente si completano. Per cui l'intento dell'interoperabilità nel settore delle costruzioni, così come in altri, consiste nel garantire la comunicazione e lo scambio di dati e informazioni tra i diversi soggetti che prendono parte alle diverse fasi di un processo edilizio, attraverso la predisposizione di infrastrutture tecnologiche e strumenti linguistici in grado di supportarne la condivisione.

L'attenzione verso l'interoperabilità in questo settore inizia nel 1999 quando il gruppo di lavoro che si occupava dell'attuale ISO/TC 59/SC 13 - *Organization of information about construction works*,⁷² convenne sulla necessità di procedere all'armonizzazione e alla classificazione della terminologia di settore secondo un modello partecipato in grado di agevolare i calcolatori elettronici nello scambio dei dati.⁷³ A seguito di quell'incontro fu pubblicata la ISO 12006 *Building construction - Organization of information about construction works*, norma di riferimento internazionale che definisce le raccomandazioni in materia di organizzazione delle informazioni nel settore delle costruzioni. La ISO 12006 si compone di due parti:

- Part 2 - *Framework for classification*;
- Part 3 - *Framework for object-oriented information exchange*.

Ai fini della presente ricerca entrambe le parti risultano essere significative: la prima in quanto maggiormente orientata a questioni di natura concettuale; la seconda perché incentrata sulla necessità di sviluppare formati standard per la distribuzione e lo scambio delle informazioni.

La Parte 2 "*Framework for classification*" propone per l'appunto un framework di riferimento per la realizzazione di sistemi di classificazione e identifica seguendo un approccio per punti di vista, una serie di tabelle che possono essere intese come macro-categorie, in cui classi di oggetti possono essere organizzati. Il modello proposto non vuole essere un'imposizione e non ha le pretese di essere esaustivo, esso può infatti essere adattato ed adeguato in funzione del contesto di applicazione e delle consuetudini vigenti al suo interno.

In particolare le tabelle specificate sono:

- *Construction information (by content)*

⁷² <

http://www.iso.org/iso/home/standards_development/list_of_iso_technical_committees/iso_technical_committee.htm?commid=49180> (ultima consultazione 16/12/2014).

⁷³ IFD Library for buildingSMART. IFD Library white paper <http://www.ifd-library.org/images/IFD_Library_White_Paper_2008-04-10_I_.pdf>

- *Construction products* (by technical function or form or material or any combination of these)
- *Construction agents* (by discipline or role or any combination of these)
- *Construction aids* (by technical function or form or material or any combination of these)
- *Management* (by management activity)
- *Construction process* (by construction activity or construction process lifecycle or any combination of these)
- *Construction complexes* (by function or user activity or any combination of these)
- *Construction entities* (by form or function or user activity or any combination of these)
- *Built spaces* (by form or function or user activity or any combination of these)
- *Construction elements* (by technical function or form or position or any combination of these)
- *Work results* (by work activity)
- *Construction properties* (by property)

La parte 3 “*Framework for object-oriented information exchange*” specifica un modello tassonomico le cui principali entità sono: i concetti (*objects*), le classi di concetti (*collections*) e le relazioni tra concetti (*relationships*). Il modello proposto è riportato nel testo della norma in formato EXPRESS⁷⁴ che permette di specificare univocamente i dati e di definire tra questi dei vincoli.

Nel 2006 le stesse organizzazioni promotrici della ISO 12006 firmarono un *agreement* con lo scopo di coniugare sinergicamente le loro competenze e soprattutto i loro interessi dando vita all’*International Framework for Dictionaries (IFD)*⁷⁵ *Library* ossia lo standard per la definizione di dizionari multilingue e/o ontologie di settore. IFD è uno degli elementi essenziali della tecnologia BuildingSMART⁷⁶ insieme all’*Industry*

⁷⁴ <http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38047> (ultima consultazione 16/12/2014).

⁷⁵ <<http://www.ifd-library.org/index.php?title=IFDWeb:About>> (ultima consultazione 16/12/2014).

⁷⁶ <<http://www.buildingsmart.org/>> (ultima consultazione 16/12/2014).

*Foundation Classes (IFCs)*⁷⁷ e all'*Information Delivery Manual/Model View Definition (IDM/MVD)*.⁷⁸

In IFD i concetti e le etichette che ad essi è possibile assegnare costituiscono entità distinte e messe in comunicazione grazie alla specificazione di una *naming relationship*. Ciascun concetto infatti, può essere identificato per mezzo di più etichette così come la medesima etichetta può veicolare più concetti soprattutto quando ad essere messe in relazione sono più lingue. Un'impostazione di questo tipo consente al concetto di poter essere più o meno dettagliato tenendo conto di fattori contestuali. Inoltre per garantire l'interoperabilità tra più applicazioni e ovviare a problematiche di ambiguità semantica, IFD identifica ciascun concetto per mezzo di un *Global Unique Identifier* (GUID).

Dal punto di vista prettamente tecnico, esistono nel settore delle costruzioni diversi standard e protocolli orientati all'interoperabilità, tra cui il succitato IFCs⁷⁹ e lo Standard for Exchange of Product model data (STEP)⁸⁰. In aggiunta a questi e in risposta alle esigenze di una maggiore rappresentazione avanzata del web semantico è stato sviluppato il formato l'ifcXML (*Industry Foundation Classes eXtensible Markup Language*).⁸¹

⁷⁷ <<http://www.buildingsmart-tech.org/specifications/ifc-releases/summary>> (ultima consultazione 16/12/2014).

⁷⁸ <<https://www.buildingsmartcanada.ca/resources/resource/terms-and-definitions/>> (ultima consultazione 16/12/2014).

⁷⁹ IFCs è un format per lo scambio di dati largamente impiegato nel contesto del Building Information Modeling (BIM). Si tratta di un diverso approccio alla progettazione e alla documentazione del progetto edilizio che sfrutta le potenzialità di strumenti informatici e protocolli per agevolare lo scambio e la condivisione di dati e informazioni.

⁸⁰ <http://www.steptools.com/library/standard/step_1.html> (ultima consultazione 16/12/2014).

⁸¹ <<http://www.buildingsmart-tech.org/specifications/ifcxml-releases>> (ultima consultazione 16/12/2014).

CAPITOLO 2

IL WEB SEMANTICO E L'APPROCCIO LINKED DATA

Il capitolo precedente incentrato sul concetto di interoperabilità, ha anticipato l'insieme delle problematiche derivanti dalla mancata applicazione di procedure comuni e ha segnalato alcune strategie attraverso cui dati e informazioni possono essere resi interoperabili nella forma e nel significato. Inoltre, è stata richiamata la necessità di utilizzare modalità standard per lo scambio e la corretta interpretazione di dati e informazioni attraverso l'adozione e l'impiego di principi propri del web semantico e dell'approccio linked data.

Tale esigenza è stata ripresa da Karen Coyle⁸² in una relazione presentata al seminario *Global Interoperability and linked data in libraries*⁸³ tenutosi a Firenze nel giugno del 2012, in cui invita le biblioteche a considerare il web semantico e l'insieme delle *best practices* da questo promosse come un'opportunità di crescita e sviluppo. Nel suo discorso Karen Coyle sottolinea i presupposti ritenuti indispensabili affinché i dati bibliografici - dotati di un livello qualitativo molto alto - diventino parte integrante del web: rendere i dati conformi ai più recenti canoni del web e collegarli a quelli ospitati in altri *repository*, cataloghi e/o collezioni⁸⁴ nella consapevolezza che da questa unione i dati si arricchiscano e acquisiscano maggiore valore.

La centralità di questi due aspetti deve indurre le biblioteche e gli istituti culturali ad estendere e a diffondere l'utilizzo di queste nuove tecnologie, incoraggiati dal fatto che seguire i paradigmi del web semantico non sia solo una questione di vita o di morte "*a matter of life or death, existence or disappearance*",⁸⁵ ma un servizio da rendere agli utenti mettendo a loro disposizione strumenti che consentano il recupero, nel modo più efficiente possibile, del materiale informativo di interesse.

⁸² <<http://kcoyle.net/>> (ultima consultazione 16/12/2014).

⁸³ Il seminario è stato promosso da: Università degli studi di Firenze, Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU), Biblioteca nazionale centrale di Firenze (BNCF), Casalini Libri, Comune di Firenze, Conferenza dei rettori delle università italiane (CRUI), Associazione italiana biblioteche (AIB), Istituto di teoria e tecniche dell'informazione giuridica del Consiglio nazionale delle ricerche (ITTIG-CNR), Fondazione Rinascimento digitale. Il seminario ha affrontato le tematiche relative all'interoperabilità di dati e informazioni sul web attraverso le best practices promosse dall'approccio linked data. <<http://www.linkedheritage.eu/linkeddatabseminar/>> (ultima consultazione 16/12/2014).

⁸⁴ «cfr.» KAREN COYLE, *Library Linked data: un'evoluzione*, «JLIS.it», vol. 4, n. 2, Gennaio 2014, DOI:10.4403/jlis.it-5443, <http://kcoyle.net/presentations/FI2012_it.pdf> (ultima consultazione 16/12/2014).

⁸⁵ *Ibidem*.

In questo capitolo saranno presentati i tratti principali che hanno determinato il passaggio dal web ipertestuale o web dei documenti al web semantico o web dei dati⁸⁶ e sarà inoltre, illustrato nel dettaglio l'approccio linked data inteso come strumento a supporto dell'interoperabilità tecnica e semantica tra dati e informazioni eterogenei.

1. Dal web tradizionale al web semantico

Il web così come lo intendiamo oggi è il risultato dell'evoluzione di una proposta che Tim Berners-Lee avanzò nel 1989 al *Conseil Européen pour la Recherche Nucléaire* (Cern) in merito all'organizzazione e alla gestione della mole di informazioni generata dalle attività e dagli esperimenti condotti dall'organizzazione, presso cui lavorava. Nello specifico si proponeva di sfruttare un sistema di collegamenti, meglio noti come *hyperlinks* per connettere, mettere in comunicazione e monitorare la conoscenza acquisita dai diversi gruppi di ricerca. Il modello basato su logiche maggiormente dinamiche e flessibili rispetto alle soluzioni fino a poco tempo prima adottate, raggiunse un ampio consenso e il modello fu progressivamente esteso al altre realtà fino a divenire di carattere mondiale.

Già a quel tempo Tim Berners-Lee prospettava una visione avanzata del web in cui la pratica di strutturare i dati e le informazioni in forma tabellare era poco adatta per gestire la loro costante evoluzione. Egli anticipò l'uso di nodi e archi per descrivere e rappresentare l'evoluzione della conoscenza.⁸⁷ I nodi intesi come insieme finito di oggetti (in matematica punti) e gli archi, ovvero le relazioni esplicitate tra coppie di oggetti sono gli elementi fondanti della teoria dei grafi.

I grafi sono delle rappresentazioni, per l'appunto grafiche, molto intuitive che consentono di descrivere in maniera chiara e comprensibile una grande quantità di oggetti e di definire tra di essi, altrettante sofisticate correlazioni.

⁸⁶ Meglio definito come "*a web of things in the world, described by data on the Web*", «cfr.» CHRISTIAN BIZER, et alii, *Linked Data - The Story So Far*, «International Journal on Semantic Web and Information Systems», vol. 5, n. 3, 2009, p. 1-22, DOI 10.4018/jswis.2009081901, <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>> (ultima consultazione 16/12/2014).

⁸⁷ Il testo completo della proposta è consultabile al seguente indirizzo: <<http://www.w3.org/History/1989/proposal.html>> (ultima consultazione 16/12/2014).

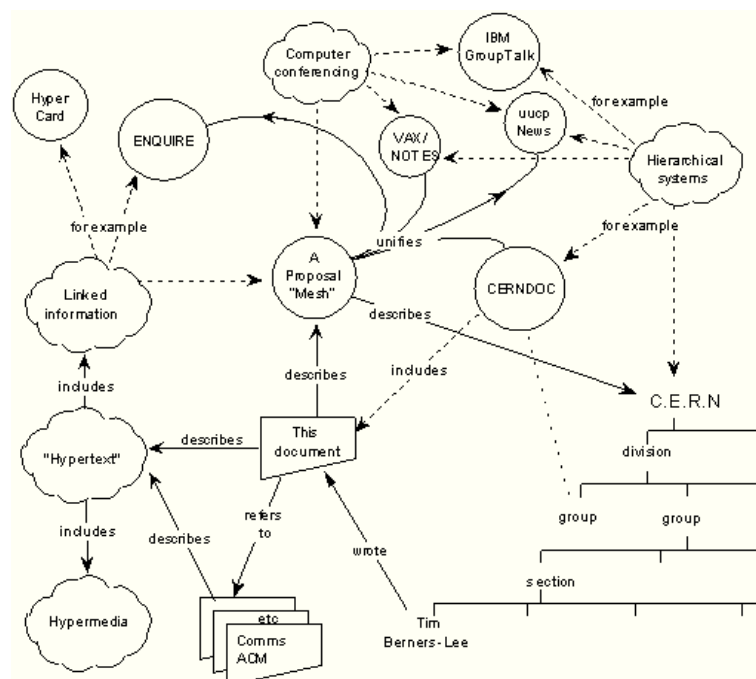


Figura 2. La visione originaria del web.⁸⁸

Il principio dell'interconnessione infatti, è stato riconosciuto come uno dei cardini del web e del suo funzionamento, indipendentemente dall'evoluzione costante che lo caratterizza: nel web ipertestuale i collegamenti erano tra documenti mentre nel web semantico la proiezione è spostata sui dati e sulle loro connessioni.

Questa inversione di prospettiva che va dall'informazione al dato, può essere intesa come la messa in pratica di un'euristica del passato, *divide et impera*, ovvero scomporre la complessità, in questo caso l'informazione, in elementi atomici, i dati, per renderla *machine readable* (leggibile dalla macchina) e *machine understandable* (comprensibili dalla macchina).

Il web attuale si basa su tre elementi essenziali:

- URL (*Uniform Resource Locator*): è una stringa di caratteri che rappresenta e identifica in maniera univoca una risorsa⁸⁹ (fisica e/o astratta) sul web;
- HTTP (*Hyper Text Trasfer Protocol*): è il protocollo applicativo adottato per il trasferimento e lo scambio di risorse (file testuali, file sonori, video, file multimediali, file eseguibili, ecc.) da un *client*⁹⁰ a un *server*,⁹¹

⁸⁸ *Ibidem*.

⁸⁹ <http://www.w3.org/TR/url/> (ultima consultazione 16/12/2014).

- HTML (*Hyper Text Markup Language*): è il linguaggio di marcatura⁹² utilizzato per descrivere e rappresentare le pagine web in un formato comprensibile all'umano e alla macchina⁹³ e consente di strutturare il testo di una pagina web ricorrendo ad un set di marcatori o *tag* predefinito, prioritariamente di tipo stilistico e riguardanti la presentazione e la strutturazione del testo. HTML è quindi un linguaggio orientato al layout delle pagine ma dispone del marcatore “” grazie al quale è possibile definire collegamenti tra le pagine.

Dati questi elementi, ne consegue che le pagine web sono scritte in linguaggio HTML, che ciascuna di esse è identificata per mezzo di URL e che i collegamenti sono garantiti dal protocollo HTTP.

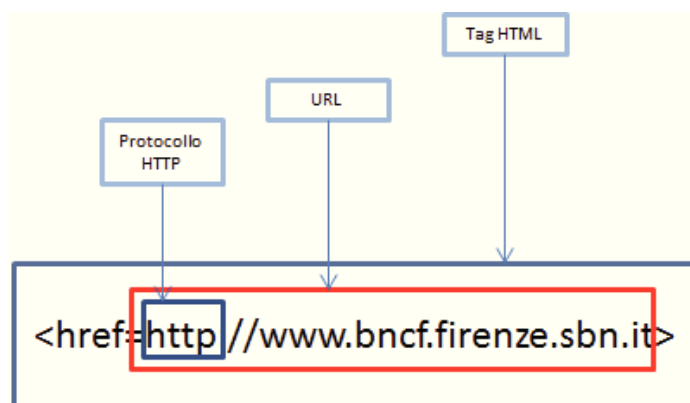


Figura 3. Elementi essenziali del web: http, URL, *tag* HTML.⁹⁴

Il web è un ambiente pensato per rispondere ad un bisogno informativo espresso da un utente attraverso l'uso di termini che nella maggior parte dei casi, provengono dal linguaggio naturale e/o da consuetudini linguistiche territoriali o di dominio.

⁹⁰ “A program that establishes connections for the purpose of sending one or more HTTP requests”. *Linked Data Platform 1.0*, <<http://www.w3.org/TR/ldp/>> (ultima consultazione 16/12/2014).

⁹¹ “A program that accepts connections in order to service HTTP requests by sending HTTP responses”, *Linked Data Platform 1.0*, <<http://www.w3.org/TR/ldp/>> (ultima consultazione 16/12/2014).

⁹² Il linguaggio di marcatura o *mark up* sono un insieme di convenzioni per la marcatura dei testi e possono essere procedurali e dichiarativi. I linguaggi di marcatura procedurali constano di un insieme di istruzioni per la formattazione del testo ai fini della visualizzazione a video o della stampa. I linguaggi dichiarativi invece, consentono di specificare la funzione logica e astratta assolta dalla porzione di testo a cui sono associati. <http://www.elearninglab.eu/learning_paths/introduzione-XML-linguaggi-markup.html> (ultima consultazione 16/12/2014).

⁹³ HTML è attualmente alla sua quinta versione HTML5. <<http://www.w3c.it/it/1/ufficio-italiano-w3c.html>> (ultima consultazione 16/12/2014).

⁹⁴ L'esempio ripreso dal file sorgente della homepage della Biblioteca Nazionale centrale di Firenze (BNCF) esprime il collegamento alla pagina dedicata all'Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU).

Il linguaggio HTML impiegato per la costruzione delle pagine web, descrive e rappresenta il loro contenuto ma non è in grado di comprenderne il significato. Con HTML infatti, possono essere definire collegamenti logici tra le pagine ma non è possibile in nessun modo specificarne la natura o veicolare il significato. Questo rappresenta uno dei principali limiti del web dei documenti e dei motori di ricerca che sfruttano questo linguaggio per recuperare le informazioni. Essi infatti, ricorrendo all'ausilio di algoritmi statistici e basandosi sulla ricorrenza del termine e sul valore di rilevanza che un dato termine assume nel documento rispetto all'intera collezione di documenti, restituiscono una quantità eccessiva di risultati, molti dei quali resteranno sicuramente inesplorati. È calcolato infatti, che il numero massimo di documenti che l'utente è disposto a leggere per verificarne l'affinità e la congruenza rispetto alla richiesta effettuata è di circa cinquanta.⁹⁵ È evidente quindi come l'attività di ricerca e recupero delle informazioni sia un'attività dispendiosa in termini di tempo.

Ad esempio effettuando una ricerca utilizzando come parola chiave il termine *quadro*, circa i primi cinquanta risultati (in *Google* le prime cinque pagine) si riferiscono al *quadro* in quanto oggetto o strumento per la tessitura e non ad esempio in quanto parte di un impianto elettrico o come caratteristica di un motore.⁹⁶

Se invece, la medesima ricerca viene effettuata a partire ad esempio da Wikipedia, che in questo caso funge da base di conoscenza⁹⁷ il risultato che si ottiene mostra la molteplicità dei significati che il termine *quadro* può veicolare a seconda del contesto d'uso.



Figura 4. Pagina di disambiguazione di Wikipedia per il termine *quadro*.

⁹⁵ WILLIAM HERSH, *Information Retrieval. A health and biomedical perspective*, «Journal of the Medical Library Association», vol. 93, n. 1, New York, NY: Springer, 2005² (Collana: Health Informatics Series), p. 201, p. 134-135. ISBN 0-387-95522-4.

⁹⁶ In questo caso la ricerca è stata tramite *Google*.

⁹⁷ < <http://searchcrm.techtarget.com/definition/knowledge-base> >.

Per cui mentre Wikipedia disambigua i termini e rende esplicite le diverse accezioni che il medesimo termine può assumere, nel web questa selettività o quanto meno questa specificazione al momento non è possibile.

Tali limitazioni portano a considerare il web dei documenti come una vasta base di dati distribuita, un grande supporto di memorizzazione in cui tutto ruota intorno al documento e/o a parti che lo compongono, in cui i collegamenti ipertestuali sono stabiliti esclusivamente tra documenti e non tra gli elementi atomici presenti al loro interno. I collegamenti non sono tipizzati e non forniscono informazioni sulla natura della relazione che intercorre tra le pagine ma guidano l'utente nella ricerca delle informazioni tramite *browsing*.

Il *browsing*, come accennato, è una modalità di recupero delle informazioni molto utile ma allo stesso tempo restrittiva, in quanto non definisce in maniera chiara il percorso di ricerca da seguire e inoltre, non dà garanzia che al termine della sessione l'utente riesca a recuperare l'informazione di interesse.⁹⁸

Il web è indubbiamente un grande contenitore di pagine, informazioni e dati utili e significativi. Tuttavia, richiede la messa a punto di soluzioni tecnologiche e non solo, in grado di favorirne una maggiore strutturazione e una migliore rappresentazione affinché i suoi contenuti siano accessibile da parte delle macchine e conseguentemente ricercabili e riutilizzabili.

I vincoli emersi da questa descrizione e inerenti alla fruizione di contenuti connota gli utenti in quanto beneficiari e consumatori degli stessi, senza considerare che nella realtà essi partecipano attivamente alla loro produzione, collaborano sinergicamente tra loro e con i provider all'interno di comunità sociali, contribuiscono cioè alla realizzazione del Web 2.0,⁹⁹ che non introduce innovazioni nella tecnologia e nei formati, ma riconosce agli utenti un ruolo attivo, produttori e non più soli consumatori di contenuti.¹⁰⁰

Questa nuova forma d interazione in cui gli utenti si riscoprono promotori di contenuti e significati, aumenta notevolmente i rischi derivanti dalla poca attenzione rivolta all'interoperabilità sia tecnica che semantica, dovuta alla diffusione di differenti formati

⁹⁸ SOUMEN CHAKRABARTI, *Mining the Web: Discovering Knowledge from Hypertext Data, Parte 2*, San Francisco, Morgan Kaufmann, 2003, p. 18. ISBN: 1-55860-754-4.

⁹⁹ Questo termine è stato coniato da O'reilly il quale lo intende come: " *un insieme di tendenze economiche, sociali e tecnologiche che formano insieme la base per la prossima generazione di Internet – un più maturo e distinto mezzo caratterizzato dalla partecipazione degli utenti, dall'apertura e dagli effetti della rete*", GIULIO PRATI, *Web 2.0. Internet è cambiato*, Trento, UNI Service, 2007, p.16. ISBN 9788861780187.

¹⁰⁰ Il termine impiegato per indicare il ruolo attivo degli utenti è *producer* ovvero produttori e consumatori allo stesso tempo.

e ad usi linguistici diversificati, ecco perché è sempre più stringente disporre di strumenti e linguaggi in grado di gestire e limitare laddove possibile questa disomogeneità.

La risposta a questa limitazione è fornita dal web semantico, il quale si propone l'interoperabilità dei contenuti delle pagine web conferendo loro una struttura ed esprimendo la natura dei collegamenti in maniera precisa e condivisa ricorrendo all'ausilio di vocabolari controllati (reti concettuali e/o ontologie). L'obiettivo è di sfruttare strumenti ed espressività di taluni linguaggi formali, per garantire non più la sola interazione del tipo uomo-macchina ma incentivare quella macchina-macchina.

Questa nuova prospettiva pertanto, non fa altro che spostare l'enfasi dall'informazione al dato e dai collegamenti tra documenti ai collegamenti tra dati. È questo un modello in cui i metadati esprimono le loro potenzialità nella rappresentazione strutturale delle pagine divenendo il mezzo per catturare il contenuto semantico ad esse associate.

Il web semantico nato come evoluzione del web dei documenti è una rete interconnessa di dati facilmente accessibile e comprensibile da qualsiasi computer grazie alla predisposizione di agenti software intelligenti.¹⁰¹ Il web semantico non è diverso dal web tradizionale e non è un'entità che si sviluppa in maniera parallela, ma una sua componente la cui intenzione è migliorarlo incrementandone le potenzialità attraverso lo sviluppo di schemi e modelli condivisi che possano essere via via impiegati da una comunità di utenza sempre più vasta.

La condivisione di linguaggi e pratiche fa sì che un utente possa ricercare all'interno di database differenti la medesima informazione in quanto essa è codificata e rappresentata seguendo regole comuni. Questa non è una forma di omologazione in cui tutti i concetti debbano essere identificati mediante la medesima etichetta, si tratta piuttosto di allineare i propri schemi e modelli a strutture maggiormente condivise e stabili. Lo scopo è quindi definire una immensa collezione di dati interpretabili e elaborabili dalle macchine.¹⁰²

Il web semantico può giocare un ruolo importante negli attuali scenari organizzativi, potrebbe ad esempio supportare il processo di *Knowledge management* e quindi le diverse fasi che lo interessano: acquisizione, organizzazione, accesso e mantenimento

¹⁰¹ «cfr.» LEE FEIGENBAUM et alii, *The Semantic Web in Action*, «Scientific American», vol. 297, December 2007, p. 90-97, <<http://thefigtrees.net/lee/sw/sciam/semantic-web-in-action>>. (ultima consultazione 16/12/2014).

¹⁰² «cfr.» FRANCESCA DI DONATO, *Lo stato trasparente. Linked open data e cittadinanza attiva*, Pisa, ETS, 2010. ISBN: 9788846728876. <<http://www.linkedopendata.it/wp-content/uploads/statotrasparente.pdf>> (ultima consultazione 16/12/2014).

della conoscenza acquisita a partire dai documenti e dall'esperienza maturata dalle persone, attraverso la messa a punto di un modello concettuale comune espresso nei formalismi del web semantico grazie ai quali a partire da una conoscenza nota sia possibile inferire nuova conoscenza.¹⁰³

L'architettura del web semantico si articola su più strati (*layer cake*) ciascuno dei quali identifica una progressione rispetto allo strato precedente. Questo approccio per piccoli passi è in (Antoniou, Van Harmelen, 2004)¹⁰⁴ dovuto alla consapevolezza che è più semplice raggiungere il consenso introducendo piccoli ma costanti innovazioni invece che proporre modelli complessi e di difficile comprensione.

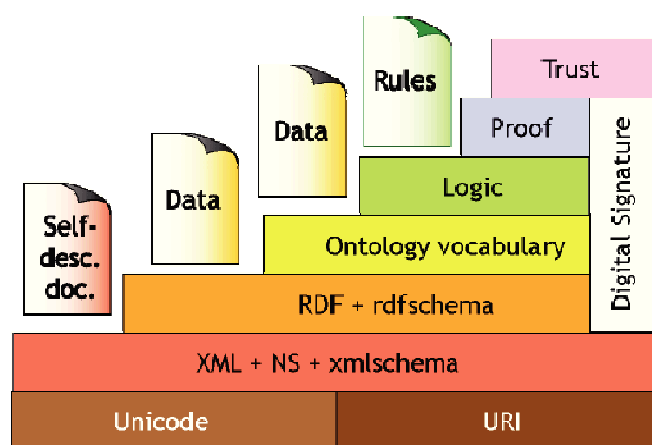


Figura 5. Visione a strati del web semantico.¹⁰⁵

L'immagine rappresenta la struttura a strati del web semantico e dà evidenza della logica graduale che vi è alla base e cioè che lo strato successivo deve essere in grado di interpretare quello precedente (ad esempio RDF deve comprendere la codifica XML) e allo stesso tempo il livello inferiore deve - anche se parzialmente - esprimere quello superiore (ad esempio RDF deve favorire la comunicazione con OWL).

Per cui gli elementi costitutivi del web semantico sono:

- ❖ UNICODE: set di caratteri che permette di identificare in maniera univoca i segni e le lettere appartenenti a tutti i sistemi linguistici esistenti.¹⁰⁶ UNICODE

¹⁰³ GRIGORIS ANTONIOU, FRANK VAN HARMELLEN, *A Semantic Web Primer*, Cambridge, The MIT Press, 2004, p. 4. ISBN 0-262-01210-3.

¹⁰⁴ *Ivi*, p. 16.

¹⁰⁵ ANTONELLA DORATI, STEFANIA COSTANTINI, *Approcci al Web Semantico*, <<http://www.websemantico.org/articoli/approcciwebsemantico.php>> (ultima consultazione 16/12/2014)..

¹⁰⁶ <<http://www.unicode.org/standard/translations/italian.html>> (ultima consultazione 16/12/2014).

associa a ciascun carattere del testo un codice rendendolo di fatto comprensibile e interpretabile dalla macchina;

- ❖ *Uniform Resource Identifier (URI)*:¹⁰⁷ stringa di caratteri che identifica univocamente risorse sul web (documenti e/o oggetti). Un URI può comprendere un *Uniform Resource Locator (URL)* che specifica la collocazione di una risorsa sul web e un *Uniform Resource Name (URN)* che esplicita il nome assegnato ad una certa risorsa;
- ❖ *eXtensible Markup Language (XML)*: metalinguaggio sviluppato per sopperire alle limitazioni di HTML nella rappresentazione delle pagine web, non fornisce informazioni sulla semantica dei marcatori ma le regole per il loro impiego. Può essere considerato come il primo vero tentativo di arricchire di semantica i *tag* associati ai contenuti delle pagine web. XML è un linguaggio: 1) dichiarativo perché usa i *tag* per veicolare la funzione logica assolta dalla porzione di testo a cui il *tag* è assegnato; 2) strutturato in quanto permette di definire e organizzare il testo in strutture logiche complesse; 3) gerarchico poiché si sviluppa per livelli creando una vera e propria gerarchia tra i *tag*;¹⁰⁸
- ❖ *XML namespaces (NS)*:¹⁰⁹ specifica che fa parte della sintassi XML e indica che un dato elemento utilizza nella sua codifica nomi e regole dichiarate nel *namespace* specificato. Ad esempio `<articolo xmlns="http://www.dominio.it/xml/articolo">` vuol dire che l'elemento articolo usa i nomi e le regole di combinazione specificate dal namespace, `http://www.dominio.it/xml/articolo`;¹¹⁰
- ❖ *XML Schemas*: definisce gli elementi, le regole combinatorie degli elementi e i vincoli che li legano e che possono essere impiegati per la descrizione di classi di documenti.¹¹¹
- ❖ *RDF*: il linguaggio di base del web semantico grazie al quale si concretizza l'interoperabilità tecnica. RDF è un modello per la rappresentazione di risorse e di collegamenti tra risorse e si basa su tre elementi fondamentali: soggetto, predicato e oggetto, la loro sequenza è detta tripla, asserzione o *statement*;

¹⁰⁷ <http://www.w3.org/TR/2008/NOTE-cooluris-20081203/#oldweb> (ultima consultazione 16/12/2014).

¹⁰⁸ ALESSANDRO LENCI, et alii, *Testo e computer*, Roma, Carocci Editore, 2005, (Collana: Studi Superiori, 829), p. 72-73. ISBN 9788843034253.

¹⁰⁹ <http://www.w3.org/TR/REC-xml-names/> (ultima consultazione 16/12/2014).

¹¹⁰ L'esempio è stato ripreso dalle slides di Michela Palucci presentate al corso "Elementi di modellazione e Programmazione di contenuti digitali", http://www.dsi.unifi.it/users/palucci/downloads/Elaborazione/Elab_25-11_xml.pdf (ultima consultazione 16/12/2014).

¹¹¹ <http://www.w3.org/TR/xmlschema-1/> (ultima consultazione 16/12/2014).

- ❖ *RDF Schema*:¹¹² è l'estensione di RDF e rappresenta il primo passo per la realizzazione di un modello ontologico. RDF Schema è un vocabolario all'interno del quale sono specificate le risorse, le proprietà e i rispettivi vincoli propedeutici alla descrizione di gruppi di risorse e delle relazioni che tra esse intercorrono;
- ❖ *Ontology vocabulary*: è questo il livello ontologico vero e proprio in cui le diverse concettualizzazioni della realtà o meglio di uno specifico dominio sono espresse in maniera consensuale, esplicita e formale. OWL¹¹³ anch'esso componente dell'insieme delle tecnologie attraverso le quali il web semantico prende forma, è un linguaggio che consente di rappresentare e formalizzare un modello ontologico rendendolo elaborabile da agenti software e di conseguenza pronto per essere sottoposto a processi di ragionamento automatico e di inferenza per l'estrazione di nuova conoscenza. Attualmente OWL si trova alla sua seconda versione (OWL2) e permette di rappresentare oggetti, classi di oggetti e relazioni tra oggetti. I suoi obiettivi prioritari sono: 1) favorire l'uso e la diffusione di una semantica comune grazie alla formalizzazione della conoscenza in un modello ontologico condiviso; 2) garantire lo scambio e l'interpretazione della conoscenza tra le diverse applicazioni tecnologiche.

I *layer* successivi a quello ontologico - *Logic*, *Proof* e *Trust* - non sono stati ancora definiti, nel senso che gli studi finora condotti e portati avanti da diversi gruppi di ricerca facenti capo al W3C, non hanno portato alla messa a punto di schemi o di linguaggi formali tali da completare e anticipare in maniera precisa il *workflow* del web semantico. Per cui questi livelli, attualmente sono da considerarsi presupposti e contestualmente condizioni, affinché il web semantico diventi una realtà effettiva e rappresenti un valore aggiunto per la rete e per tutta la comunità di Internet.

Gli strati che come detto non richiamano una tecnologia o un linguaggio preciso sono:

- ❖ *Logic*: individua il livello logico dell'architettura del web semantico e si propone di sviluppare un linguaggio e/o un modello per il superamento dei limiti dei livelli sottostanti per ridurre possibili problematiche legate all'incongruenza di quanto previamente specificato;
- ❖ *Proof*: a questo livello l'obiettivo consiste nel verificare che la conoscenza implicitamente veicolata dai modelli, resa esplicita mediante

¹¹² <<http://www.w3.org/TR/rdf-schema/>> (ultima consultazione 16/12/2014).

¹¹³ <<http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>> (ultima consultazione 16/12/2014).

un processo di inferenza, sia conforme e rispecchi il modello ontologico predisposto;

- ❖ *Trust*: è il livello che si propone di definire l'affidabilità della conoscenza estratta a partire dal *layer* precedente (*Proof*) associando a ciascuna informazione da cui essa deriva un valore di fiducia, come indice di tutela per l'utente. Questo aspetto si ricollega ad un altro molto avvertito nel contesto del web semantico e che riguarda la definizione di un modello in cui specificare le informazioni inerenti alla provenienza dei dati.¹¹⁴

Trasversale ai diversi livelli della struttura del web semantico prevista da Tim Berners Lee si colloca la:

- ❖ *Digital signature* (firma digitale): intesa come strumento di garanzia per l'autenticità, l'affidabilità e la veridicità delle informazioni pubblicate sul web. La firma digitale potrebbe svolgere nel web semantico un ruolo chiave contribuendo ad una sua maggiore diffusione grazie alla garanzia di veridicità delle informazioni. Questa possibilità potrebbe essere un criterio di discriminazione delle informazioni pubblicate sul web ad uso degli utenti, le cui richieste sono sempre più accurate e precise.

Questa descrizione mostra come il web semantico intenda essere uno strumento in sostegno all'utente, un facilitatore che lo accompagni alla scoperta delle informazioni di interesse presenti sul web, che tende a crescere a ritmi esponenziali e i cui contenuti non sempre si distinguono per qualità, accuratezza e autenticità. Grazie al web semantico gli utenti potranno usufruire non solo di strumenti maggiormente innovativi e informazioni più ricche ma potranno essi stessi contribuire a definire collegamenti semantici tra i dati, arricchendo di conseguenza i dati di partenza.

¹¹⁴ Questo campo di ricerca meglio noto come *provenance* mira a definire un modello grazie al quale descrivere le attività e i passaggi che ha subito una certa risorsa digitale prima di raggiungere l'utente finale. <<http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/>> (ultima consultazione 16/12/2014).

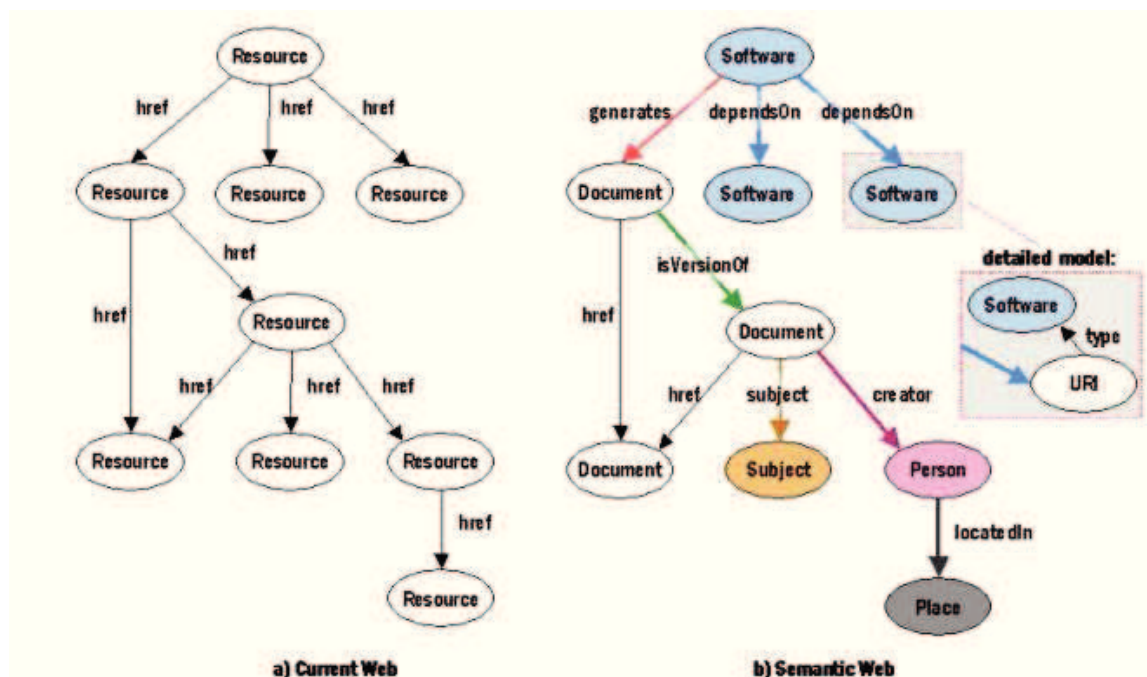


Figura 6. Due visioni a confronto: il web ipertestuale e il web semantico.¹¹⁵

In conclusione, il web semantico riprendendo le parole di (Berners-Lee, et alii, 2001)¹¹⁶ a complemento del titolo: “*A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*” mira a:

- strutturare il contenuto maggiormente significativo delle pagine web affinché gli agenti informatici possano, sfruttando le correlazioni semantiche, navigare tra i contenuti alla ricerca di informazioni rilevanti;
- garantire una sinergica collaborazione tra i diversi utenti e tra questi e le macchine grazie all’uso di formalismi e regole comuni;
- collegare parti di conoscenza legate a più domini del sapere secondo una logica universale e decentralizzata;¹¹⁷
- migliorare il grado di precisione dei risultati ottenuti a seguito di una ricerca *on the fly*¹¹⁸ maggiormente dinamica che restituisce risultati più aggiornati e quindi attendibili.

¹¹⁵ «cfr.» MARJA-RIITTA KOIVUNEN, ERIC MILLER, *W3C Semantic Web Activity*, in *Proceedings of the Semantic Web Kick-off Seminar*, Finland, 2nd November 2001, 2002, <<http://www.w3.org/2001/12/semweb-fin/w3csw>> (ultima consultazione 16/12/2014).

¹¹⁶ TIM BERNERS-LEE, *Op. cit.*

¹¹⁷ *Ibidem.*

¹¹⁸ «cfr.» OKTIE HASSANZADEH, et alii, *Data Management Issues on the Semantic Web*, in *ICDE '12 Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*, Washington, DC, USA

2. L'approccio Linked Data come strumento di supporto al web semantico

Come summenzionato, il web semantico è inteso come l'evoluzione del web dove i contenuti presentati in una forma maggiormente strutturata sono veicolati sotto forma di dati. Si assiste quindi alla reingegnerizzazione del processo di gestione e recupero dell'informazione in cui partendo dal livello più basso ovvero il dato, si procede progressivamente ad una sua ottimizzazione arricchendolo grazie alla specificazione di metadati che gli attribuiscono una struttura e un significato.

Questi elementi connotano il web semantico come web dei dati, il cui obiettivo consiste nella costruzione di una rete di dati formalizzati e resi interpretabili dalla macchina ricorrendo all'uso di linguaggi formali appositamente realizzati e all'impiego di buone pratiche per la loro produzione e pubblicazione sul web, meglio note come linked data.

Il termine linked data si riferisce all'insieme delle buone pratiche per pubblicare e collegare sul web, dati strutturati secondo formati standard e procedure condivise, di modo da essere comprensibili agli utenti e ai calcolatori elettronici.¹¹⁹ La loro applicazione non fa altro che favorire la progressione del web dei dati poc'anzi citato, la cui missione consiste nel favorire l'arricchimento dei dati di partenza integrandoli con quelli presenti e attestati in altre fonti e/o da domini della conoscenza.

, IEEE Computer Society, 2012, p. 1204-1206, DOI 10.1109/ICDE.2012.141, <<http://disi.unitn.it/~velgias/docs/HassanzadehKV12.pdf>> (ultima consultazione 16/12/2014).

¹¹⁹ «cfr.» TOM HEATH, CHRISTIAN BIZER, *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, 2011, DOI 10.2200/S00334ED1V01Y201102WBE001.

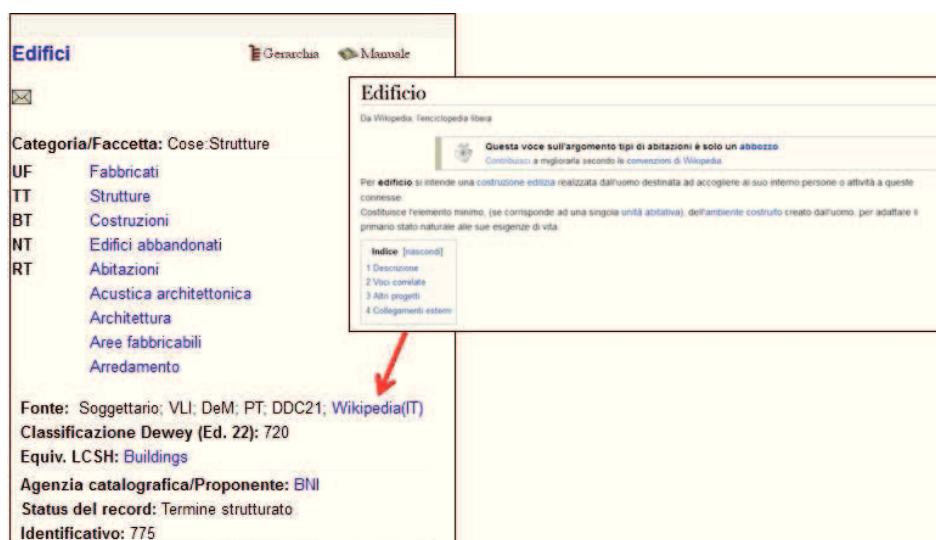


Figura 7. Esempificazione della modalità di arricchimento semantico del Nuovo Soggettario.

L'immagine 7 riporta un esempio di uno dei possibili modi in cui l'arricchimento si realizza. Nello specifico riguarda l'estensione delle informazioni inerenti al concetto di *Edificio* veicolato dall'etichetta *Edifici* all'interno del Nuovo Soggettario grazie alla specificazione di una correlazione con la voce *Edificio* di Wikipedia, definita ricorrendo ad un formalismo del web semantico pensato precisamente per definire associazioni tra sistemi di organizzazione della conoscenza. In questa situazione il Nuovo Soggettario avrebbe potuto avvalersi dell'uso di una nota d'ambito¹²⁰ o di una definizione per specificare meglio il significato espresso dal concetto *Edificio* all'interno dello schema, soluzione che è stata messa da parte in favore della creazione di un collegamento verso la definizione presente in Wikipedia.

Questo esempio mostra la fiducia che la comunità di esperti dedita all'aggiornamento e alla gestione del Nuovo Soggettario ha riposto nell'approccio linked data. La strategia di definire siffatti collegamenti ad altri dataset¹²¹ risulta essere ancor più significativa in quanto applicata dal Nuovo Soggettario, sistema di organizzazione della conoscenza per eccellenza, impiegato per la sua autorevolezza a fini di indicizzazione di risorse documentali e da poco utilizzato in via di sperimentazione per l'indicizzazione automatica di risorse digitali.¹²²

¹²⁰ "Note that defines or clarifies the semantic boundaries of a concept as it is used in the structured vocabulary." «cfr.» ISO, *Op. cit.*

¹²¹ "Un dataset (o data set) è una collezione di dati." <<http://it.wikipedia.org/wiki/Dataset>> (ultima consultazione 16/12/2014).

¹²² ELISABETTA VITI, *Interoperabilità fra thesauri generali e thesauri specialistici in ambito economico finanziario. il caso del Nuovo Soggettario. Sintesi e risultati di un progetto di ricerca*, Seminario nazionale di Biblioteconomia. Didattica e ricerca nell'università italiana e confronti internazionali,

2.1. Caratteristiche e principi dell'approccio linked data

Il termine *linked data* indica "l'insieme delle buone pratiche per pubblicare e connettere dati strutturati sul web". Queste buone pratiche sono state adottate da un numero sempre maggiore di data provider negli ultimi tre anni, contribuendo a creare uno spazio globale di dati che contiene bilioni di asserzioni".¹²³ Sono queste le parole adottate da (Bizer et alii, 2009)¹²⁴ per spiegare il concetto di linked data e incoraggiate dalla risposta positiva proveniente dagli utenti del web. Per cui linked data è un concetto complesso che racchiude al suo interno le seguenti potenzialità: pubblicare i dati secondo certi formalismi (di seguito vedremo quali impiegare a seconda del contesto di applicazione); definire correlazioni tipizzate tra dati assimilabili ma afferenti a dataset differenti; individuare e specificare ulteriori correlazioni a partire da quelle già espressamente indicate.

L'approccio linked data si esplica tenendo conto dei seguenti principi:

- adottare URI per identificare univocamente "things"¹²⁵ attraverso una stringa di caratteri;
- utilizzare il protocollo HTTP per favorire il recupero di "things", garantendo il soddisfacimento del bisogno informativo espresso dall'utente attraverso un meccanismo di deferenziazione;¹²⁶
- in fase di ricerca corredare ciascun URI di informazioni rilevanti ricorrendo agli standard predisposti quali RDF e SPARQL;
- includere link ad altri URI così da poter scoprire ulteriori "things" e contribuire attivamente alla realizzazione di una ragnatela di dati.

L'approccio linked data promuove la condivisione e il riutilizzo dei dati, soprattutto di quelli prodotti e forniti da provider autorevoli (regioni, comuni, biblioteche, musei, archivi, editori, ecc.) per sviluppare dei servizi sempre più innovativi e rispondenti alle esigenze dell'utenza finale.

<http://w3.uniroma1.it/seminario-biblioteconomia/wp-content/uploads/2013/05/Viti_Elisabetta.pdf> (ultima consultazione 16/12/2014).

¹²³ Questa definizione è la traduzione non ufficiale di "*refers to a set of best practices for publishing and connecting structured data on the Web. These best practices have been adopted by an increasing number of data providers over the last three years, leading to the creation of a global data space containing billions of assertions - the Web of Data*". «cfr.» CHRISTIAN BIZER, et alii, *Op. cit.*

¹²⁴ *Ibidem.*

¹²⁵ Con questo termine ci si riferisce a concetti che identificano oggetti concreti e astratti del mondo reale

¹²⁶ Si tratta di un meccanismo che consente di accedere ai campi di altri dataset mediante appositi puntatori, garantendo quindi l'accesso a dati sempre aggiornati.

I principi succitati sono ormai largamente impiegati per la realizzazione di dataset *linked data compliant*. Per incentivarne la pubblicazione e valutarne il grado di rispondenza rispetto ai vincoli linked data, Tim Berners Lee ha ideato nel 2010 una metodologia basata su cinque criteri, ognuno dei quali prevede l'assegnazione di una stella.¹²⁷

★	make your stuff available on the web (whatever format)
★★	make it available as structured data (e.g. excel instead of image scan of a table)
★★★	non-proprietary format (e.g. csv instead of excel)
★★★★	use URLs to identify things, so that people can point at your stuff
★★★★★	link your data to other people's data to provide context

Figura 8. Il criterio delle 5 stelle.

- ★★★★★ pubblicare e rendere disponibili i dati in un qualsiasi formato;
- ★★★★★ pubblicare e rendere disponibili i dati in forma strutturata, privilegiando un foglio Excel rispetto ad un file pdf o ad un'immagine di modo che possano essere *machine-readable*;
- ★★★★★ pubblicare e rendere disponibili i dati privilegiando formati non proprietari ma aperti;
- ★★★★★ pubblicare e rendere disponibili i dati nei formati proposti dal W3C ovvero RDF, SKOS, OWL, ecc.;
- ★★★★★ utilizzare i formalismi del W3C per definire e descrivere i collegamenti tra la propria collezione di dati e i dati ospitati in altri dataset.

Al di là dei tecnicismi che l'uso di questo approccio impone, è importante valutare in che modo questa visione possa contribuire a ridimensionare il problema dell'interoperabilità tra i dati, dovuta da un lato al proliferare di formati proprietari che a causa di interessi economici non sono aperti e condivisi; e dall'altro derivanti dalla

¹²⁷ Il termine vocabolario viene utilizzato in questo contest in senso generale e senza fare distinzione di alcune genere. «cfr.» KRZYSZTOF JANOWICZ, et alii, *Five Stars of Linked Data Vocabulary Use*, «Semantic Web 0», IOS Press, 2014, <<http://www.semantic-web-journal.net/system/files/swj653.pdf>> (ultima consultazione 16/12/2014).

complessità e dall'ambiguità del linguaggio naturale. Entrambe le dimensioni necessitano di essere prese in considerazione in egual misura, per far sì che agli utenti del web, ed in particolar modo a coloro che hanno intenzione per necessità o per desiderio recuperare informazioni di interesse, sia reso un servizio che fornisca dei risultati maggiormente affidabili e quanto più possibili precisi, grazie anche alla predisposizione di sezioni in cui l'accesso ai dati possa essere immediato e semplificato.

Un numero sempre più crescente di utenti sta adottando i paradigmi linked data e ha contribuito a realizzare bilioni di asserzioni RDF inerenti a diversi ambiti della conoscenza (geografia, scienze, salute e benessere, ecc.), molte di queste asserzioni rispecchiano i collegamenti di tipo *cross-domain* (ad esempio messa in comunicazione dei dati inerenti alla localizzazione geografica e alle persone).¹²⁸ Questa modalità di interazione in cui i dati nati in contesti e per obiettivi diversi sono messi insieme e riutilizzati al fine di realizzare servizi innovativi, è una pratica largamente utilizzata.

In (Sabou, 2012)¹²⁹ ci si preoccupa se al momento di decidere la destinazione per un viaggio sia necessario tenere in considerazione oltre ai tradizionali indicatori (ad esempio localizzazione, mezzi di trasporto, pernottamento, ecc.), anche altri elementi quali la situazione economica della zona e/o la sua politica energetica. Si tratta di un chiaro esempio di *cross-domain link* in cui i dati sulla localizzazione, sulle modalità di trasporto sono correlati ai dati economici e a quelli sulla sostenibilità ambientale, per cui il turista può disporre di un numero maggiore di elementi per poter decidere di scegliere o meno di recarsi in un certo paese. In aggiunta, l'esempio mostra un ulteriore campo di impiego dei linked data e che riguarda la realizzazione di sistemi di supporto alle decisioni.

La crescita esponenziale di ambiti in cui l'approccio linked data trova largo consenso unitamente alla quantità di dati che è necessario gestire e il cui valore molto spesso è ancora da scoprire, richiedono che i dati siano liberamente accessibili, aperti e non sottoposti a restrizioni (ad esempio copyright) a meno che queste limitazioni non siano espressamente definite ricorrendo ad una delle licenze *Creative Commons* che consentono di specificare il grado di apertura e la misura in cui è possibile utilizzare i dati da parti terze.¹³⁰

¹²⁸ TOM HEATH, CHRISTIAN BIZER, *Op. cit.*

¹²⁹ MARTA SABOU, et alii, *Supporting Tourism Decision making with linked data*, in *8th Int. Conference on Semantic Systems I-SEMANTICS/I-CHALLENGE*, ACM, Graz, Austria, 2012, p. 201-204, <<http://eprints.weblizard.com/54/>> (ultima consultazione 16/12/2014).

¹³⁰ <<http://creativecommons.org/>> (ultima consultazione 16/12/2014).

Questa necessità ha quindi esteso il concetto da linked data a linked open data (LOD). Quest'ultimo richiede che i dati siano strutturati secondo i formalismi e i principi del web semantico e resi liberamente disponibili e accessibili di modo da poter essere reimpiegati per il raggiungimento di ulteriori obiettivi e per la realizzazione di servizi integrati.

2.2. I linguaggi dell'approccio Linked Data

La pubblicazione dei dati sul web secondo i principi linked data impone come anticipato l'adozione di taluni formalismi appositamente predisposti dal W3C, un consorzio che si occupa di definire standard e linee guida per l'organizzazione e la pubblicazione di contenuti sul web.

Per rispondere a esigenze di interoperabilità sia tecnica che semantica è necessario che i dataset si presentino in formato RDF, un metalinguaggio per la descrizione e la strutturazione dei dati di modo da renderli conformi ai principi del web semantico e quindi pronti per entrare a far parte del web dei dati.

Le specifiche RDF consentono ai dati di poter essere interpretati e compresi non solo da un utente umano ma anche da una macchina, ed è proprio su quest'ultima modalità di interazione che RDF fa leva, ovvero garantire la comunicazione tra macchine sviluppando applicazioni software in grado di recuperare dati non solo sulla base di criteri sintattici ma introducendo anche elementi di natura semantica.

RDF rappresenta i dati in maniera chiara e concisa di modo da poter essere sottoposti al processo di recupero ricorrendo ad un altrettanto preciso linguaggio appositamente realizzato dal *Data Access Working Group*: *SPARQL Protocol and RDF Query Language* (acronimo ricorsivo *SPARQL*).

Nei paragrafi successivi i due formati principali della visione linked data saranno presentati ricorrendo all'ausilio di esempi esplicativi per evidenziare i principi del loro funzionamento.

2.2.1. Resource Description Framework (RDF)

RDF rappresenta nella piramide dell'architettura del web semantico il punto di partenza per la sua realizzazione, ciò implica che i dati per essere esposti sul web debbano essere in formato RDF, il quale si basa su tre concetti ritenuti sostanziali:

- risorsa (*resource*): concetto che identifica qualsiasi oggetto concreto o astratto di cui si vuol parlare, ad esempio un libro, un convegno, un autore, un paese, e altri. Nella logica RDF ciascuna risorsa deve essere identificata in modo univoco tramite un URI;
- proprietà (*property*): concetto che esprime l'insieme delle relazioni possibili tra due o più risorse, ad esempio "titolo", "è curato da", "è nato nel", "è figlio di", e altri. Le proprietà così come le risorse devono essere identificate tramite un URI; ciò implica che la proprietà *title* definita dal *Dublin Core* dovrà essere dichiarata con il rispettivo URI e cioè <http://purl.org/dc/terms/title>. Le proprietà come in questo caso, se inserite all'interno di un vocabolario controllato e condiviso concorrono a garantire una maggiore interoperabilità, a ridurre i numerosi casi di ambiguità semantica e a migliorare l'efficacia della comunicazione tra utente e macchina e tra le macchine;
- asserzione (*statement*): concetto che identifica un'espressione di senso compiuto costituita da un *soggetto*, un *predicato* e un *oggetto*. Soggetto e oggetto identificano una risorsa con una sola eccezione, mentre il soggetto deve sempre essere espresso da un URI, l'oggetto invece, può essere identificato sia da un URI che da un *literal*, vale a dire un valore fisso del tipo stringa e/o numero.

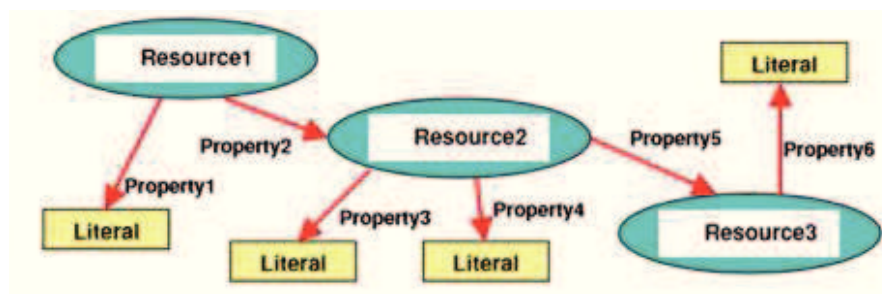


Figura 9. Rappresentazione del modello di descrizione RDF.¹³¹

¹³¹ ORESTE SIGNORE, *RDF per la rappresentazione della conoscenza*, <<http://www.w3c.it/papers/RDF.pdf>> (ultima consultazione 16/12/2014).

Ad esempio si supponga di voler rappresentare sotto forma di grafo RDF l'asserzione *Umberto Eco è autore di "Il nome della rosa"*, nella logica RDF essa sarà tripartita come di seguito:

Umberto Eco → **Soggetto**

è autore di → **Predicato**

"Il nome della rosa" → **Oggetto**

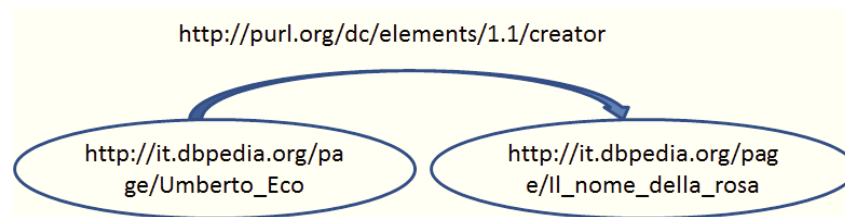


Figura 10. Grafo RDF dell'asserzione *Umberto Eco è autore di "Il nome della rosa"*.¹³²

In questo caso l'oggetto *"Il nome della rosa"* è rappresentato da un URI ma potrebbe anche essere semplicemente un *literal*, indicando di conseguenza un nodo finale della rete che di fatto, interrompe il processo di navigazione.

L'asserzione *Umberto Eco è autore di "Il nome della rosa"* potrebbe a sua volta essere legato ad altre asserzioni, ad esempio *Umberto Eco è nato nel 1932*, *Umberto Eco è nato ad Alessandria*, *Alessandria si trova in Piemonte* e così via. L'interconnessione tra le diverse asserzioni arricchisce il dato di partenza ovvero *Umberto Eco* con approfondimenti via via sempre più dettagliati ma che riguardano i dati associati a quello di partenza, in un certo senso racconta il contesto che ruota intorno ad un dato. Nel caso specifico si recupereranno informazioni inerenti alla città natale dell'autore, agli avvenimenti storici che sono avvenuti nel corso dell'anno della sua nascita, ecc. Questa esemplificazione dimostra l'idea che è alla base del web ipertestuale prima e del web semantico adesso: costruire una rete semantica di grandi dimensioni che consenta di scoprire un numero sempre maggiore di dati rispetto a quelli preventivati in fase di avvio della ricerca.

Per cui i collegamenti, la loro tipizzazione e il loro orientamento costituiscono il valore aggiunto di RDF rispetto a XML che non deve essere inteso come un suo componente

¹³² Nell'esempio gli URI che identificano il *soggetto* e l'*oggetto* sono ripresi dalla pagina web di DBpedia mentre il *predicato* è espresso tramite il set di metadati DCMI ovvero DC:Creator.

ma semplicemente lo strato posto alla base del web semantico. RDF infatti, non si limita a definire un insieme di regole sintattiche da seguire per la strutturazione di un documento, ma consente di descrivere il contenuto semantico dei documenti e grazie alla tipizzazione delle relazioni rende esplicita la natura delle relazioni che possono intercorrere tra i diversi dati RDF.

Nonostante queste peculiarità, RDF non consente di dichiarare il vocabolario da impiegare in fase di strutturazione dei dati, ragion per cui necessita di un meccanismo che metta gli sviluppatori nella condizione di poter definire le classi di oggetti, le proprietà degli oggetti i vincoli tra oggetti e tra proprietà. Questo compito è demandato a RDF Schema, un *data model* che specifica i concetti, la sintassi e la semantica RDF rendendoli *machine-accessible*.¹³³

Per meglio comprendere l'innovazione di RDF Schema rispetto a RDF si veda l'esempio riportato in Figura 10, in cui viene specificata la relazione che intercorre tra le due risorse *Umberto Eco* e *Il nome della rosa*, ma non si evince il livello di astrazione superiore che identifica *Umberto Eco* in quanto *Persona* e *Il nome della rosa* in quanto *Opera intellettuale*.

RDF Schema infatti, pone le basi per la realizzazione di un modello ontologico *machine understandable* e garantisce i presupposti per la definizione di un modello ontologico.

Gli elementi costitutivi di RDF Schema sono:

1. classi (*rdfs:Class*): si tratta di gruppi di risorse identificate da IRI (*Internationalized Resource Identifiers*)¹³⁴ i cui costituenti sono definiti *istanze*.

Le classi possono riferirsi a:

- risorse (*rdfs:Resource*): classe di oggetti che è possibile descrivere in RDF;
- *literal* (*rdfs:Literal*): classe di valori che possono riferirsi a dati fissi e non variabili ad esempio i numeri, i caratteri, e altri;
- *datatype* (*rdfs:Datatype*): classe dei vari tipi di dato quali numeri reali, numeri interi, valori booleani, e altri;
- classi di proprietà (*rdf:Property*): insieme delle proprietà che è possibile definire tra le risorse;

¹³³ GRIGORIS ANTONIOU, FRANK VAN HARMELEN, *Op. cit.*

¹³⁴ Consiste nella generalizzazione dell'URI attraverso l'uso di una sequenza di caratteri più ampia, <<http://www.w3.org/International/O-URL-and-ident.html>> (ultima consultazione 16/12/2014).

2. proprietà (*rdf:Property*): sono stabilite tra un soggetto e un oggetto e possono riguardare:
- sotto-classi (*rdfs:subClassOf*): proprietà che specifica che una certa classe si articola in una o più sottoclassi;
 - sotto-proprietà (*rdfs:subPropertyOf*): proprietà che dichiara che una certa proprietà può contenere al suo interno un'altra proprietà;
 - *label* (*rdfs:label*): proprietà attraverso la quale è possibile specificare in una forma leggibile all'utente umano una risorsa;
 - ecc..¹³⁵

Una tripla RDF prima di essere pubblicata sul web deve essere serializzata ovvero convertita in un formato maggiormente comprensibile da una macchina. I linguaggi di serializzazione raccomandati dal W3C sono:

- *Resource Description Framework in Attributes (RDFa)*:¹³⁶ si propone l'annotazione semantica del contenuto dei documenti esprimendolo in maniera semplice e ricorrendo a linguaggi di marcatura già esistenti come *eXtensible HyperText Markup Language (XHTML)*:¹³⁷
- *Resource Description Framework/ eXtensible Markup Language (RDF/XML)*: ha come obiettivo mappare i nodi e gli archi previsti dalla notazione RDF in elementi e attributi della sintassi XML. RDF/XML è un sistema largamente impiegato che tuttavia risulta di difficile interpretazione all'occhio umano ragion per cui nei casi in cui sia richiesto un intervento umano, altri linguaggi di serializzazione dovrebbero essere presi in considerazione;
- *Notation 3 Logic (N3)*:¹³⁸ questa annotazione seppur molto intuitiva dal punto di vista dell'utente risulta poco adatta all'elaborazione automatica e può essere definita per tale peculiarità una sintassi informale;

¹³⁵ Una descrizione dettagliata unitamente all'elenco completo delle classi e delle proprietà di RDF Schema sono visualizzabili alla pagina web riportati di seguito. <<http://www.w3.org/TR/rdf-schema/#bib-RDF11-CONCEPTS>> (ultima consultazione 16/12/2014).

¹³⁶ <<http://www.w3.org/TR/rdfa-primer/>> (ultima consultazione 16/12/2014).

¹³⁷ «È una famiglia di attuali e futuri tipi di documenti che riproduce, ingloba ed estende l'HTML 4. I tipi di documenti della famiglia XHTML si basano su XML, e sono disegnati fondamentalmente per poter lavorare insieme agli user agent basati su XML. [...]». «cfr.» STEVEN PEMBERTON et alii, *XHTML™ 1.0: Extensible HyperText Markup Language*, 2000, <<http://www.w3.org/TR/xhtml1/>> (ultima consultazione 16/12/2014).

¹³⁸ <<http://www.w3.org/DesignIssues/Notation3.html>> (ultima consultazione 16/12/2014).

- *N-Triples*:¹³⁹ è un formato di testo per la codifica di un grafo RDF, in cui ciascuna linea identifica tramite URI una tripla del tipo soggetto-predicato-oggetto;
- *Terse RDF Triple Language (Turtle)*:¹⁴⁰ questa sintassi è un'estensione di *N-Triples* e riprende alcuni elementi ritenuti significativi della sintassi *N3*. In *Turtle* un grafo RDF è scritto in maniera compatta e le triple soggetto-predicato-oggetto sono separate da uno spazio e dal segno di interpunzione.

2.2.2. SPARQL Protocol and RDF Query Language (SPARQL)

SPARQL *Protocol and RDF Query Language*, è allo stesso tempo: 1) un linguaggio di interrogazione (*query language*), per interrogare appunto, i dati in formato RDF conservati all'interno di *triple store* distribuiti sulla rete; 2) un protocollo che garantisce l'accesso ai dati RDF. Tale linguaggio è stato realizzato in ambito W3C dal *Data Access Working Group*, in qualità di linguaggio di interrogazione non è in grado di inferire nuova conoscenza a partire dai grafi RDF ma restituisce a seguito di una richiesta esattamente i dati dichiarati nei modelli.

Le specifiche di SPARQL si basano sulla logica RDF e sono scritte come triple del tipo soggetto-predicato-oggetto, tale costrutto è definito *basic graph pattern*. Tra la sintassi RDF impiegata per la rappresentazione dei dati e la sintassi SPARQL impiegata invece per l'interrogazione dei dati esiste dal punto di vista logico una chiara similitudine.

A titolo semplificativo la sequenza “*?film dbpedia-owl:starring ?attore*” si compone di un soggetto (“*?film*”), di un predicato (*dbpedia-owl:starring*) e di un oggetto (*?attore*), dove soggetto e oggetto identificano le variabili e sono per l'appunto precedute dal segno di interpunzione (?) mentre il predicato funge da costante.

Dal punto di vista sintattico SPARQL è assimilabile al linguaggio di interrogazione *Structured Query Language (SQL)* impiegato per la gestione e l'interrogazione delle basi di dati. Tale analogia così come si evince dall'esempio riportato in Figura 11 si riscontra nell'uso dell'istruzione SELECT per specificare la lista degli attributi (in questo caso il segno di asterisco (*) fa riferimento a *movie*) e dell'istruzione WHERE usata per specificare la condizione che il risultato deve rispettare.

¹³⁹ <<http://www.w3.org/TR/2004/REC-rdf-testcases-20040210/#ntriples>> (ultima consultazione 16/12/2014).

¹⁴⁰ <<http://www.w3.org/TeamSubmission/turtle/>> (ultima consultazione 16/12/2014).

```

SELECT * WHERE {
  ?movie a
    <http://dbpedia.org/ontology/Film> .
  ?movie <http://dbpedia.org/ontology/country>
    <http://it.dbpedia.org/resource/Italia> .
  .

```

Figura 11. Query SPARQL - *Tutti i film di produzione italiana.*¹⁴¹

Un ulteriore esempio che interroga invece il dataset dedicato ai dati del Senato della Repubblica Italiana, riguardanti l'attività politica e istituzionale mostra in forma completa l'insieme degli elementi costitutivi di una query SPARQL.

```

PREFIX osr: <http://dati.senato.it/osr/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT DISTINCT ?senatore ?nome ?cognome
  ?inizioMandato ?legislatura ?tipoMandato
WHERE {
  ?senatore a osr:Senatore.
  ?senatore foaf:firstName ?nome.
  ?senatore foaf:lastName ?cognome.
  ?senatore osr:mandato ?mandato.
  ?mandato osr:legislatura ?legislatura.
  ?mandato osr:inizio ?inizioMandato.
  ?mandato osr:tipoMandato ?tipoMandato.
  OPTIONAL { ?mandato osr:fine ?df. }
  FILTER(!bound(?df))
} ORDER BY ?cognome ?nome

```

Figura 12. Query SPARQL - *Elenco dei senatori attualmente in carica.*¹⁴²

Ne consegue che gli elementi costitutivi di una query SPARQL sono:

- PREFIX (*prefix*): assegna un prefisso ad un *namespace*, ad esempio foaf:firstName invece di riportare l'URI per esteso ogniqualvolta si richiamino

¹⁴¹ L'esempio è ripreso dal sito della DBpedia italiana. <<http://it.dbpedia.org/esempi/>> (ultima consultazione 16/12/2014).

¹⁴² <<http://dati.senato.it/>> (ultima consultazione 16/12/2014).

elementi appartenenti al medesimo schema ontologico. PREFIX ricopre una doppia funzione, identifica il modello ontologico tramite il quale è possibile recuperare un certo dato e alleggerisce e semplifica la sintassi della query;

- SELECT: specifica l'insieme degli attributi che devono essere riportati nel risultato (senatore, nome, cognome, inizio mandato, legislatura, tipo mandato);
- FROM: indica il dataset da interrogare (ad esempio DBpedia, i dati del Senato, e altri);
- WHERE: definisce i criteri di selezione ovvero i vincoli che i risultati devono soddisfare;
- OPTIONAL: definisce che alcune variabili sono opzionali, ammette quindi anche i casi in cui taluni valori non sono disponibili;
- FILTER: dichiara delle restrizioni sulle variabili;
- ORDER BY: consente di ordinare i risultati in base a certi criteri (ad esempio cognome, nome, ecc.).

La risposta ad una interrogazione SPARQL è rappresentata in forma tabellare, in cui le intestazioni delle colonne sono le variabili dichiarate nell'istruzione SELECT (senatore, nome, cognome, inizio mandato, legislatura, tipo mandato) e le righe costituiscono i risultati restituiti dalla query. Anche i risultati così come la query si presentano sotto forma di triple.

senatore	nome	cognome	inizioMandato	legislatura	tipoMandato
http://dati.senato.it/senatore/29033	"Piero" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	"Aiello" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	2013-03-15	17	"elettivo" ^{^^} < http://www.w3.org/2001/XMLSchema#string >
http://dati.senato.it/senatore/29034	"Alberto" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	"Airola" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	2013-03-15	17	"elettivo" ^{^^} < http://www.w3.org/2001/XMLSchema#string >
http://dati.senato.it/senatore/29035	"Donatella" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	"Albano" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	2013-03-15	17	"elettivo" ^{^^} < http://www.w3.org/2001/XMLSchema#string >
http://dati.senato.it/senatore/29036	"Gabriele" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	"Albertini" ^{^^} < http://www.w3.org/2001/XMLSchema#string >	2013-03-15	17	"elettivo" ^{^^} < http://www.w3.org/2001/XMLSchema#string >

Figura 13. Rappresentazione tabulare di una query SPARQL.¹⁴³

L'interrogazione dei dataset RDF avviene a partire da sezioni appositamente predisposte, *endpoint* SPARQL, da cui è possibile accedere a singoli *repository* di dati e/o interrogare contestualmente dati provenienti da sorgenti distribuite sul web.

¹⁴³ La tabella rappresenta un estratto dei risultati ottenuti eseguendo la query riportata in Figura 13.

3. La diffusione dell'approccio Linked Data

L'approccio linked data, unitamente all'insieme delle soluzioni tecnologiche proposte dal web semantico si è diffuso in diversi contesti di applicazione per il raggiungimento di una varietà di obiettivi: messa in comunicazione di dati generati all'interno del medesimo istituto, azienda, ente; interconnessione tra dati prodotti da più provider per la messa a punto ad esempio di servizi orientati ai cittadini; fruibilità di dati provenienti da fonti eterogenee; condivisione dei dati realizzati in contesti e domini differenti; ecc..

Gli aspetti positivi dovuti ad un loro utilizzo, sono stati e continuano tuttora, ad essere valutati e verificati non solo in ambito accademico-scientifico ma anche in relazione ad altri contesti orientati ad esempio all'organizzazione del patrimonio culturale, alla gestione e all'apertura dei dati governativi (*Open Government*),¹⁴⁴ alla sistematizzazione e alla condivisione della conoscenza di dominio.

La diffusione dell'approccio linked data è confermata dalle numerose iniziative volte a garantire l'interoperabilità e la condivisione dei dati e la cui consistenza è quantificata in base al numero di dataset pubblicati seguendo i principi linked data, i quali sono graficamente rappresentati all'interno di una nuvola, nota come *LOD data cloud*.

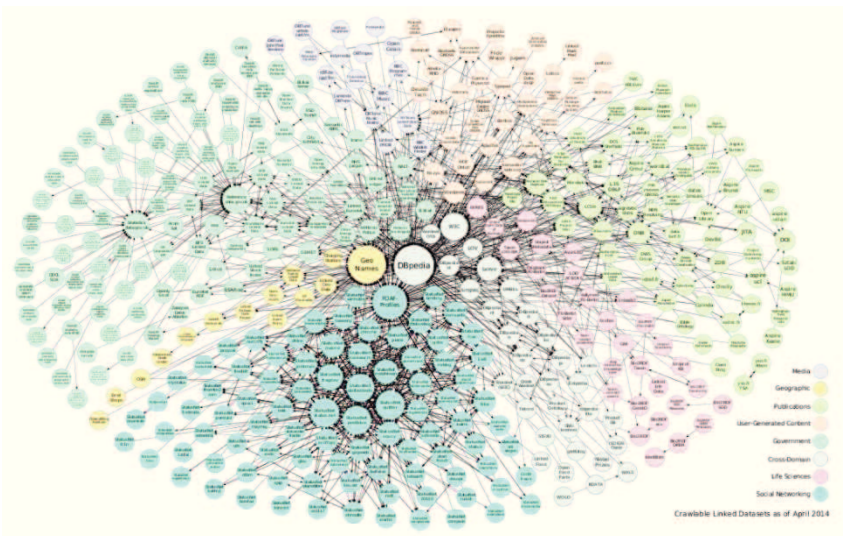


Figura 14. *LOD data cloud* aggiornata a Aprile 2014.¹⁴⁵

¹⁴⁴ "Letteralmente "governo aperto" è un nuovo concetto di Governance a livello centrale e locale, basato su modelli, strumenti e tecnologie che consentono alle amministrazioni di essere "aperte" e "trasparenti" nei confronti dei cittadini.". «cfr.» ASSOCIAZIONE ITALIANA PER L'OPEN GOVERNMENT, *Come di fa Open Data? Istruzioni per l'uso per Enti e Amministrazioni Pubbliche*, a cura di Ernesto Belisario, Gianluigi Cogo, Stefano Epifani, Claudio Forghieri, 2011, p. 7, <<http://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFJAA&url=http%3A%2F%2Fwww.retecamere.it%2Fdownload%2F180.html&ei=ZlXoVN35E6vcywOmgYFA&usg=AFQjCNFCvoRf2SEbsnmNoMRBFXgw8LLn6w&bvm=bv.79142246,d.bGQ>> (ultima consultazione 16/12/2014).

¹⁴⁵ <<http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>> (ultima consultazione 16/12/2014).

La nuvola raccoglie l'insieme dei dataset pubblicati a partire dall'archivio on line CKAN¹⁴⁶ realizzato dalla *Open Knowledge Foundation*¹⁴⁷ in cui per ciascun dataset è riportata una breve descrizione e altre informazioni aggiuntive quali origine, autore, ecc.. I dataset presenti al suo interno tenuto conto del campo disciplinare di appartenenza, sono organizzati nelle seguenti categorie: *Government, Publications, Life sciences, User-generated content, Cross-domain, Media, Geographic, Social web*, a ciascuna delle quali è stata assegnata un colore. Ad esempio la categoria *Publications* contrassegnata in verde contiene dataset quali:

- *data.bnf.fr - Bibliothèque Nationale de France*:¹⁴⁸ è un progetto in cui le risorse create in diversi formati, quali InterMarc¹⁴⁹ per il catalogo generale e l'XML- *Encoded Archival Description* (EAD)¹⁵⁰ per gli inventari di archivio, sono strutturate, descritte e pubblicate secondo i formalismi del web semantico e presentati in un formato *machine understandable*. Il dataset prevede inoltre, collegamenti a dataset esterni quali GeoNames,¹⁵¹ Virtual International Authority File (VIAF),¹⁵² *International Standard Name Identifier* (ISNI)¹⁵³ e altri;
- *Linked Open Vocabularies* (LOV):¹⁵⁴ è un progetto che si propone di raccogliere e raggruppare la molteplicità di vocabolari (terminologie, liste di controllo, thesauri, ontologie, ecc.) attualmente realizzati e impiegati, con l'obiettivo di favorire l'accesso, la visibilità e la fruibilità di questo immenso patrimonio. Inoltre, esso riporta la descrizione per i vocabolari RDFS e per le ontologie OWL dei dataset presenti nella nuvola;

¹⁴⁶ CKAN è l'acronimo di Comprehensive Knowledge Archive Network. <<http://ckan.org/>> (ultima consultazione 16/12/2014).

¹⁴⁷ Si tratta di una comunità no-profit diffusa in tutto il mondo che divulga il principio di apertura dell'informazione attraverso la realizzazione di una rete che consenta alle persone di lavorare insieme e costruire sinergicamente nuova conoscenza. <<https://okfn.org/>> (ultima consultazione 16/12/2014).

¹⁴⁸ <<http://linkeddatacatalog.dws.informatik.uni-mannheim.de/dataset/data-bnf-fr>> (ultima consultazione 16/12/2014).

¹⁴⁹ Si tratta del formato utilizzato dalla Biblioteca Nazionale Francese (BnF) per la gestione dei dati bibliografici conservati la suo interno. <http://www.bnf.fr/en/professionals/intermarc_format_eng/s.intermarc_b_presentation_eng.html> (ultima consultazione 16/12/2014).

¹⁵⁰ È uno standard XML impiegato per la descrizione per la codifica di strumenti di corredo archivistici. <http://en.wikipedia.org/wiki/Encoded_Archival_Description> (ultima consultazione 16/12/2014).

¹⁵¹ È un database liberamente consultabile di dati geografici e contiene circa 10 milioni di nomi geografici. <<http://www.geonames.org/>> (ultima consultazione 16/12/2014).

¹⁵² È un servizio che mette in relazione diversi authority file realizzati da più biblioteche. <<http://viaf.org/>> (ultima consultazione 16/12/2014).

¹⁵³ L'ISNI è un numero che consente di identificare in maniera univoca gli autori e l'insieme degli altri *contributors* intervenuti nella realizzazione di un'opera. <<http://www.isni.org/>> (ultima consultazione 16/12/2014).

¹⁵⁴ <<http://lov.okfn.org/dataset/lov/>> (ultima consultazione 16/12/2014).

- WordNet 2.0 (W3C):¹⁵⁵ è l'equivalente RDF del database lessicale WordNet¹⁵⁶ realizzato dalla *Princeton University* in cui nomi, verbi, aggettivi e avverbi sono raggruppati in base a principi di affinità semantica e lessicale in un insieme definito *synset*. I *synset* sono interconnessi per mezzo di relazioni concettuali del tipo genere-specie, parte-tutto, esemplificativa.

La nuvola riporta dati inerenti ai diversi ambiti della conoscenza il cui nucleo è rappresentato dal succitato database GeoNames e dalla base di conoscenza DBpedia.¹⁵⁷ In particolare, DBpedia sfruttando la logica dei collegamenti ipertestuali è divenuta il collettore dell'insieme dei dati attualmente disponibili in formato RDF per diversi ordini di motivi: è una base di conoscenza molto visitata, è in costante aggiornamento, è disponibile in circa 250 lingue, garantisce una copertura semantica molto estesa e abbraccia i diversi ambiti del sapere.

I collegamenti RDF espressi da e verso DBpedia mettono a disposizione di chiunque una vasta quantità di dati, per cui gli utenti che iniziano la loro navigazione partendo da un singolo dataset possono, grazie ai collegamenti specificati, navigare tra le diverse sorgenti di dati, proseguendo potenzialmente fino al termine del web dei dati.¹⁵⁸

Come succitato, la nuvola dà evidenza delle numerose iniziative avviate a livello internazionale e i cui obiettivi prioritari riguardano l'apertura, la condivisione l'uso e lo scambio di dati, informazioni e conoscenza a favore dell'interoperabilità intesa nella sua dimensione tecnica e semantica. Molte di queste iniziative hanno portato alla costituzione di gruppi di lavoro a cui partecipano esperti internazionali provenienti da diversi ambiti disciplinari, si tratta sia di specialisti del web semantico e sia di specialisti di specifici domini della conoscenza i quali si fanno portavoce delle esigenze e dei bisogni di una certa comunità. A titolo esemplificativo si riportano di seguito alcuni dei gruppi di lavoro costituiti in sede W3C e che possono essere intesi come veri e propri pionieri della visione linked data e in generale del web semantico:

- *Government Linked Data (GLD) Working Group*:¹⁵⁹ è nato nel giugno del 2011 e si propone di fornire agli enti pubblici di ogni ordine e dimensione (dalle amministrazioni locali a quelle nazionali e comunitarie) gli standard e le linee guida per la pubblicazione sul web di dati amministrativi conformi ai principi

¹⁵⁵ <<http://www.w3.org/2006/03/wn/wn20/>> (ultima consultazione 16/12/2014).

¹⁵⁶ <<http://wordnet.princeton.edu/>> (ultima consultazione 16/12/2014).

¹⁵⁷ <<http://dbpedia.org/About>> (ultima consultazione 16/12/2014).

¹⁵⁸ «cfr.» CHRISTIAN BIZER, et alii, *Linked Data on the Web (LDOW2008)*, Workshop Summary at the 17th International World Wide Web Conference, Beijing, China, April 21-25, 2008, p. 1265-1266.

¹⁵⁹ <<http://www.w3.org/2011/gld/charter>> (ultima consultazione 16/12/2014).

linked data, di modo da favorire l'interazione con il tessuto sociale del paese (cittadini, imprese, ecc.). Lo scopo che tale gruppo di lavoro persegue, mira a garantire una maggiore trasparenza ed efficienza delle attività amministrative; una proficua collaborazione con i cittadini mediante la condivisione di idee e proposte in merito al miglioramento delle attività; la semplificazione delle procedure attualmente in uso grazie alla predisposizione di strumenti in grado di agevolare l'accesso ai servizi, ecc.. I paesi che per primi hanno aderito ai nuovi canoni proposti dal web dei dati sono stati gli Stati Uniti d'America (USA)¹⁶⁰ e la Gran Bretagna¹⁶¹ per i quali la trasparenza, la partecipazione collettiva e l'apertura dei dati sono entrati a far parte dell'agenda politica. Anche l'Italia¹⁶² ha aderito ai principi di apertura e condivisione dei dati, tant'è vero che un gruppo di lavoro della Commissione di Coordinamento del Sistema di Pubblica Connettività (SPC) ha redatto in sinergia con alcuni rappresentanti di altrettante autorevoli amministrazioni le *Linee guida per l'interoperabilità semantica attraverso i linked open data*¹⁶³ da seguire, affinché i dati siano conformi rispetto alle raccomandazioni indicate dal *European Interoperability Framework* (EIF);¹⁶⁴

- *W3C Library Linked Data Incubator Group*:¹⁶⁵ è nato nel 2010 al fine di favorire e incrementare l'interoperabilità dei dati realizzati dalle biblioteche e sostenere la collaborazione tra la comunità del web semantico e quella bibliotecaria.¹⁶⁶ Le attività portate avanti da questo gruppo hanno avuto come obiettivo rilanciare il ruolo delle biblioteche nell'era del web semantico aggiornando e adeguando gli standard, gli schemi e le pratiche di lavoro in uso, rispetto a quanto raccomandato dal web semantico, al fine di consentire ai dati bibliografici di essere integrati nel web. Contestualmente il gruppo ha effettuato una disamina delle migliori iniziative avviate in ambito bibliotecario e in altri contesti ad esso correlati, con l'obiettivo di evidenziare empiricamente i benefici riguardanti l'uso delle buone pratiche promosse dall'approccio linked data. L'adozione di questa visione consente di fatto alle biblioteche di incrementare il valore dei propri dati; ridurre la risonanza delle descrizioni delle risorse; scoprire e specificare importanti connessioni tra più fonti fino a questo momento sconosciute;

¹⁶⁰ <<http://www.data.gov/>> (ultima consultazione 16/12/2014).

¹⁶¹ <<http://data.gov.uk/>> (ultima consultazione 16/12/2014).

¹⁶² <<http://www.dati.gov.it/>> (ultima consultazione 16/12/2014).

¹⁶³ COMMISSIONE DI COORDINAMENTO SPC, *Linee guida per l'interoperabilità semantica attraverso i linked open data*, 2012, <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf> (ultima consultazione 16/12/2014).

¹⁶⁴ <http://en.wikipedia.org/wiki/European_Interoperability_Framework> (ultima consultazione 16/12/2014).

¹⁶⁵ <<http://www.w3.org/2005/Incubator/lld/XGR-lld-usecase-20111025/>> (ultima consultazione 16/12/2014).

¹⁶⁶ <<http://www.w3.org/2005/Incubator/lld/>> (ultima consultazione 16/12/2014).

descrivere i dati considerando separatamente la dimensione semantica e quella strutturale; aumentare la visibilità delle biblioteche; conferire alle biblioteche lo status di luogo dell'informazione in cui poter accedere liberamente e gratuitamente ad un'immensa collezione di risorse fisiche e digitali.¹⁶⁷ Uno degli aspetti più significativi derivanti dall'uso delle tecnologie del web semantico riguarda la realizzazione di servizi multilingue di modo da abbracciare una varietà di utenza sempre più vasta garantendo a tutti l'accesso ad informazioni di interesse, per cui le risorse possono essere descritte e arricchite sinergicamente tra più istituti bibliotecari per evitare la duplicazione delle informazioni e conferire loro una maggiore autorevolezza.

- *Open Linguistics Working Group* (OWLG):¹⁶⁸ è un progetto che coinvolge esperti provenienti da diversi domini della conoscenza e che generalmente si occupano di dati di natura linguistica. Esso persegue i seguenti obiettivi riguardanti essenzialmente l'apertura dei dati linguistici, l'estrazione di significato a partire dalla loro rappresentazione, la condivisione multidisciplinare degli stessi. Per dati linguistici si intendono risorse lessicali e semantiche, corpora annotati, database linguistici, liste di metadati, ecc. che rappresentate in RDF e OWL divengono strutturalmente e concettualmente interoperabili. Grazie alla rappresentazione RDF anche i dati linguistici possono essere interconnessi ad altri dati, arricchendo il dato di partenza. Tale gruppo non è il solo che opera in tal senso ve ne sono degli altri quali *Multilingual Europe Technology Alliance* (META-NET),¹⁶⁹ *Fostering Language Resources Network* (FLaReNet)¹⁷⁰, ecc..

¹⁶⁷ MAURO GUERRINI, *La biblioteca spiegata agli studenti universitari*, Milano, Editrice Bibliografica, 2012 (Collana: Conoscere la biblioteca, 8). ISBN/EAN 9788870757231.

¹⁶⁸ <<http://linguistics.okfn.org/resources/llod/>> (ultima consultazione 16/12/2014).

¹⁶⁹ È un progetto che nasce con l'idea di valorizzare il multilinguismo peculiare dell'Unione Europea attraverso la definizione dei presupposti per l'arricchimento di una società in cui una molteplicità di lingue convivono, interagiscono e si influenzano vicendevolmente. <<http://www.meta-net.eu/>> (ultima consultazione 16/12/2014).

¹⁷⁰ È un progetto inerente alla definizione di una politica comune in merito alla realizzazione di risorse linguistiche e di tecnologie linguistiche. Gli obiettivi riguardano la promozione di una strategia condivisa attraverso l'individuazione delle priorità a breve e medio termine e delle strategie da adottare a lungo termine per favorire la competitività del settore sia in ambito comunitario che internazionale. <<http://flarenet.eu/>> (ultima consultazione 16/12/2014).

CAPITOLO 3

ANALISI DELLO STATO DELL'ARTE

La diffusione dei principi e delle soluzioni proposte dal web semantico unitamente alle strategie di condivisione e riutilizzo previste dall'approccio linked data, hanno stimolato l'avvio di numerose iniziative, soprattutto in ambito comunitario e internazionale concretizzatesi nella messa a punto di sistemi di organizzazione e gestione della conoscenza in formato SKOS/RDF, tramite il quale è possibile definire vocabolari *ex novo* oppure convertire vocabolari già esistenti in un formato tale da assicurare una maggiore condivisione e interoperabilità sia sul piano tecnologico che su quello semantico.

I sistemi di organizzazione della conoscenza, in qualità di strumenti di controllo terminologico a supporto dell'indicizzazione e del recupero dell'informazione, per primi sono chiamati a mettere a disposizione della collettività il loro patrimonio lessicale e la struttura concettuale più o meno sofisticata che ne è alla base, di modo da poter essere riutilizzati da terze parti per definire nuovi strumenti, arricchire quelli già esistenti e sviluppare servizi che integrino i dati prodotti dai diversi provider.

La più nota iniziativa avviata nel 2004 sul tema dell'allineamento di vocabolari è l'*Ontology Alignment Evaluation Initiative* (OAEI),¹⁷¹ un vero e proprio punto di riferimento per quanti si accostano allo studio dei diversi sistemi e delle soluzioni predisposti per l'allineamento e la mappatura tra fonti eterogenee da far confluire in un modello comune, che può essere interrogato e sottoposto a processi di ragionamento automatico.¹⁷²

Ulteriori esperienze, seppur di portata inferiore, hanno riguardato l'allineamento di vocabolari al fine della loro pubblicazione secondo i principi del linked data.

Il presente capitolo si propone di: delineare lo stato dell'arte dei più significativi sistemi di organizzazione della conoscenza attualmente disponibili in SKOS/RDF ed arricchiti

¹⁷¹ <<http://oaei.ontologymatching.org>> (ultima consultazione 16/12/2014).

¹⁷² JÉRÔME EUZENAT, *Ontology Alignment Evaluation Initiative: Six Years of Experience*, «Journal on Data Semantics XV», Berlin, Springer Verlag, 2011 (Collana: Lecture Notes in Computer Science, 6720), p. 158-192, DOI: 10.1007/978-3-642-22630-4_6, <<http://disi.unitn.it/~p2p/RelatedWork/Matching/OAEI-JODS.pdf>> (ultima consultazione 16/12/2014).

di correlazioni semantiche ad altri vocabolari; presentare taluni studi che si sono occupati di problematiche inerenti all'allineamento come prerequisito per la condivisione e lo scambio dei dati; illustrare i principali sistemi di organizzazione della conoscenza afferenti al dominio delle costruzioni.

Tale ricognizione ha l'obiettivo di mettere in evidenza innanzitutto, come il settore delle costruzioni, nel contesto italiano, non sia adeguatamente supportato dalla presenza di risorse terminologiche e semantiche di dominio e di come quelle esistenti non prevedano alcuna forma di allineamento rispetto a concetti espressi in altri vocabolari.

1. I KOS disponibili in modalità Linked Data

Il panorama nazionale e internazionale dei sistemi di organizzazione della conoscenza pubblicati in linked data, mostra come l'ultimo decennio sia stato caratterizzato da un numero sempre maggiore di iniziative e progetti volti alla predisposizione di vocabolari (sistemi di classificazione, thesauri, liste controllate, ecc..) in RDF e al loro successivo allineamento rispetto a vocabolari esterni non necessariamente afferenti al medesimo dominio.

Di seguito è proposta una rassegna non esaustiva dei KOS attualmente pubblicati in linked data, i quali si differenziano: per il dominio applicativo, per la consistenza, per la copertura semantica, per gli scopi cui sono destinati.

I KOS presi in considerazione non si riferiscono direttamente al settore delle costruzioni. Tuttavia, dalla loro analisi è stato possibile trarre importanti spunti di riflessione per la predisposizione dell'approccio metodologico di supporto alla definizione di correlazioni semantiche tra concetti provenienti da fonti differenti a cui sarà dedicato il capitolo successivo.

Nuovo Soggettario

Il Nuovo Soggettario¹⁷³ è in ambito nazionale il progetto più interessante avviato in tal senso. Si tratta di un thesaurus multidisciplinare impiegato nell'indicizzazione per soggetto di risorse bibliografiche disponibili su supporto cartaceo e digitale (testi antichi, libri, riviste, audiovisivi, materiale fotografico, ecc.). Il Nuovo Soggettario è stato

¹⁷³ <<http://thes.bncf.firenze.sbn.it/ricerca.php>> (ultima consultazione 16/12/2014).

sviluppato dalla Biblioteca Nazionale Centrale di Firenze (BNCF) nel 1956, la quale tuttora ne detiene la gestione e il mantenimento, è largamente impiegato presso le biblioteche pubbliche e private e in altre istituzioni e/o organizzazioni che aderiscono al Servizio Bibliotecario Nazionale (SBN), la rete bibliotecaria italiana nata con lo scopo di creare un catalogo unico, l'OPAC¹⁷⁴ SBN.¹⁷⁵ Questo thesaurus rappresenta per il panorama italiano, lo strumento di indicizzazione più autorevole in quanto garantisce il controllo terminologico e propone una rappresentazione strutturata della conoscenza trasversale ai diversi ambiti del sapere.

I suoi punti di forza risiedono nella visione multidisciplinare della conoscenza e nella logica di tipo analitico-sintetica che ne è alla base. Dati questi presupposti tale strumento garantisce una copertura semantica generale che agevola il processo di identificazione e rappresentazione del contenuto concettuale dei documenti per mezzo di un soggetto, e la successiva collocazione e organizzazione del soggetto sulla scorta dell'analisi dei singoli elementi che lo costituiscono.

Il Nuovo Soggettario, mette in comunicazione ambiti del sapere che apparentemente non presentano punti di contatto, attraverso la predisposizione di una macro-struttura inclusiva, che prevede i seguenti livelli:

- categoria: macro-raggruppamento in cui si articolano le quattro faccette principali: *Agenti, Azioni, Cose e Tempo*;
- faccetta: classe generale all'interno della quale è ricondotto l'insieme di termini che presentano tratti ed elementi comuni, ad esempio per la categoria *Azioni* le faccette predisposte sono *Attività, Discipline e Processi*;
- sotto-faccetta: classe più specifica rispetto alla faccetta, che può essere espressa per mezzo di un termine preferito o da un'etichetta di snodo, ad esempio la faccetta *Attività* prevedete le sotto-faccette *Arte* (come termine preferito), [Attività comportamentali e personali] (come etichetta di snodo), ecc..¹⁷⁶

Questa struttura è stata predisposta a partire dal 2007 con l'intento di rendere il Nuovo Soggettario uno strumento:

¹⁷⁴ OPAC è l'acronimo tramite il quale si designa l'*On-line Public Access Catalogue* ovvero il catalogo informatizzato e liberamente accessibile, tramite il quale è possibile interrogare simultaneamente l'insieme dei cataloghi che vi afferiscono. <<http://www.sbn.it/opacsbn/opac/iccu/informazioni.jsp>> (ultima consultazione 16/12/2014).

¹⁷⁵ L'OPAC SBN è nato per volontà dell'allora Ministero per i beni culturali e ambientali (Mbca), oggi Ministero dei Beni e delle Attività Culturali e del Turismo (MiBAC) e gestito dall'Istituto Istituito Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche (ICCU).

¹⁷⁶ Biblioteca Nazionale Centrale di Firenze, *Nuovo soggettario. Guida la sistema italiano di indicizzazione per soggetto. Prototipo del thesaurus*, Ministero dei Beni e delle Attività Culturali e del Turismo, Direzione generale per i beni librari e gli istituti culturali.

- ❖ adeguato per l'indicizzazione e il recupero dell'informazione;
- ❖ interoperabile grazie alla tecnologia adottata;
- ❖ intenzionato ad arricchirsi di terminologia e conoscenza conservata in altri sistemi di organizzazione della conoscenza.

L'identità di questo strumento, la sua autorevolezza e l'ampio utilizzo nel contesto italiano soprattutto da parte di biblioteche, archivi e musei hanno portato nel 2010 alla sua pubblicazione in RDF/XML garantendo, di fatto, il colloquio con altri vocabolari non necessariamente riguardanti l'ambito bibliotecario.¹⁷⁷

Il Nuovo Soggettario è conforme alla più recente normativa in materia di thesauri e interoperabilità con altri vocabolari (ISO 25964 *Information and Documentation – Thesauri and interoperability with other vocabularies*) e predispone per ciascuna voce del thesaurus una scheda descrittiva dalla quale si evincono le seguenti informazioni:

- la categoria e la faccetta di appartenenza di un dato concetto, presentate una di seguito all'altra, ad esempio *Cose:Strutture*;
- la nota d'ambito (scope note o SN): utilizzata per definire i confini semantici di un dato concetto; per specificarlo soprattutto qualora il termine possa veicolare più significati, ad esempio, SN *Complessi delle strutture portanti che completano la sommità di un fabbricato*;
- la struttura sistematizzata vera e propria:
 - Top Term (TT): specifica il termine di testa dalla quale parte la suddivisione gerarchica, ad esempio TT Strutture;
 - Used for (UF): rinvia da un termine preferito ad un termine non preferito, ad esempio Canali UF Canali artificiali;
 - Use (USE): rinvia da un termine non preferito ad un termine preferito, ad esempio Canali artificiali (USE) Canali;
 - Broader Term (BT): rinvia da un concetto più specifico ad un termine più generico, ad esempio: Coperture BT Elementi architettonici;
 - Narrower Term (NT): rinvia da un concetto più generale ad un termine più specifico, ad esempio Elementi architettonici NT Coperture;
 - Related Term (RT): associa concetti per i quali non è possibile stabilire relazioni di equivalenza o gerarchiche, ad esempio Coperture RT Capriate;

¹⁷⁷ «cfr.» ANNA LUCARELLI, *Esperienze di linked data e web semantico in BNCF*, 6' Incontro ISKO Italia Firenze 20 maggio 2013, <<http://www.iskoi.org/doc/firenze13/lucarelli.pdf>> (ultima consultazione 16/12/2014).

- l'agenzia catalografica/proponente, nella fattispecie la Bibliografia Nazionale Italiana (BNI);¹⁷⁸
- lo status del record che specifica il grado di lavorazione di un termine, ad esempio Termine strutturato, BASE e MIN;¹⁷⁹
- la fonte in cui il termine è attestato, ad esempio il Dizionario ragionato di termini tecnici dell'edilizia,¹⁸⁰ Wikipedia, AGROVOC, *Library of Congress Subject Headings* (LCSH) ecc., verso i quali sono esplicitati collegamenti basati sul principio della corrispondenza esatta.

The screenshot displays the 'Murature' entry in the Nuovo Soggettario. On the left, a hierarchical tree shows the classification: **TT** Attività, **BT** [Attività tecniche relative alle strutture], **RT** Agglomeranti cementizi, Arcate in muratura, Camini in muratura, Case, Edifici in muratura, Edilizia, Materiali da costruzione, Mattoni, Muri, and Volte in muratura. Below this, a red box highlights the source information: 'Fonte: Soggettario, VLI, DeM, Dev: [Wikipedia\(IT\)](#)', 'Equiv. LCSH: [Masonry](#)', 'Agenzia catalografica/Proponente: BNI', 'Status del record: Termine strutturato', and 'Identificativo: 23872'. On the right, an RDF/XML snippet is shown, detailing the SKOS properties and external links, including a 'closeMatch' to the LCSH term 'Masonry'.

Figura 15. RDF/XML e collegamenti semantici esterni nel Nuovo Soggettario

Il Nuovo Soggettario è uno strumento in divenire, poiché la base terminologica viene costantemente incrementata e la sua struttura concettuale si specializza e si arricchisce di nuove relazioni. A questa forma di ampliamento interno, si affianca una nuova modalità di accrescimento del thesaurus, che prevede l'allineamento a risorse terminologiche e semantiche esterne, per mezzo della specificazione di correlazioni concettuali introdotte dai metadati predisposti da SKOS. Inoltre, il Nuovo Soggettario prevede accessi multilingue, che si esplicitano a partire dalla predisposizione di

¹⁷⁸ <<http://www.bncf.firenze.sbn.it/pagina.php?id=187>> (ultima consultazione 16/12/2014).

¹⁷⁹ Per la descrizione dettagliata dei tre status si rinvia alla pagina *Sigle e simboli* consultabile al seguente indirizzo web: <<http://thes.bncf.firenze.sbn.it/sigle.php?menuR=1&menuS=2>> (ultima consultazione 16/12/2014).

¹⁸⁰ «cfr.» VINCENZO BACIGALUPI, *Dizionario ragionato di termini tecnici dell'edilizia*, Roma, Kappa, 1998. ISBN 8878903108.

collegamenti con il *Library of Congress Subject Headings* (LCSH), il *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU), ecc..

La Figura 15 mostra i campi coinvolti nel processo di allineamento: 1) la fonte in cui un dato termine è attestato; 2) l'equivalenza (*Equiv*) specificata verso i corrispettivi concetti espressi in altre lingue. In aggiunta mette in evidenza la natura dei collegamenti verso i vocabolari esterni, i quali sono introdotti a partire dal tag SKOS <skos:closeMatch> concettualmente più ampio di <skos:exactMatch>

La multidisciplinarietà che contraddistingue il Nuovo Soggettario, la sua disponibilità in RDF/XML congiuntamente al suo rilascio con licenza *Creative Commons* Attribuzione 2.5, che ne prevede l'ampia utilizzazione purché ne sia riconosciuta e citata la paternità,¹⁸¹ rappresentano i motivi per cui il Nuovo Soggettario è stato scelto ed utilizzato nell'ambito della presente ricerca, come strumento verso cui indirizzare le correlazioni semantiche finalizzate alla definizione di uno schema di corrispondenze.

Il Nuovo Soggettario arricchendosi ulteriormente tramite la specificazione di collegamenti esterni verso altri KOS che a loro volta si collegano ad altri dataset, contribuisce alla realizzazione della ragnatela di dati interoperabili.

AGROVOC

AGROVOC¹⁸² è il thesaurus multilingue riguardante i settori dell'agricoltura, della pesca, della sicurezza alimentare. È pubblicato dal *Food and Agriculture Organization* (FAO) delle Nazioni Unite ed è curato da una comunità di esperti ed organizzazioni internazionali che operano in domini affini a quelli dell'organizzazione.¹⁸³ AGROVOC è utilizzato in tutto il mondo da parte di ricercatori, bibliotecari, gestori di informazione e altri, per l'indicizzazione, il recupero, e l'organizzazione dei dati in sistemi informativi e pagine web.¹⁸⁴

La prima versione, è stata resa disponibile su supporto cartaceo nel 1980 in tre lingue: inglese, francese e spagnolo, nel 2004 è stato trasferito su un *Database Management System* (DMS), un sistema software per l'organizzazione, la gestione e l'interrogazione

¹⁸¹ Per i dettagli in merito a questa tipologia di licenza si rinvia alla consultazione alla pagina web di seguito riportata: <<https://creativecommons.org/licenses/by/2.5/it/>> (ultima consultazione 16/12/2014).

¹⁸² <<http://aims.fao.org/>> (ultima consultazione 16/12/2014).

¹⁸³ MARINA CARACCILO, et alii, *Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case*, «Int. J. Metadata, Semantics and Ontologies», vol. 7, n. 1, 2012, <http://eprints.rclis.org/17735/1/IJMSO%207_1_Paper%206_PublishedVersion%5B1%5D.pdf> (ultima consultazione 16/12/2014).

¹⁸⁴ <<http://aims.fao.org/agrovoc#.VICN3sIIT1U>> (ultima consultazione 16/12/2014).

di dati strutturati e nel 2009 in seguito all'avvento delle tecnologie del web semantico, la sua struttura relazionale è stata tradotta in *Ontology Web Language Description Logic* (OWL DL),¹⁸⁵ ritenuto più adatto per la massima espressività e specificità del sistema.¹⁸⁶

Nel 2009 AGROVOC è stato convertito in *RDF/Simple Knowledge Organization System eXtension for Labels* (SKOS-XL)¹⁸⁷ e quindi pubblicato come *linked open dataset*. Attualmente è disponibile in 21 lingue ed è stato mappato verso circa tredici vocabolari, alcuni dei quali di carattere generale come DBpedia, la CDD;¹⁸⁸ altri inerenti a specifici domini come il *National Agricultural Library's Agricultural Thesaurus* (NAL Thesaurus)¹⁸⁹ riferito all'agricoltura, altri ancora afferenti a domini diversi con i quali sussiste un'implicazione reciproca, ad esempio il *Thesaurus for the Social Sciences*¹⁹⁰ e il *Thesaurus for Economics*.¹⁹¹

La struttura concettuale di AGROVOC si basa su tre elementi fondamentali:

- i concetti: le unità di pensiero, le idee astratte generate da un processo cognitivo, identificati a loro volta da un URI;
- i termini: le forme lessicali attraverso cui si esprimono i concetti, specificati per ciascuna lingua;
- le varianti: l'insieme delle forme che è possibile impiegare per esprimere i concetti, ad esempio le varianti morfologiche, le abbreviazioni, le sigle, gli acronimi, ecc..

I concetti in AGROVOC sono identificati univocamente a partire dal prefisso URI <http://aims.fao.org/aos/agrovoc/>, per cui ad esempio l'URI http://aims.fao.org/aos/agrovoc/c_15903 identifica e distingue tra gli altri il concetto di *pesce*.

La Figura 16 dà evidenza della struttura concettuale di un termine e diversamente da quanto avviene nel Nuovo Soggettario, le relazioni non sono specificate ricorrendo

¹⁸⁵ OWL DL è un sotto-linguaggio di OWL che garantisce la massima espressività dei sistemi di *reasoning*. *OWL Web Ontology Language Guide*, <<http://www.w3.org/TR/owl-guide/>> (ultima consultazione 16/12/2014).

¹⁸⁶ <<http://aims.fao.org/standards/agrovoc/concept-scheme#.VINRMXvg-wk>> (ultima consultazione 16/12/2014).

¹⁸⁷ Estensione di SKOS che fornisce un ulteriore supporto nella descrizione e nel collegamento di entità lessicali. ALISTAIR MILES, SEAN BECHHOFFER, *SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant*, 2009. <<http://www.w3.org/TR/skos-reference/skos-xl.html>>, data ultima consultazione 16/12/2014).

¹⁸⁸ <<http://dewey.info/>> (ultima consultazione 16/12/2014).

¹⁸⁹ <<http://agclass.nal.usda.gov/>> (ultima consultazione 16/12/2014).

¹⁹⁰ <<http://www.gesis.org/en/services/research/thesauri-und-klassifikationen/social-science-thesaurus/>> (ultima consultazione 16/12/2014).

¹⁹¹ <<http://zbw.eu/stw/versions/latest/about>> (ultima consultazione 16/12/2014).

all'ausilio delle canoniche sigle impiegate per la rappresentazione della struttura concettuale di un termine, ma dalla loro specificazione per esteso, ad esempio *broader concept* invece di *BT* e *alternative label* al posto di *UF*.

AGROVOC propone una differenziazione netta tra il piano concettuale e quello dell'espressione terminologica: tra i concetti sono stabilite correlazioni gerarchiche e/o associative mentre le relazioni espresse tra un concetto e la rispettiva variante morfologica e/o equivalente in una o più lingue interessano più che altro il piano puramente terminologico.

PREFERRED TERM	fish
CONCEPT TYPE	Concept
BROADER CONCEPT	animal products Fishery products
ALTERNATIVE LABEL	fish meat fresh fish wet fish
SCOPE NOTE	Use for fish as food; for the animal use <2943>
IS USED AS	foods
IN OTHER LANGUAGES	Poisson (aliment) (fr) Ikan (ms) ปลา (นลินี) (th) ငါး (te) Pescado (es)
URI	http://aims.fao.org/aos/agrovoc/c_15903
Download this concept:	RDF/XML TURTLE
CLOSELY MATCHING CONCEPT	魚類 (ja)
EXACT MATCH	Fische (de) STW Thesaurus for Economics http://d-nb.info/gnd/4272625-6 http://eurovoc.europa.eu/2435 http://lod.nal.usda.gov

Figura 16. Esempificazione della struttura di un concetto in AGROVOC ed esplicitazione di link esterni.

AGROVOC come summenzionato è collegato a circa 13 vocabolari di varia natura anch'essi disponibili in SKOS/RDF. L'approccio adottato per l'allineamento ai vocabolari esterni ha previsto diverse fasi descritte in (Morshed, et. alii, 2011):¹⁹²

- i vocabolari coinvolti nel processo di *mapping* sono stati anzitutto collocati all'interno di un *triple store* e sono stati tutti considerati nella loro interezza ad eccezione del repertorio RAMEAU;
- il *matching* tra concetti è stato definito in modo automatico a partire da coppie di concetti preferiti: concetto AGROVOC e concetto appartenente ad un altro

¹⁹² MORSHED ALAN, et. alii, *Thesaurus alignment for Linked Data publishing, Proceedings International Conference on Dublin Core and Metadata Applications*, The Hague, the Netherlands, 21-23 September, 2011, p. 65-75. DOI: 10.1504/IJMSO.2012.048511; ISSN: 1744-2621.

vocabolario. Per valutare il grado di corrispondenza tra coppie di concetti sono state impiegate diverse misure di similarità, ad esempio la distanza di Hamming,¹⁹³ ecc..

- verifica e validazione da parte degli esperti di dominio dei *matching* definiti in maniera automatica.

Al termine di questo processo i dati accompagnati dalle loro correlazioni concettuali sono resi in formato *machine-accessible* e consultabili attraverso un *endpoint* SPARQL e resi disponibili agli utenti tramite una pagine HTML generata con Pubby.¹⁹⁴

http://aims.fao.org/aos/agrovoc/c_15903			fish		
Property	Value	pref.label	alt.label	Lang	
rdfs:type	skos:Concept	fish	fresh fish	en	
skos:broader	http://aims.fao.org/aos/agrovoc/c_2941 http://aims.fao.org/aos/agrovoc/c_438	fish	fish meat		
skos:exactMatch	http://d-nb.info/gnd/4272625-6 http://lod.nal.usda.gov/nalt/845 http://eurovoc.europa.eu/2435 http://zbw.eu/stw/descriptor/14158-2	Pesce (alimento)	wet fish		
skos:closeMatch	http://dbpedia.org/resource/Fish	Pesce in salamoia	Pesce fresco	it	
skos:broadMatch	http://www.caas.net.cn/caas/cat/c_57994 http://www.caas.net.cn/caas/cat/c_21812 http://www.caas.net.cn/caas/cat/c_53035 http://www.caas.net.cn/caas/cat/c_58511	鱼肉	Carne di pesce		
dcterms:created	2012-02-03T16:43:28Z	魚	鲜鱼	zh	
dcterms:modified	2013-08-27T22:12:24Z	ryba (mäso)	Čerstvé ryby	sk	
skos:scopeNote	Para o animal usar <2943> <2943> Hal, mint táplálék; Hal, mint állat: <2943> Per gli animali usare <2943> Używać dla ryb jako pożywienia; dla ryb jako zwierząt użyć <2943> <2943> ใช้สำหรับปลาใช้เป็นอาหาร, สำหรับปลาใช้เป็นสัตว์กินไม่ได้ <2943> Gıda olarak balık için kullan; hayvan için <2943> kodlu terimi kullan Pro rybu jako potraviny; pro zvíře USE RYBY Use for fish as food; for the animal use <2943> ใช้สำหรับปลาเป็นอาหาร; สำหรับสัตว์ ใช้ <2943>	rybie mäso	Chladené ryby (v fade)		
		Poisson (aliment)	Poisson frais	fr	
		魚	鮮魚	ja	
		Carne de peixe	Peixe fresco	pt	
		Mięso ryb	Ryby (mięso)	pl	
		puša	Ryba świeża		
			Ryba surowa		
			мясо рыбы	ru	
			свежая рыба		
			сырая рыба		

Figura 17. I *mapping* generati.

EUROVOC

EuroVoc¹⁹⁵ è un thesaurus multilingue e multidisciplinare che abbraccia le tematiche di interesse dell'Unione Europea ed è disponibile in diverse lingue.¹⁹⁶ È nato con l'intento di organizzare e gestire l'informazione prodotta da questa istituzione, ma è un valido strumento di controllo che si adatta all'uso in altri contesti, soprattutto in ambito parlamentare.

¹⁹³“the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. Put another way, it measures the minimum number of substitutions required to change one string into the other, or the number of errors that transformed one string into the other.”. <https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Hamming_distance.html> (ultima consultazione 16/12/2014).

¹⁹⁴ Applicazione web realizzata in Java per la trasformazione di SPARQL endpoint in Linked Data Server. <<http://wifo5-03.informatik.uni-mannheim.de/pubby/>> (ultima consultazione 16/12/2014).

¹⁹⁵<<http://eurovoc.europa.eu/>> (ultima consultazione 16/12/2014).

¹⁹⁶ Bulgaro, ceco, croato, danese, estone, finlandese, francese, greco, inglese, italiano, lettone, lituano, maltese, neerlandese, polacco, portoghese, rumeno, slovacco, sloveno, spagnolo, svedese, tedesco, ungherese e serbo.

Eurovoc è il risultato di uno studio promosso nel 1982 dal Parlamento europeo e dalla Commissione europea riguardante l'analisi e il confronto tra i linguaggi di indicizzazione adottati dalle varie istituzioni per organizzare e gestire le proprie collezioni documentali. La prima edizione è stata pubblicata nel 1984, in sette lingue¹⁹⁷ e si componeva di un thesaurus tematico ed di un thesaurus alfabetico. Alla prima versione ne sono seguite sette, che hanno previsto il progressivo arricchimento e approfondimento della struttura, attraverso: l'aggiunta di nuovi descrittori in particolare quelli riferiti alle entità geografiche dei nuovi paesi che entrano a far parte dell'Unione Europea; la migliore precisazione delle equivalenze tra le diverse lingue; la disponibilità in tutte le lingue ufficiali dell'Unione europea.¹⁹⁸ La versione più aggiornata di Eurovoc è l'edizione 4.3 del 2008 la quale è stata interessata dall'aggiunta di nuove entità geografiche e ha interessato la revisione dei settori riguardanti la geografia e le organizzazioni internazionali

Il thesaurus EuroVoc è stato nel corso del tempo adottato da un numero sempre crescente di utenti, è infatti stimato che esso non sia esclusivamente in uso presso le sole istituzioni comunitarie ma è ampiamente impiegato da biblioteche e centri di documentazione sia a livello comunitario che internazionale. La sua larga diffusione è dovuta a diversi ordini di motivi: è un linguaggio di indicizzazione controllato per cui valido sia per l'indicizzazione del contenuto dei documenti che per il loro successivo recupero; è multilingue, ragion per cui consente di creare punti di accesso differenziati al medesimo documento; si propone di creare una rete di collaborazione operativa tra coloro che si occupano delle medesime tematiche di modo da evitare la ridondanza del lavoro svolto.¹⁹⁹

EuroVoc è stato sviluppato conformemente agli standard internazionali in materia di sviluppo di thesauri, è curato dall'Ufficio delle Pubblicazioni dell'Unione Europea²⁰⁰ ed è attualmente disponibile in formato SKOS / RDF.

Il thesaurus è in continuo aggiornamento in considerazione della sua stretta correlazione con le attività condotte nel quadro comunitario e delle osservazioni costanti provenienti dai diversi utenti. Tali rivisitazioni si concretizzano nell'introduzione di nuovi descrittori, nell'eliminazione di taluni descrittori ritenuti obsoleti e nell'aggiornamento della struttura classificatoria.

¹⁹⁷ Danese, tedesco, greco, inglese, francese, italiano e olandese. <<http://eurovoc.europa.eu/drupal/?q=it/node/414&cl=it>> (ultima consultazione 16/12/2014).

¹⁹⁸ <<http://eurovoc.europa.eu/drupal/?q=it/node/399>> (ultima consultazione 16/12/2014).

¹⁹⁹ Comunità Europee, *Thesaurus EUROVOC Volume 3: Presentazione multilingue*, Lussemburgo: Ufficio delle pubblicazioni ufficiali delle Comunità europee, 1995.

²⁰⁰ È un organismo interistituzionale che si occupa della pubblicazione giornaliera della Gazzetta Ufficiale dell'Unione europea e dell'edizione delle pubblicazioni realizzate nell'ambito delle attività condotte dalle diverse istituzioni. <http://publications.europa.eu/index_it.htm> (ultima consultazione 16/12/2014).

Dal punto di vista della struttura EuroVoc prevede due livelli di dettaglio:

- i settori: classi generali identificate da un numero e da una notazione esplicativa, ad esempio 04 Vita politica, 66 Energia, 52 Ambiente, ecc.;
- i microthesauri: raggruppamenti di termini inerenti a un sotto-dominio della classe di apparenza. I microthesauri sono introdotti da quattro cifre, di cui le prime due richiamano il settore in cui si colloca il micro-thesaurus, ad esempio 0406 quadro politico, 6606 politica energetica, 5206 politica dell'ambiente, ecc.. e la relazione di appartenenza al microthesaurus è introdotta dalla sigla MT (microthesaurus), ad esempio *filosofia politica* MT 0406 *quadro politico*.

Questi due livelli assicurano l'univocità di un dato concetto e limitano di conseguenza le problematiche inerenti all'ambiguità semantica nel caso in cui lo stesso termine sia portatore di più significati, pertanto, l'uso e il valore del concetto viene circoscritto ad un dato settore e di conseguenza al relativo microthesaurus.

Inoltre, EuroVoc conformemente alla logica dei thesauri specifica al suo interno relazioni di equivalenza stabilita tra sinonimi, quasi-sinonimi ed esplicitate dalle sigle USE e UF, relazioni gerarchiche introdotte da BT e NT e relazioni associative. Ciascun concetto di EuroVoc può essere a sua volta accompagnato da: una nota operativa che descrive l'uso del termine nel thesaurus SN (*scope note*); una nota di definizione che specifica il senso del concetto DEF; una nota storica che traccia i mutamenti che il concetto ha subito all'interno del thesaurus HN (*history note*); un'ulteriore spiegazione qualora vengano utilizzati come descrittori termini omografi, ad esempio *Pesca* (frutto) rispetto a *Pesca* (attività). Inoltre EuroVoc, prevede limitatamente al settore della geografia, che il medesimo concetto possa essere assegnato a più categorie (relazione di poligerarchia), ad esempio Iraq può appartenere al microthesaurus 7226 Asia-Oceania e alla categoria Vicino e Medio Oriente e contemporaneamente al microthesaurus 7231 geografia economica e alle categorie paesi del CUEA, paesi dell'OPEC, paesi membri del Mercato comune arabo.

Come anticipato EuroVoc è disponibile in RDF/SKOS per il quale è stato predisposto un apposito modello ontologico in cui è stata codificata l'intera struttura del thesaurus.

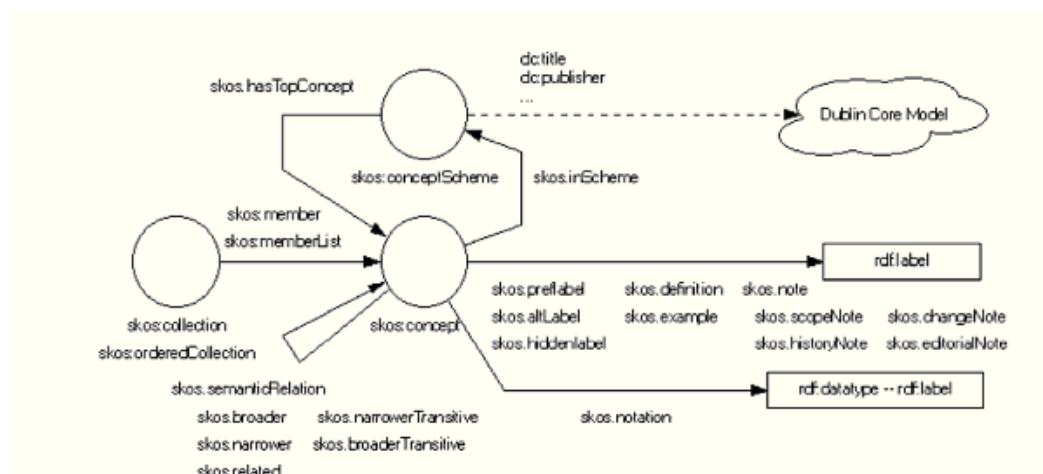


Figura 18. Il modello ontologico SKOS di EuroVoc.²⁰¹

Infine, il thesaurus EuroVoc è allineato ad altri vocabolari quali AGROVOC, il Sistema Automatizzato della Biblioteca della Commissione Europea (Eclas),²⁰² il *General Multilingual Environmental Thesaurus* (Gemet) e il *United Nations Bibliographical Information System* (Unibis).²⁰³ I collegamenti tra EuroVoc e tali risorse sono state specificate ricorrendo al tag <skos:exactMatch> e sono disponibili sul Portale Open Data dell'Unione europea.²⁰⁴

General Multilingual Environmental Thesaurus (GEMET)

GEMET²⁰⁵ è un thesaurus multilingue relativo alle tematiche ambientali, è stato realizzato per rispondere a specifiche esigenze espresse dalla *European Environment Agency* (AEA)²⁰⁶ nell'ambito del progetto *European Topic Centre on Catalogue of Data*

²⁰¹ JAVIER LACASTA, et alii, *Terminological Ontologies: Design, Management and Practical Applications*, «Semantic Web and Beyond», vol. 9, Springer, 2010, p. 34. ISBN-13: 9781461426448 - ISBN-10: 1461426448

²⁰² È il catalogo elettronico della biblioteca centrale della Comunità europea in cui sono presenti i riferimenti bibliografici alle pubblicazioni e ai documenti sviluppati nel quadro comunitario. <http://ec.europa.eu/libraries/index_en.htm> (ultima consultazione 16/12/2014).

²⁰³ Un thesaurus multilingua che contiene la terminologia inerente alle tematiche trattate dalle Nazioni Unite ed è stato sviluppato dalla Biblioteca americana Dag Hammarskjöld e dal Dipartimento della Pubblica Informazione delle Nazioni Unite. <[https://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf/\\$\\$search?OpenForm](https://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf/$$search?OpenForm)> (ultima consultazione 16/12/2014).

²⁰⁴ <<http://open-data.europa.eu/en/data/dataset/eurovoc>> (ultima consultazione 16/12/2014).

²⁰⁵ <<http://www.eionet.europa.eu/gemet/en/about/>> (ultima consultazione 16/12/2014).

²⁰⁶ In italiano Agenzia europea dell'ambiente il cui compito è diffondere e divulgare informazioni in merito a tematiche ambientali. <<http://www.eea.europa.eu/>> (ultima consultazione 16/12/2014).

Sources (ETC/CDS)²⁰⁷ ed è reso disponibile a partire dalla rete *European Environment Information and Observation Network* (EIONet).²⁰⁸

GEMET è il risultato di un processo di fusione (tecnicamente *merge*) e di armonizzazione della terminologia e della struttura concettuale impiegate in altrettanti thesauri multilingue, ritenuti autorevoli e quindi affidabili,²⁰⁹ al fine di sviluppare una terminologia comune per il dominio dell'ambiente. I termini presenti al suo interno sono organizzati secondo: un ordine categorico-gerarchico, articolato in 3 super-gruppi e a sua volta suddiviso in 30 gruppi, che conferisce coerenza e uniformità alla struttura sistematizzata; un ordine tematico espresso per mezzo di circa 40 temi che rendono i termini maggiormente fruibili.²¹⁰

In GEMET ciascuna entrata del thesaurus è accompagnata da una definizione e dalla specificazione degli equivalenti in talune lingue ufficiali dell'Unione Europea²¹¹ e in altre lingue,²¹² ciò al fine di tenere traccia e monitorare il plurilinguismo proprio dell'Unione europea favorendo di conseguenza la comunicazione e lo scambio delle informazioni tra parlanti di lingue diverse. In aggiunta a questi elementi GEMET prevede l'indicazione dell'ordine tematico e del gruppo di appartenenza nonché l'introduzione delle classiche relazioni thesaurali: gerarchiche (del tipo genere-specie e parte-tutto), di equivalenza e associative, queste ultime di solito specificate in relazione alla disciplina di studio a cui afferiscono.

Anche GEMET come i KOS precedentemente descritti, è attualmente disponibile in SKOS/RDF e segue i principi linked data. Di fatto, al suo interno ciascun concetto è identificato univocamente per mezzo di un URI il cui prefisso è <http://www.eionet.europa.eu/gemet/>.

²⁰⁷ Il progetto si propone la messa a punto di un set di metadati riguardanti l'ambiente al fine di supportare lo sviluppo di catalogo di metadati univoco e di una terminologia altrettanto precisa e chiara. *Internet for Environmental Communication, Workshop Papers* - Vienna, 1996, p. 1, <<http://www.umweltbundesamt.at/fileadmin/site/publikationen/CP017.pdf>> (ultima consultazione 16/12/2014).

²⁰⁸ È una rete costituita a livello europeo di cui fanno parte la *European Environment Agency* (AEA), i Centri Tematici Europei ovvero gruppi di esperti specializzati in tematiche ambientali che operano a supporto dei piani strategici definiti a livello AEA e altri esperti provenienti da diversi parti del mondo. <<http://www.eionet.europa.eu/>> (ultima consultazione 16/12/2014).

²⁰⁹ I termini presenti all'interno di GEMET provengono dai seguenti thesauri: *Umwelt Thesaurus dell'Umweltbundesamt* (UBA); il *Thesaurus Italiano per l'Ambiente* (TIA); il *Multilingual Environment Thesaurus* (MET); l'*EnVoc Thesaurus*; *Thesaurus de Medio Ambiente*; *Lexique environnement -Planète*; i descrittori presenti nel *Europe's Environment, The Dobris Assessment*; il *Thesaurus Eurovoc*. <<http://www.eionet.europa.eu/oldgemet/about?langcode=de>> (ultima consultazione 16/12/2014)..

²¹⁰ EDOARDO CHITI, RICCARDO GULDO, *Il regime linguistico dei sistemi comuni europei*, numero monografico della «Rivista di Diritto Pubblico», Milano, Giuffrè, 2008, pp. 57-58. ISBN: 88-14-14191-6 - EAN: 9788814141911

²¹¹ Le lingue codificate da GEMET sono :bulgaro, ceco, danese, estone, finlandese, francese, greco, italiano, olandese, polacco, portoghese, slovacco, sloveno, spagnolo, svedese, tedesco, ungherese; al contrario non esistono traduzioni in rumeno, maltese, irlandese, lettone e lituano.

²¹² Le altre lingue in cui GEMET è predisposto sono: basco, inglese americano, norvegese, cinese, russo, ecc.

In questo modo l'arricchimento del thesaurus con le traduzioni provenienti da altre lingue ne trae beneficio, il concetto così come la sua struttura organizzativa rimane immutato in quanto le singole traduzioni sono introdotte come semplici etichette. GEMET è inoltre, interoperabile rispetto ad altri sistemi di organizzazione e gestione della conoscenza quali *UMTHES*²¹³ *Umwelt-Thesaurus dell'Umweltbundesamtes Dessau*,²¹³ DBpedia, e AGROVOC, EuroVoc, ecc., verso i quali sono specificati i collegamenti esterni.

pesca (attività)

The industry of catching, processing and selling fish
Definition is not available for the current language.

Themes

- pesca

Groups

- AGRICOLTURA; ATTIVITA' FORESTALI; ALLEVAMENTO; PESCA

Related terms

Related:

- [legislazione sulla pesca](#)
- [malattia dei pesci](#)
- [morta di pesci](#)
- [politica della pesca industriale](#)

Narrower:

- [caccia con trappole](#)
- [luogo di pesca](#)
- [pesca accessoria](#)
- [pesca eccessiva](#)
- [pesca in mare](#)
- [pesca interna](#)
- [pesca professionale](#)
- [resa del pescato](#)
- [struttura di pesca](#)

Other relations

Has close match: [UMTHES: Fischerei](#)
 Wikipedia article: [Fischerei](#)

Translations

Arabic:	الصيد
Basque:	amantza
Bulgarian:	Рибарство
Catalan:	pesquera
Chinese:	漁業
Croatian:	ribarstvo
Czech:	rybářství
Danish:	fiskei
Dutch:	visserij
English:	fishery
English (US):	fishery
Estonian:	kalandus, kalamegistus
Finnish:	kalapuu
French:	caladus
French (general):	pêche (général)
German:	Fischerei
Hebrew:	פִּשְׁתִּיּוֹת
Latvian:	zivniecība
Lithuanian:	žuvininkystė
Maltese:	sajd (attività)
Norwegian:	fiskei
Polish:	rybactwo
Portuguese:	pescaria
Romanian:	pescul
Russian:	рыболовство
Slovak:	rybné hospodárstvo
Slovenian:	ribarstvo in predelava rib
Spanish:	pesquerías
Swedish:	fiskei
Turkish:	balıkçılık (genel)
Ukrainian:	рибальство

Figura 19. Struttura organizzativa di GEMET e evidenza delle correlazioni esterne.

La Figura 19 riporta la sistematizzazione del termine italiano *Pesca (attività)*, per il quale le informazioni specificate sono:

- la definizione espressa in inglese poiché non disponibile in italiano (*The industry of catching, processing and selling fish*);
- il tema a cui il concetto afferisce (Attività);
- i gruppi di appartenenza, Agricoltura; Attività forestali; Allevamento; Pesca;
- i termini correlati introdotti dall'etichetta *Related*, ad esempio legislazione sulla pesca, malattia della pesca;

²¹³ Thesaurus in lingua tedesca inerente alla tutela dell'ambiente, è gestito dalla Federal Environment Agency, Germania e contiene numerosi equivalenti in lingua inglese. <http://www.portalu.de/portal/search-detail.psm?cmd=doShowObjectDetail&docuuiid=B81A2791-6073-11D3-9E41-00C095EF00F7&plugid=/kug-group:kug-iplug-udk-db_uba> (ultima consultazione 16/12/2014).

- i termini più specifici introdotti dall'etichetta *Narrower*, ad esempio caccia con trappole, luogo di pesca;
- i collegamenti esterni introdotti: da *HasCloseMatch* seguito dal concetto a cui è equivalente, ad esempio *UMTHES: Fischerei* e da *Wikipedia article* che rinvia alla voce *Fishery* di Wikipedia.

EARTH - Environmental Applications Reference Thesaurus (Earth Thesaurus)

Earth Thesaurus²¹⁴ è un progetto sviluppato a partire dal 2001 dal Consiglio Nazionale delle Ricerche – Istituto sull'Inquinamento Atmosferico (CNR-IIA) EkoLab²¹⁵ con l'obiettivo di predisporre una terminologia condivisa per il dominio dell'ambiente ed estendere GEMET.²¹⁶ Analogamente a GEMET anche Earth include al suo interno terminologia proveniente da altri KOS riguardanti le tematiche ambientali²¹⁷ ma a differenza di quest'ultimo è predisposto unicamente in lingua inglese e in italiano

Earth è un sistema multidimensionale basato su un sistema di categorie logiche, che rappresentano il livello più alto nella struttura a cui si riconducono tutti gli altri termini presenti nel thesaurus. Le categorie logiche sono:²¹⁸

- entità: costituiscono gli oggetti di interesse (*things*) e sono suddivisi in entità materiali e immateriali;
- attributi: identificano le caratteristiche delle entità e sono articolati in composizione e proprietà;
- aspetti dinamici: definiscono le operazioni e i processi che possono riguardare le entità e sono scomposti in attività, condizioni e processi;
- dimensioni: definiscono lo spazio e il tempo in cui si collocano le entità e si suddividono per l'appunto in tempo e spazio.

Earth è un thesaurus in costante aggiornamento dato il dominio di afferenza caratterizzato da un'evoluzione costante e dall'emergere di continue problematiche

²¹⁴ <<http://thesaurusonline.iaa.cnr.it/vocabs/earth/en/index.php>> (ultima consultazione 16/12/2014).

²¹⁵ È stato costituito nel 1987 le sue attività riguardano l'organizzazione della conoscenza e lo sviluppo di sistemi per l'organizzazione, la gestione e la fruizione di informazioni su tematiche ambientali. <<http://ekolab.iaa.cnr.it/>> (ultima consultazione 16/12/2014)..

²¹⁶ «cfr.» RICCARDO ALBERTONI, et. alii, *EARTH: an Environmental Application Reference Thesaurus in the Linked Open Data Cloud*, «Semantic Web», vol.5, n. 2, 2014, Amsterdam, IOS Press, 2014. ISSN 1570-0844 <http://www.semantic-web-journal.net/system/files/swj520_0.pdf>. (ultima consultazione 16/12/2014).

²¹⁷ La terminologia presente in Earth proviene: dall'*Italian Thesaurus of Earth Sciences*; dall'*Inland Water Terminology*; dall'*Emergency Management Terms Thesaurus*, ecc.. *Ibidem*.

²¹⁸ FULVIO MAZZOCCHI, PAOLO PLINI, *Development of enviromental application reference thesaurus (EARTH)*, ISKO. 7° Capítulo español. Congreso, Barcelo, 2005, p.450, <<http://bd.ub.edu/isko2005/mazzocchi.pdf>> (ultima consultazione 16/12/2014).

legate all'ambiente. Anche questo thesaurus così come quelli precedentemente descritti è stato pubblicato in linked data e prevede numerosi *exact match* con i termini presenti in GEMET, anche perché molti termini sono in comune visto che Earth ne rappresenta l'estensione. Earth ha inoltre, sfruttato i collegamenti esterni di GEMET verso altri sistemi di organizzazione della conoscenza quali AGROVOC, EUROVOC, DBpedia e UMBEL mentre i restanti concetti non condivisi con GEMET sono stati correlati ai medesimi sistemi in maniera semi-automatica ricorrendo prima all'ausilio del software SILK²¹⁹ e successivamente alla valutazione dell'attendibilità e della consistenza da parte degli esperti di dominio.

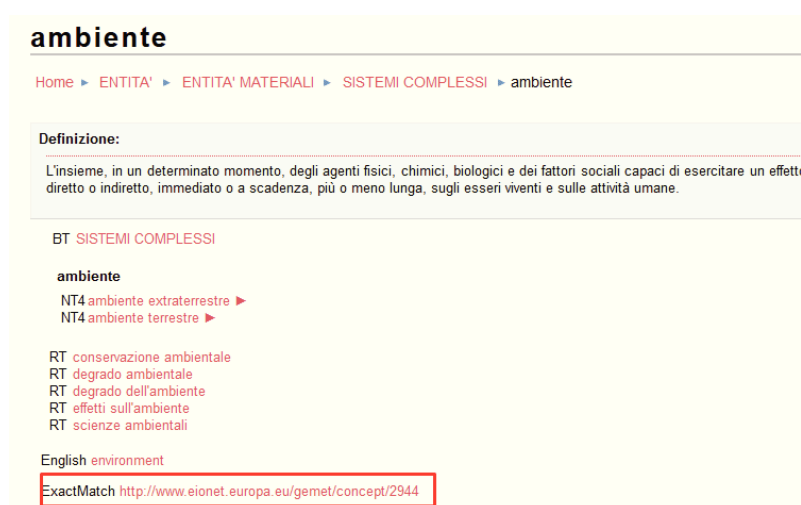


Figura 20. Strutturazione in EARTH e specificazione del collegamento esterno a GEMET.

La Figura 20 dà evidenza della sistematizzazione del termine *Ambiente*, esso afferisce innanzitutto alla categoria *Entità* suddivisa a sua volta in *Entità materiali* e *Sistemi complessi*, ha una definizione, le relazioni thesaurali tipiche introdotte da BT, NT e RT, l'equivalente in inglese e la specificazione del collegamento a risorse esterne, in questo caso del tipo *exact match*.

Library of Congress Subject Headings (LSCHE)

LSCHE è un vocabolario controllato, l'unico repertorio delle voci di soggetto valido e riconosciuto come standard a livello internazionale ed è curato dalla *Library of*

²¹⁹ È un tool che consente di scoprire relazioni tra dataset pubblicati in linked data.
<<http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/>> (ultima consultazione 16/12/2014).

*Congress*²²⁰ dal 1898, che lo impiega per l'indicizzazione di milioni di risorse bibliografiche di vario genere.

Nato per rispondere ad esigenze proprie del catalogo interno della *Library of Congress* ha riscosso nel tempo un enorme successo tanto da divenire un vero e proprio strumento di riferimento per l'intera comunità bibliotecaria internazionale.

La prima versione del repertorio è stato pubblicato a partire dal 1909 e da allora è stato interessato da numerosi avvenimenti, dettagliatamente riportati in (Stone, 2000),²²¹ che ne hanno influenzato la struttura e il contenuto, fino ad arrivare all'attuale trentaseiesima edizione.²²² LCSH è aggiornato periodicamente di modo da riflettere gli sviluppi sociali, tecnologici e terminologici, è stato calcolato che ogni anno entrano a far parte della lista circa 5,000 soggetti proposti sia dai catalogatori interni della *Library of Congress* che dai partecipanti al *Subject Authority Cooperative Program* (SACO), le sottomissioni approvate entrano a far parte della versione on line.

Ciascuna intestazione di soggetto può essere costituita da uno o più termini espressi al singolare se riguardano i concetti, al plurale se invece si riferisce ad oggetti (ad esempio *Hope* o *Railroad cars*), a questa regola sono, tuttavia, previste delle deroghe. Inoltre, le intestazioni di soggetto possono essere accompagnate da: un numero di classe che consente di circostanziare il soggetto rispetto all'aspetto considerato (ad esempio *Hope* BD216 - *Railroad cars* TF600-TF606); una nota d'ambito (SN) specificata, se necessario per garantire la coerenza nell'uso del soggetto. Nel 1985 il repertorio LSCH è stato interessato da una rivisitazione e da un aggiornamento interno che ha portato alla sua attuale configurazione strutturale, divenendo un thesaurus in cui ciascuna intestazione di soggetto è concettualmente strutturata sulla base delle relazioni thesaurali tipiche: di equivalenza, di gerarchia, di associazione introdotte rispettivamente da USE e UF, BT e NT, RT. Sempre per quanto riguarda la struttura, un altro elemento fondamentale nella LSCH è rappresentato dalle suddivisioni usate per combinare nella medesima intestazione di soggetto più concetti soprattutto nel caso in cui i *topic* siano complessi. Tali suddivisioni possono riguardare: l'argomento trattato, la forma, il periodo cronologico in cui si colloca, la dimensione spaziale, (ad

²²⁰ La *Library of Congress* è la più antica istituzione culturale degli Stati Uniti e la più grande biblioteca del mondo che ospita milioni di libri, manoscritti, fotografie, ecc.. <<http://id.loc.gov/authorities/subjects.html>> (ultima consultazione 16/12/2014).

²²¹ «cfr.» GAETANA COGNETTI, Recensione a *The LCSH Century: One Hundred Years with the Library of Congress Subject Headings System*, «Bollettino AIB», vol. 42, n. 1, 2002. URN:NBN:IT:AIB-5257 <<http://bollettino.aib.it/article/view/7155>> (ultima consultazione 16/12/2014).

²²² La versione più recente della LSCH è del giugno 2014. <<http://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html#About>> (ultima consultazione 16/12/2014)..

esempio Brooklyn (New York, N.Y.) --History, New York (State) -- History--Colonial period, ca. 1600-1775, New York (State) --History—1865-).

La LSCH consta di circa 30.000 descrittori pubblicati in linked data in formato SKOS e in *Metadata Authority Description Schema in RDF* (MADS/RDF)²²³ a cui si aggiunge un vocabolario in OWL.

Le intestazioni di soggetto sono identificate univocamente per mezzo di un URI che si presenta nel seguente formato *http://id.loc.gov/authorities/{identifier}* inoltre per ciascuna intestazione sono specificati collegamenti a risorse esterne che si evincono a partire dai tag *Exact Matching Concepts from Other Schemes* e *Closely Matching Concepts from Other Schemes* rispettivamente identificate in SKOS come `<skos:closeMatch>` e `<skos:exactMatch>`.



Figura 21. Intestazione di soggetto della LBSH e specificazione dei collegamenti esterni.

La Figura 21 riporta la sistematizzazione del termine *Cement* nella LCSH in cui sono specificate le correlazioni a risorse esterne che in questo caso rinviano verso il NAL

²²³ Data model per esprimere i dati di autorità e i vocabolari impiegati in ambito bibliotecario. <http://www.loc.gov/standards/mads/rdf/> (ultima consultazione 16/12/2014).

Thesaurus²²⁴ tramite un *exact match*; il catalogo open data data.bnf.fr,²²⁵ e verso il catalogo della *Deutschen Nationalbibliothek*.²²⁶

DBpedia

DBpedia ospita al suo interno in maniera strutturata le informazioni contenute in Wikipedia²²⁷, la più grande risorsa informativa al mondo realizzata e mantenuta dagli stessi utenti che quotidianamente la consultano. Wikipedia prima di essere considerata uno strumento di diffusione e divulgazione dell'informazione è un progetto di carattere internazionale che ha dato vita ad una nuova forma di cooperazione ed interazione che si propone di sviluppare, partendo dal basso, ovvero gli utenti, una grande base di conoscenza multidisciplinare a disposizione della collettività.

Di fatto Wikipedia non può essere annoverata tra le risorse terminologiche e semantiche di tipo autorevole poiché al suo interno è possibile riscontrare alcune incoerenze di carattere terminologico e semantico, soprattutto in relazione a contesti disciplinari precisi e altamente specialistici e dovuti in sostanza all'assenza di un adeguato consorzio di esperti addetto al processo di verifica e validazione delle voci che quotidianamente sono inserite al suo interno. Nonostante Wikipedia non offra alcuna forma di controllo terminologico, sono migliaia gli utenti che frequentemente vi accedono per avere informazioni di vario genere. Le sue potenzialità risiedono innanzitutto nell'approccio che segue per il suo costante arricchimento e successivamente nella numerosità di dati presenti al suo interno.

DBpedia si presenta quindi come la forma strutturata di Wikipedia, è disponibile in diverse lingue, tra cui anche in italiano: DBpediaItalia. La versione inglese di DBpedia è al centro della *data cloud* presentata nel Capitolo 3 e data la numerosità dei collegamenti e il carattere multidisciplinare che la contraddistinguono essa rappresenta il punto di contatto tra i diversi dataset attualmente resi disponibili sulla rete in formato RDF.²²⁸

²²⁴ <<http://agclass.nal.usda.gov/mtwdk.exe?k=default&l=60&n=1&s=5&t=2&w=564>> (ultima consultazione 16/12/2014).

²²⁵ <<http://data.bnf.fr/11964832/cementome/>> (ultima consultazione 16/12/2014).

²²⁶ <<https://portal.dnb.de/opac.htm?method=simpleSearch&cqlMode=true&query=idn%3D040675742>> (ultima consultazione 16/12/2014).

²²⁷ <https://it.wikipedia.org/wiki/Pagina_principale> (ultima consultazione 16/12/2014).

²²⁸ A titolo esemplificativo si ricordano solo alcuni dei dataset attualmente collegati a DBpedia: *Freebase*, *OpenCyc*, *UMBEL*, *GeoNames*, *Musicbrainz*, *CIA World Fact Book*, *DBLP*, il progetto *Gutenberg*, ecc..

Tale progetto è stato avviato a partire dal 2007 a seguito di una collaborazione tra l'Università di Mannheim, l'Università di Lipsia, la *Freie Universität Berlin* e l'*OpenLink Software*.

Attualmente, la versione inglese di DBpedia è stata rilasciata con release 3.9²²⁹ e contiene circa 4.58 milioni voci, organizzate in classi (circa 685) e descritte per mezzo di proprietà (circa 2,795) organizzate all'interno di un modello ontologico predisposto a seguito della dettagliata analisi dei contenuti presenti in Wikipedia.

Ogni entità codificata all'interno di DBpedia è univocamente identificata per mezzo di un URI a cui corrisponde una pagina in Wikipedia, ad esempio alla risorsa *Giacomo Leopardi* di DBpedia identificata dall'URI http://fr.dbpedia.org/page/Giacomo_Leopardi coincide una pagina Wikipedia http://en.wikipedia.org/wiki/Giacomo_Leopardi.

The screenshot shows the DBpedia entry for Giacomo Leopardi. At the top, it says 'Giacomo Leopardi' and 'From Wikipedia, the free encyclopedia'. Below this, a note states: '"Leopardi" redirects here. For other uses, see Leopardi (disambiguation)'. A warning box indicates that the article needs additional citations for verification. The main text of the article is visible, starting with 'Giacomo Taldegardo Francesco di Sales Saverio Pietro Leopardi (italian: [ˈdʒakomo leoˈpardi]; June 29, 1798 – June 14, 1837) was an Italian poet, philosopher, essayist and philologist. Although he lived in a secluded town in the ultra-conservative Papal States, he came in touch with the main thoughts of the Enlightenment, and, by his own literary evolution, created a remarkable and renowned poetic work, related to the Romantic era.' Below the text, there is an 'About: Giacomo Leopardi' section with an 'Entity of Type' field set to 'person' and a 'Data Space' field set to 'fr.dbpedia.org'. At the bottom, there is a table with 'Property' and 'Value' columns, showing various biographical details in multiple languages.

Figura 22. Corrispondenza tra una pagina Wikipedia e una risorsa DBpedia.

In Wikipedia oltre alle informazioni di carattere testuale ve ne sono delle altre rappresentate tramite infobox riguardanti: la categoria di appartenenza, le immagini, i riferimenti esterni, i collegamenti ad altre pagine, ecc.. Anche questa tipologia di

²²⁹ L'ultima versione risale al 9 settembre 2009. <<http://it.wikipedia.org/wiki/DBpedia>> (ultima consultazione 16/12/2014).

informazioni sono a loro volta estratte e codificate ricorrendo ai formalismi del web semantico e quindi tradotte in triple RDF.

All'interno di DBpedia, ciascuna entità può essere organizzata ricorrendo ad i seguenti sistemi.²³⁰

- YAGO:²³¹ è un modello derivato dall'allineamento tra Wikipedia, WordNet e GeoNames, prevede una copertura semantica molto ampia e consente di esprimere entità, fatti, relazioni tra fatti e proprietà inerenti alle relazioni,²³²
- *Upper Mapping and Binding Exchange Layer* (UMBEL):²³³ è un vocabolario di riferimento per la realizzazione di ontologie di dominio ed è basato sul concetto di classe e predicato. In particolare, le classi previste da questo modello sono: *RefConcept*, *SuperType*, e *Qualifier* mentre i predicati che è possibile specificare sono circa 38 tra cui *correspondsTo*, *isAbout*, *isRelatedTo*, *relatesTo*, ecc.,²³⁴
- DBpedia Ontology: è l'ontologia su cui si basa DBpedia, è stata realizzata manualmente a partire dal contenuto informativo presente nelle *infobox*.

DBpedia è distribuita con una doppia licenza: *Creative Commons Attribution-ShareAlike 3.0*²³⁵ che ne prevede la distribuzione, il riutilizzo e la trasformazione anche per fini commerciali, indicando sempre la paternità e le eventuali modifiche apportate; *GNU Free Documentation License*, una licenza sviluppata dalla *Free Software Foundation*²³⁶ nell'ambito del progetto *GNU's Not Unix* (GNU)²³⁷ a supporto della distribuzione di materiale didattico e manuali inerenti alle applicazioni software.²³⁸

Come anticipato DBpedia è disponibile in circa 125 lingue tra cui l'italiano DBpediaItalia. DBpediaItalia è stata sviluppata da un gruppo di ricercatori afferenti alla

²³⁰ «cfr.» TOMMASO DI NOIA, et alii, *Semantic web. Tra ontologie e Open Data*, Milano, Apogeo, 2013. ISBN: 8850332149 - ISBN-13: 9788850332144.

²³¹ <<http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/>> (ultima consultazione 16/12/2014)..

²³² FABIAN M. SUCHANEK, et alii, *YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia, proceedings of the 16th International World Wide Web Conference (WWW 2007) Banff, Canada, 2007*. ISBN:978-1-59593-654-7 - DOI:10.1145/1242572.1242667 <<http://suchanek.name/work/publications/www2007.pdf>> (ultima consultazione 16/12/2014).

²³³ <<http://www.umbel.org/>> (ultima consultazione 16/12/2014).

²³⁴ <http://techwiki.umbel.org/index.php/UMBEL_Vocabulary> (ultima consultazione 16/12/2014).

²³⁵ <<https://creativecommons.org/licenses/by-sa/3.0/it/>> (ultima consultazione 16/12/2014).

²³⁶ È un'organizzazione no-profit che opera a livello mondiale e promuove la cultura del libero accesso ad applicazioni software da parte degli utenti. <https://www.fsf.org/?set_language=it> (ultima consultazione 16/12/2014).

²³⁷ Il progetto è stato avviato nel 1984 con lo scopo di sviluppare un sistema operativo del tipo Unix di modo che potesse essere utilizzato liberamente da chiunque. Esso garantisce all'utente quattro livelli di libertà: libertà 0: l'utente può decidere di eseguire il programma per perseguire qualsiasi fine; libertà 1: l'utente può intervenire sul codice sorgente e modificarlo per adattarlo alle proprie esigenze; libertà 2: può distribuire la propria versione; libertà 3: può rendere noti e condividere le proposte migliorative apportate al sistema. <<https://www.gnu.org/home.it.html>> (ultima consultazione 16/12/2014).

²³⁸ <<http://www.gnu.org/copyleft/fdl.html>> (ultima consultazione 16/12/2014).

Fondazione Bruno Kessler di Trento²³⁹ insieme alla startup SpazioDati,²⁴⁰ con l'obiettivo di convertire il contenuto disponibile in lingua italiana in formato RDF di modo da poter essere integrato di nuove voci e collegato ad altri dataset. Ai fini della presente ricerca la versione italiana di DBpedia ha un enorme valore in quanto è stata impiegata nel processo di definizione delle correlazioni semantiche.

2. L'organizzazione della conoscenza di dominio

L'organizzazione della conoscenza nel settore delle costruzioni è una tematica ampiamente discussa in ambito europeo e internazionale, meno all'interno del contesto italiano. Ai fini della presente ricerca saranno presi in considerazione i progetti più significativi che si propongono l'organizzazione e la sistematizzare della conoscenza di dominio di modo da renderla fruibile, utilizzabile e comprensibile da tutti.²⁴¹

L'analisi dello stato dell'arte parte dal *Construction index/ Samarbetskommitten for Byggnadsfrågor*²⁴² (CI/SfB), sistema di classificazione nato negli anni '60 al fine di supportare l'organizzazione del materiale documentale degli uffici, la redazione di relazioni tecniche e di computi metrici estimativi, la divulgazione e il trasferimento delle informazioni inerenti ai prodotti edili. Lo schema è stato largamente utilizzato all'interno del settore poiché fornisce una terminologia controllata e in aggiunta presenta una struttura flessibile e ampliabile, caratteristiche queste che ne favoriscono l'aggiornamento rispetto alle innovazioni tecnologiche. Ciò nonostante la sua ultima revisione risale al 1976, ragion per cui il sistema non codifica le novità avvenute nel settore da quella data in poi. CI/SfB può essere inteso come il primo esempio nel settore, di sistema di classificazione per punti di vista in cui i termini attestati nella letteratura di dominio sono raggruppati in faccette. Lo schema si compone di cinque categorie o tavole (*table*) generali:

- Tavola 0 (Ambiente fisico o Ambiente naturale e costruito): classifica le informazioni relative alla progettazione e alla pianificazione urbana e territoriale e inerenti ad esempio agli impianti industriali, alle strutture ricettive, alle

²³⁹ È un ente di ricerca affermato sia in ambito nazionale che internazionale, con una forte attenzione allo sviluppo locale, le sue attività sono essenzialmente rivolte allo sviluppo di tecnologie dell'informazione e al trasferimento tecnologico. <<http://www.fbk.eu/it/front-page>> (ultima consultazione 16/12/2014).

²⁴⁰ <<http://www.spaziodati.eu/it/#startup>> (ultima consultazione 16/12/2014).

²⁴¹ Per una ricognizione maggiormente esaustiva e completa delle iniziative e dei vocabolari controllati realizzati in ambito nazionale e internazionale si rimanda a Folino, *Op. cit.*.

²⁴² Si tratta del Comitato di coordinamento per i problemi edilizi svedese.

strutture residenziali, ecc.. Il codice identificativo di questa tavola si compone di una coppia di numeri che va da 0 a 9;

- Tavola 1 (Elementi): classifica gli elementi che entrano a far parte di un'opera edilizia, ad esempio il muro, le scale, il tetto, ecc.. Il codice è costituito da una coppia di numeri tra parentesi tonde “(27) *Roofs*”;
- Tavola 2 (Lavori): classifica le diverse tipologie di lavorazione ad esempio tubi, giunti, cavi e reti. Il codice identificativo di questa tavola è rappresentato da una lettera alfabetica posta in maiuscolo “A *Constructions, forms*”;
- Tavola 3 (Materiali): classifica i materiali impiegati nella realizzazione di un'opera edilizia, ad esempio legno, metalli, fibre inorganiche, ecc.. Anche in questo caso il codice è costituito da una lettera alfabetica ma specificata in minuscolo “o *Glass*”;
- Tavola 4 (Attività e requisiti): classifica l'insieme delle attività che congiuntamente intervengono nella realizzazione di un opera edilizia, ad esempio organizzazione degli uffici e del progetto, trasporto, ispezioni di qualità, ecc.. Il codice identificativo in questo caso è di tipo alfanumerico “A6 *Buying, delivery*”.²⁴³

Le tavole organizzate al loro interno in faccette permettono la combinazione delle informazioni in relazione a precise esigenze. Di fatto questo sistema consente di combinare non solo termini appartenenti a tavole diverse ma anche termini appartenenti alla medesima tavola. Nonostante CI/SfB sia un sistema dotato di una certa flessibilità e capacità combinatoria, risulta essere non esaustivo e non sufficientemente completo, a causa della complessità della notazione alfanumerica e della tipologia di informazioni che contempla, principalmente inerenti alla realizzazione di opere civili per scopi residenziali. Per cui esso risulta poco adatto alla gestione della complessità e della varietà delle informazioni del settore, soprattutto in contesti digitali.

Date tali limitazioni, il sistema di classificazione CI/SfB è stato superato dallo schema di classificazione *Unified Classification for the Construction Industry* (Uniclass), realizzato nel 1997 dal *National Building Specification* (NBS)²⁴⁴ per conto del *Construction Project Information Committee* (CPIC)²⁴⁵ e implementato nel Regno Unito,

²⁴³ <http://www.tecnologica.altervista.org/php5/index.php/Piano_di_classificazione_SfB> (ultima consultazione 16/12/2014).

²⁴⁴ Entità che si occupa della definizione di specifiche e soluzioni tecniche a supporto del settore delle costruzioni riconosciute nel Regno Unito. <<http://www.thenbs.com/index.asp>> (ultima consultazione 16/12/2014).

²⁴⁵ È un'organizzazione che comprende i rappresentanti dei più importanti enti e organizzazioni che operano nel settore: il Royal Institute of British Architects (RIBA), il Royal Institution of Chartered Surveyors (RICS), l'Institution of Civil Engineers (ICE), il UK Contractors Group (UKCG), il Chartered

con lo scopo di organizzare la letteratura di settore e la mole di informazioni che interessa un progetto edilizio. Uniclass si sviluppa a partire da CI/SfB ma viene alimentato con informazioni provenienti da altri sistemi quali il *Common Arrangement of Works Sections* (CAWS), il *Civil Engineering Standard Method of Measurement* (CESSM) e altri ancora.²⁴⁶ Uniclass è un'implementazione della ISO 12006-2 e può essere impiegato per classificare i diversi aspetti legati alla progettazione e alla realizzazione di un'opera, alla gestione dei costi e delle informazioni che la riguardano. Il sistema Uniclass è stato aggiornato nel 2013, la versione attualmente utilizzabile è Uniclass2, che rispetto alla precedente risulta semplificata nel numero delle tavole,²⁴⁷ che sono passate da 15 a 8 (ulteriori revisioni sono previste), maggiormente consistente e aggiornata nei contenuti. Uniclass2 si propone di mappare l'insieme delle informazioni inerenti all'intero ciclo del progetto classificandole su base strutturale, la notazione alfanumerica che accompagna ciascuna entrata dello schema definisce una struttura arborescente come di seguito esemplificato:

Ac_25_30 Scientific Laboratory Activities

Ac_25_30_02 Analytical Chemistry Researching

Ac_25_30_03 Animal Researching

Ac_25_30_04 Astronomical Researching

*Ac_25_30_05 Astronomical Observing.*²⁴⁸

Uniclass2 è un sistema aperto, liberamente accessibile, disponibile in formato *Comma-Separated Values* (CSV)²⁴⁹ e pubblicato sotto la licenza *Creative Commons*²⁵⁰ *Attribution-NoDerivs 3.0 Unported* (CC BY-ND 3.0) che consente di “riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato per qualsiasi fine, anche

Institute of Architectural Technologists (CIAT) e il Chartered Institution of Building Services Engineers (CIBSE).

²⁴⁶ «cfr.» ANN ROSS, JONATHAN HETREED, *Architect's Pocket Book*, Architectural Press, 2011. ISBN: 9780080969596.

²⁴⁷ Le tavole di cui tale schema si compone sono: Complexes, Entities, Activities, Spaces, Entities by Form, Elements, Systems, Products, CAD, Project Phases. Per informazioni dettagliate in merito alla tipologia delle informazioni ospitate e classificate all'interno del Sistema di classificazione Uniclass2 si rimanda al seguente indirizzo: <<http://www.cpic.org.uk/uniclass2/>> (ultima consultazione 16/12/2014).

²⁴⁸ L'esempio è stato estratto dalla Tavola *Ac Activities* nella versione stampabile aggiornata al 03/12/2013.

²⁴⁹ “È un formato di file basato su file di testo utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o database) di una tabella di dati.” <http://it.wikipedia.org/wiki/Comma-separated_values> (ultima consultazione 16/12/2014).

²⁵⁰ Creative Commons è un'associazione no-profit che si occupa di tutelare le opere intellettuali predisponendo talune licenze in cui il creatore può decidere quali diritti tenere riservati e quali mettere a disposizione della collettività. <http://it.wikipedia.org/wiki/Creative_Commons> (ultima consultazione 16/12/2014).

*commerciale*²⁵¹ purché ne sia riconosciuta la paternità e siano segnalate eventuali modifiche (integrazioni, rettifiche, ecc.) effettuate. Infine, Uniclass2 è un sistema dinamico, per cui laddove si evidenziassero dei *gap* informativi o classificatori è possibile previa attenta valutazione del CPIC incrementarne e aggiornarne la struttura.

MasterFormat® è un indice standardizzato di numeri e titoli²⁵² che si propone l'organizzazione dei dati relativi alle *construction specifications*, ovvero documenti descrittivi in cui sono riportati l'insieme delle informazioni di natura tecnica ed economica relative ad un dato progetto. MasterFormat® è uno strumento diffuso soprattutto nel Nord America, la prima edizione è stata pubblicata nel 1963 ma data la sua diffusa applicazione è stato più volte esaminato e rivisto. La versione che ha introdotto significative differenze rispetto a quella originale, è del 2004 in cui il numero delle divisioni di cui l'indice si compone sono passate da 16 a 50 mentre il sistema notazionale da 5 a 6 caratteri, modifiche che mettono in luce l'aumentare della complessità del settore, delle esigenze avvertite al suo interno e delle continue innovazioni che lo interessano. Gli aggiornamenti sono effettuati con cadenza biennale da una comunità di volontari supportati da proposte e iniziative avanzate dagli utenti di tale strumento, attualmente la versione più aggiornata è del 2014.

Uniformat^{TM253} è una metodo di organizzazione delle informazioni di settore per la descrizione, l'analisi economica e la gestione del ciclo di vita di un'opera. La sua struttura si basa su elementi funzionali o di loro parti che ne determinano il funzionamento, a prescindere dai materiali impiegati, dal metodo di costruzione e dalle specifiche di progettazione adottati per la realizzazione. UniformatTM è uno strumento principalmente usato per redigere computi metrici estimativi e/o descrizioni dei progetti preliminari²⁵⁴ e per preventivare e monitorare i costi inerenti alla realizzazione di un'opera. La versione più aggiornata è del 2010.

Electronic Product Information Cooperation (EPIC) è nato nel 1990 per volontà di dieci paesi europei ed è uno standard per lo scambio delle informazioni tra i diversi provider del settore che si propone di agevolare la diffusione e la distribuzione delle informazioni in considerazione dello sviluppo delle nuove tecnologie.

EPIC, UniformatTM, MasterFormat® confluiscono in OmniClassTM *Construction Classification System* (OCCS).²⁵⁵ Si tratta del sistema di classificazione più importante

²⁵¹ <<http://creativecommons.org/licenses/by-nd/3.0/it/>> (ultima consultazione 16/12/2014).

²⁵² <<http://csc-dcc.ca/Document+Store/MasterFormat/>> (ultima consultazione 16/12/2014).

²⁵³ <<http://www.csinet.org/uniformat>> (ultima consultazione 16/12/2014).

²⁵⁴ <<http://www.csinet.org/Home-Page-Category/Formats/UniFormat.aspx>> (ultima consultazione 16/12/2014).

²⁵⁵ <<http://www.omniclass.org/>> (ultima consultazione 16/12/2014).

tra quelli annoverabili nel settore sia perché basato sui principi della ISO 12006-2:2001 *Building construction - Organization of information about construction works - Part 2: Framework for classification of information* sia per la copertura semantica che fornisce. OmniClass™ è stato elaborato per rispondere a specifiche esigenze di organizzazione e gestione delle informazioni nei domini dell'architettura, dell'ingegneria e della costruzione avvertite in contesto nordamericano e offre una visione della conoscenza di dominio per punti di vista. Esso infatti, si compone di 15 tabelle, ciascuna delle quali analizza e struttura gli aspetti di dominio ritenuti fondamentali:

- Tabella 11 - *Construction Entities by Function*: organizza le entità in base allo scopo e/o all'uso, ad esempio *Library* intesa in senso edile come una struttura per la raccolta, la conservazione e il recupero di risorse analogiche e digitali;
- Tabella 12 - *Construction Entities by Form*: struttura le entità considerandole indipendentemente dal loro uso e/o applicazione, ad esempio *Bridge* definito come struttura costruita per superare un ostacolo e fornire un passaggio a veicoli o persone;
- Tabella 13 - *Spaces by Function*: classifica gli spazi in base alla destinazione d'uso, ad esempio *Waste and Recycling Spaces* definito come spazio utilizzato per la raccolta dei rifiuti e dei materiali riciclabili;
- Tabella 14 - *Spaces by Form*: classifica gli spazi intesi come costituiti da confini ben precisi, collocandoli all'interno di aree più grandi, ad esempio *Transition Spaces* (Spazi di passaggio) è specializzato in *Corridor* (Corridoio), *Vestibule* (Vestibolo/Atrio), *Nave* (Navata);
- Tabella 21 - *Elements*: classifica i componenti principali di un'opera, ad esempio *Wall*, *Roof* e *Floor*;
- Tabella 22 - *Work Results*: il concetto di *Work Results* indica un'entità conclusa che esiste a seguito di un processo di lavorazione, ad esempio *Ceramic Tiling* (Rivestimento in ceramica) e *Railways* (Ferrovie);
- Tabella 23 - *Products*: i *Products* indica gli elementi più piccoli impiegati nella realizzazione di un'opera; gli elementi possono essere unici o scaturire dall'assemblaggio di più prodotti, ad esempio *Concrete* (Calcestruzzo) e *Door* (Porta);
- Tabella 31 - *Phases*: classifica le diverse fasi che interessano la filiera edilizia, ad esempio *Design Phase* (Fase di progettazione) *Coordination Phase* (Fase di coordinazione);
- Tabella 32 - *Services*: organizza l'insieme dei servizi previsti e forniti per la realizzazione e il mantenimento di un'opera, ad esempio *Financing*

(Finanziamenti per disporre di capitali da impiegare per la realizzazione di un'opera);

- Tabella 33 – *Disciplines*: riguarda le aree in cui si collocano le attività svolte dai diversi soggetti che lavorano alla realizzazione di un'opera, ad esempio *Urban Planning* (Pianificazione urbana), *Environmental Planning* (Pianificazione ambientale), ecc.;
- Tabella 34 - *Organizational Roles*: indica il ruolo ricoperto da coloro che intervengono nella realizzazione di un'opera, ad esempio *Architect* (Architetto), *Committee* (Committente);
- Tabella 35 - *Tools*: organizza le risorse impiegate per sviluppare un progetto che non rientrano a far parte dell'opera, ad esempio *Hammers* (Martelli), *Compressors* (Compressori), ecc.;
- Tabella 36 - *Information*: classifica i dati prodotti e trasmessi per la realizzazione di un'opera edilizia, ad esempio *Guides*, *Catalogues*, ecc.;
- Tabella 41 – *Materials*: specifica i materiali e/o le sostanze da utilizzare *tout court* o per la realizzazione dei prodotti da impiegare successivamente nel processo edilizio, ad esempio *Granite* (Granito), *Quartz* (Quarzo), ecc.;
- Tabella 49 - *Properties*: classifica le caratteristiche delle entità costruttive, ad esempio *Color* (Colore), *Width* (Larghezza), *Length* (Lunghezza), ecc..

Dal punto di vista della struttura OmniClass™ si configura come uno schema di classificazione all'interno del quale la struttura di tipo gerarchico e quella a faccette (le 15 tabelle rappresentano le categorie in cui il dominio è stato scomposto e sono un chiaro esempio di organizzazione del sapere per punti di vista) sono combinate. Le tabelle riportano al loro interno le informazioni inerenti alle diverse fasi del processo edilizio e possono essere utilizzate singolarmente, per esprimere concetti semplici, o in combinazione, per esprimere concetti più complessi. In OmniClass™ a ciascun concetto è associato un codice che lo identifica in maniera univoca e lo colloca all'interno della struttura gerarchica. La struttura a faccette e il patrimonio terminologico che caratterizzano OmniClass™ ne hanno agevolato l'impiego e l'adattamento su scala mondiale. Anche l'Italia ha manifestato un certo interesse verso questo sistema di classificazione, tant'è vero che tra l'Istituto per le Tecnologie della Costruzione del Consiglio Nazionale delle Ricerche (ITC-CNR) e il *Construction Specification Institute* (CSI) e il *Construction Specification Canada* (CSC) è stata stipulata una convenzione ha manifestato con l'obiettivo di valutarne una possibile traduzione in italiano.

In aggiunta ai KOS di dominio fin qui illustrati, e che dato il loro grado di esaustività e diffusione risultano essere più noti, ve ne sono degli altri altrettanto significativi poiché sviluppati per garantire la gestione delle informazioni in contesti digitali.

e-Cognos ontology è un modello ontologico sviluppato nell'ambito dell'omonimo progetto inerente alla rappresentazione e alla formalizzazione della conoscenza di dominio estratta dai documenti, maturata dagli esperti e dalle organizzazioni, al fine di sviluppare una *Knowledge Management Infrastrutture* (e-ICKMI). Tale modello è stato sviluppato a partire dall'*Industry Foundation Classes (IFC) model*,²⁵⁶ dal *bcXML Meta Schema*²⁵⁷ e dal vocabolario *Building and civil engineering* (BS6100).²⁵⁸ L'ontologia e-Cognos si compone di due tassonomie: una tassonomia dei concetti ripresa da IFC model e una tassonomia di relazioni. Questo modello ontologico può essere impiegato a supporto del processo di modellizzazione della conoscenza nel settore.

LexiCon, sviluppato dal *Dutch Specification Institute* (STABU), può essere inteso come l'implementazione della ISO 12006-3:2007 *Building construction - Organization of information about construction works - Part 3: Framework for object-oriented information*. Si tratta di un vocabolario strutturato di oggetti, proprietà e unità (dall'inglese *units*) ciascuno seguito da una definizione,²⁵⁹ organizzati sulla base delle seguenti categorie: *Subjects, Activities, Collections, Properties, Measures* e *Units*. A partire dal 2006 LexiCon insieme a BARBi, il linguaggio di riferimento utilizzato in Norvegia per l'organizzazione dell'informazione, realizzato dal *Norwegian Building Standards Organization*, molto simile a LexiCon nella struttura e nei contenuti, è confluito all'interno dell'*International Framework for Dictionaris (IFD) Library* al fine di realizzare un unico modello di riferimento.²⁶⁰

Per quanto riguarda il panorama italiano, l'Ente di Normazione Italiano (UNI)²⁶¹ ha risposto a questo proliferare di iniziative con una specifica norma, ovvero la UNI 11337:2009 - *Edilizia e opere di ingegneria civile - Criteri di codificazione di opere e prodotti da costruzione, attività e risorse - Identificazione, descrizione e*

²⁵⁶ Meta modello sviluppato dalla *Alliance for Interoperability (IAI)* per rappresentare i dati nel settore delle costruzioni e agevolare l'interoperabilità al suo interno. <<http://www.buildingsmart-tech.org/specifications/ifc-overview>> (ultima consultazione 16/12/2014).

²⁵⁷ Un meta-modello realizzato nell'ambito del settore delle costruzioni (bc indica in inglese *Building Construction*) basato su XML.

²⁵⁸ Si tratta di un vocabolario in lingua inglese che raccoglie la terminologia in uso nel settore delle costruzioni.

²⁵⁹ JASON UNDERWOOD, UMIT ISIKDAG, *Handbook of Research on Building Information Modeling and Construction Informatics: Concepts and Technologies*, Information Science Publishing, 2009, p. 114. ISBN 978-1-60566-928-1.

²⁶⁰ LexiCon insieme a e-Cognos è stata impiegato nel progetto europeo FUNSIEC finalizzato allo studio di fattibilità per sviluppo dell'*Open Semantic Infrastructure for Construction Sector* (OSIECS).

²⁶¹ <<http://www.uni.com/>> (ultima consultazione 16/12/2014).

*interoperabilità*²⁶² che definisce i criteri di indirizzo da adottare per l'identificazione e la descrizione univoca dei soggetti, degli oggetti e delle attività che intervengono in un processo edilizio al fine di favorire attraverso l'uso di una terminologia normalizzata, lo scambio condiviso e interoperabile delle informazioni generate e trasmesse nell'intero ciclo della filiera edilizia.

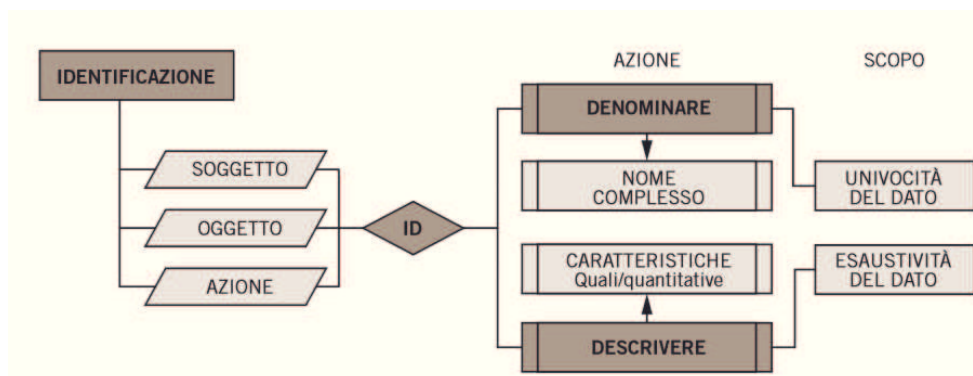


Figura 23. criteri di indirizzo della UNI 11337.²⁶³

La norma contempla tutti gli elementi coinvolti nella filiera costruttiva e si compone di due macro-parti:

- la norma vera e propria che definisce per l'appunto i criteri di indirizzo;
- le specifiche tecniche (TS) alle quali è demandato il compito di dettagliare i criteri di indirizzo, in riferimento alla denominazione, alla descrizione dei soggetti, degli oggetti e delle attività²⁶⁴ alla gestione e alla diffusione delle informazioni.

Si tratta quindi di un impianto normativo che specifica: le linee guida da tenere in considerazione in fase di identificazione di soggetti, oggetti e azioni e gli strumenti da utilizzare per procedere alla loro precisa identificazione ed esaustiva descrizione. Nonostante la norma sia tuttora in vigore, le novità che stanno riguardando il settore unitamente alla necessità di garantire la comunicazione non ambigua tra i diversi

²⁶² <http://store.uni.com/magento-1.4.0.1/index.php/uni-113372009.html?josso_back_to=http://store.uni.com/josso-security-check.php&josso_cmd=login_optional&josso_partnerapp_host=store.uni.com> (ultima consultazione 16/12/2014).

²⁶³ «cfr.» UNI 11337:2009 *Edilizia e opere di ingegneria civile - Criteri di codificazione di opere e prodotti da costruzione, attività e risorse - Identificazione, descrizione e interoperabilità*.

²⁶⁴ L'attività di redazione delle specifiche tecniche è stata demandata ad uno specifico tavolo di lavoro costituito in sede UNI "UNI GL 9 "Codificazione dei prodotti e dei processi in edilizia"" che ha già redatto la UNI TS2 ora in inchiesta pubblica, mentre le attività per la stesura della UNI TS1 sono in corso d'opera.

operatori, hanno accentuato la necessità di procedere ad una sua rivisitazione e aggiornamento.

La norma UNI 11337:2009 stabilisce che ciascun soggetto, oggetto e azione coinvolto nella filiera edilizia debba essere denominato per mezzo di un *nome complesso*²⁶⁵ composto da un *codice identificativo*²⁶⁶ e da un *enunciato esplicativo*²⁶⁷ che aggrega l'insieme delle caratteristiche discriminanti all'interno della moltitudine degli elementi che è possibile considerare. Le raccomandazioni previste dalla norma dovranno essere maggiormente impiegate nella redazione dei documenti tecnici (capitolati speciali d'appalto, prezziari, relazioni tecniche di progetto, computi metrici estimativi) e in tutti i documenti richiesti dal settore al fine di disporre di una terminologia comune e di una notazione univoca.

Per quanto concerne le specifiche tecniche, parti integranti dell'apparato normativo di cui alla UNI 11337:2009, è stato costituito in ambito UNI un tavolo di lavoro "UNI GL 9 - *Codificazione dei prodotti e dei processi in edilizia*" che si sta occupando della redazione di due delle tre specifiche tecniche previste a corredo della norma. In particolare, la UNI/TS2 - *Edilizia e opere di ingegneria civile. Modelli di raccolta, organizzazione e archiviazione dell'informazione tecnica per i prodotti da costruzione* definisce i criteri in materia di descrizione dei prodotti da costruzione²⁶⁸ e fornisce un modello strutturato per la raccolta delle informazioni inerenti ai prodotti sottoposti e non a marcatura CE. Per quanto riguarda invece, la UNI/TS1, si sta costituendo il tavolo tecnico per avviare i lavori in merito alla sua stesura. In aggiunta, a completamento di questo corpo normativo è in atto l'elaborazione di una specifica tecnica dedicata all'interoperabilità delle informazioni supportata dalla metodologia del *Building Information Modeling* (BIM).

Questo corpo normativo rappresenta quindi, nel contesto italiano l'iniziativa più importante che mira a fornire un *framework* condiviso e strategico per rendere l'informazione interoperabile, coerente e predisposta alla condivisione e alla comunicazione.

Basato sui principi e sui criteri predisposti dalla summenzionata UNI 11337:2009 è il lessico INNOVance, realizzato da TC-CNR nell'ambito dell'omonimo progetto.

²⁶⁵ "Nome articolato in più informazioni contenente ogni carattere significativo all'identificazione univoca di un qualsiasi soggetto, oggetto, od azione [...]." UNI, Op. cit., p. 2.

²⁶⁶ "Sequenza di simboli (codici alfanumerici e di interpunzione) con funzioni identificative e di ordinamento; sistema simbolico per rappresentare un nome complesso", *Ibidem*.

²⁶⁷ "Sequenza di termini, con funzioni identificative, esplicative e di ordinamento; sistema terminologico utilizzato per rappresentare un nome complesso" *Ibidem*.

²⁶⁸ "Ogni prodotto fabbricato al fine di essere incorporato in modo permanente negli edifici e nelle altre opere di ingegneria civile", UNI, Op. Cit., p. 4.

L'obiettivo generale del progetto consiste nel definire la prima banca dati nazionale per il settore delle costruzioni, in cui raccogliere, organizzare e codificare in maniera univoca i soggetti, gli oggetti e le azioni che interessano la filiera edilizia per ottimizzare la capacità e la qualità della comunicazione. Per rispondere a queste esigenze è stato sviluppato un lessico che fornisce la terminologia di settore inerente ai prodotti da costruzione e propone allo stesso tempo una rappresentazione della conoscenza di dominio, messa in evidenza dall'esplicitazione di relazioni concettuali tra i termini in esso ospitati, grazie alle quali, i concetti sono disposti in una rete semantica sia sul piano orizzontale che verticale.

In considerazione della copertura semantica circoscritta ad uno specifico dominio, e date le sue caratteristiche organizzative e strutturali, il lessico INNOVance si connota come un lessico specialistico strutturato, in cui i termini veicolano concetti precisi e univoci. Tale configurazione lo rende assimilabile ad un *thesaurus*, che può quindi rappresentare la sua naturale evoluzione e costituire inoltre la base per lo sviluppo di sistemi di organizzazione della conoscenza maggiormente strutturati e formalizzati come le ontologie.²⁶⁹

Più precisamente, il lessico ospita e sistematizza la terminologia concernente i prodotti da costruzione recanti la marcatura CE e attestata in un corpus documentale²⁷⁰ appositamente predisposto. Il corpus è una collezione di testi ritenuta sufficientemente rappresentativa della lingua o della varietà linguistica che si intende indagare e costituisce: "*la principale fonte della Linguistica Computazionale*"²⁷¹ grazie alla quale è possibile procedere ad effettuare analisi di tipo computazionale sulla lingua e/o su una varietà linguistica, osservandone ad esempio l'evoluzione, la variabilità in funzione del contesto d'uso, ecc.. Il corpus realizzato, raggruppa testi settoriali ritenuti autorevoli, quali le norme tecniche armonizzate, le norme tecniche di prodotto, la legislazione europea e nazionale, i manuali predisposti dalle associazioni di categoria e ulteriori documenti tecnici ritenuti affidabili e ha rappresentato la base per l'estrazione dei termini candidati a divenire entrate del lessico. La fase di estrazione terminologica non si è avvalsa di una procedura automatica e/o semi-automatica ma i termini di interesse sono stati individuati manualmente, in quanto i testi sono stati resi disponibili in formato

²⁶⁹ «cfr.» ELENA CARDILLO, et alii, *Towards the reuse of standardized thesauri into ontologies*, Workshop on Ontology and Semantic Web Patterns (WOP2014), co-located with the 13th International Semantic Web Conference (ISWC2014), Riva del Garda, 2014 <http://ontologydesignpatterns.org/wiki/images/0/03/Paper_8.pdf> (ultima consultazione 16/12/2014)..

²⁷⁰ "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" JOHN SINCLAIR, *Trust the text: language, corpus and discourse*, Londra, Routledge, 2004, p.14. ISBN 0-203-59407-X.

²⁷¹ ALESSANDRO LENCI, et alii, *Op. cit.*, p. 26.

cartaceo o in formati digitali protetti, che hanno di fatto, ostacolato la messa a punto di una metodologia automatica e/o semiautomatica.

Inoltre, a queste difficoltà di natura tecnica si sono aggiunte questioni proprie del linguaggio di dominio, dovute ad esempio alla presenza di numerose forme complesse, si tratta di locuzioni piuttosto lunghe e articolate come *acciaio d'armatura resistente alla corrosione, elemento di legno massiccio con sistema di assemblaggio*, ecc..

Nel lessico INNOVance sono stati sfruttati: i principi classificatori/tassonomici, per cui i termini sono stati inseriti in una rete di sovra-ordinati e sotto-ordinati semantici (dal generale al particolare); e le potenzialità delle correlazioni orizzontali grazie alle quali è possibile predisporre sullo stesso piano i concetti che nella scala gerarchica condividono il medesimo padre e specificare per ciascuno di essi i sinonimi, i quasi-sinonimi e/o le varianti morfologiche e lessicali.

Date queste premesse, il lessico INNOVance presenta una struttura arborescente, ed include sia la terminologia riguardante i prodotti da costruzione propriamente detti, e sia la terminologia inerente alle caratteristiche prestazionali riportate nella sezione ZA delle norme di prodotto.²⁷²

Il lessico pur essendo stato realizzato tenendo conto delle raccomandazioni riportate nella norma ISO 25964-1:2011 *Information and documentation – Thesauri and interoperability with other vocabularies - Part1: Thesauri for information retrieval*, si presenta come uno strumento ibrido e cioè non è una semplice lista controllata e normalizzata di termini ma una rappresentazione discretamente strutturata della conoscenza di dominio. Dal punto di vista della struttura il lessico INNOVance si caratterizza per la specificazione di relazioni paradigmatiche di seguito specificate:

- relazione di equivalenza: collega termini rappresentativi di un medesimo concetto, è stabilita tra un termine preferito (TP) e uno o più termini non preferiti (TNP). Tale relazione è espressa tramite l'etichetta *Sinonimi*, ad esempio *finestra a battente con apertura verso l'esterno* è la forma preferita di *finestra all'inglese*;
- relazione gerarchica: basata su criteri di inclusione, è stabilita tra coppie di concetti di cui uno più generale e l'altro più specifico. Questo legame è specificato rispettivamente attraverso le etichette *Iperonimo* e *Iponimo*, ad esempio *accessorio* è *Iperonimo* di *accessorio per intonato* e viceversa *accessorio per intonato* è *Iponimo* di *accessorio*;

²⁷² L'allegato ZA è un'appendice informativa in cui sono descritti i requisiti tecnici che i prodotti da costruzione devono soddisfare per ottenere la marcatura CE.

- relazione di associazione: correla coppie di concetti per i quali non è possibile stabilire vincoli gerarchici o preferenziali ma che presentano una forte implicazione reciproca. In questo lessico i termini inerenti ai prodotti da costruzione, sono stati associati ai termini concernenti le caratteristiche prestazionali, ad esempio al termine *palo* sono stati associati così come previsto dall'allegato ZA della norma di prodotto che lo interessa, termini come *resistenza a compressione*, *resistenza meccanica*, ecc..

L'insieme delle relazioni fin qui presentate, unitamente ad informazioni aggiuntive quali: la definizione estratta laddove presente dalla normativa di riferimento, la traduzione (principalmente in lingua inglese), la fonte del termine, la specificazione se trattasi o meno di un termine preferito, ecc., sono ospitate all'interno di una scheda terminologica appositamente predisposta, dinamica e interattiva che permette attraverso rinvii reciproci tra i termini di navigare all'interno della struttura classificatoria.

Per un'esemplificazione della struttura della scheda terminologica e delle informazioni al suo interno specificate si riporta il caso del termine *additivo*:

Termine preferito [TP]

Additivo

Traduzioni

Admixture

Definizione

Materiale aggiunto in piccole quantità per ottenere modificazioni specifiche delle proprietà.

(Fonte: UNI EN 998-2:2010 "Specifiche per malte per opere murarie. Parte 2: Malte da muratura")

Iponimi

Additivo accelerante di indurimento

Additivo accelerante di presa

Additivo accelerante di presa e fluidificante

Additivo accelerante di presa non alcalino

...

Termini correlati

Comportamento alla corrosione

Consistenza

Contenuto di ioni cloruro

Contenuto di solfuri

La predisposizione di una scheda terminologica così concepita consente di tenere traccia del contesto all'interno del quale un dato concetto si colloca, lo disambigua, lo considera in relazione agli altri concetti e dà evidenza del ruolo che il concetto è chiamato a svolgere all'interno della struttura sistematizzata.

Il lessico INNOVance è un sistema in divenire, l'intenzione è di continuare nella sua implementazione arricchendo ulteriormente la sua terminologia e la sua struttura, di modo da favorire un suo impiego al di là dei precisi scopi per il quale è stato realizzato.

CAPITOLO 4

METODOLOGIA PER LA REALIZZAZIONE DELLO SCHEMA DI CORRISPONDENZE

L'approccio metodologico adottato, volto a promuovere il dialogo tra i termini provenienti da risorse terminologiche e semantiche eterogenee è stato applicato al lessico specialistico strutturato INNOVance il quale come descritto nel capitolo 3, predispone la terminologia di settore impiegata per designare i prodotti da costruzione. Tale attività, condotta grazie all'ausilio dell'*editor Protégé*,²⁷³ ha beneficiato dell'analisi dello stato dell'arte per l'identificazione dei vocabolari verso i quali definire i collegamenti esterni. Nello specifico sono stati selezionati due tra i KOS disponibili sul web in lingua italiana e in formato linked data: il Nuovo Soggettario e la base di conoscenza DBpedia nella versione italiana.

La scelta dei vocabolari da impiegare nella definizione delle correlazioni semantiche, ha tenuto conto di una varietà di contesti, in cui il linguaggio proprio del settore viene quotidianamente impiegato. Esso infatti, è usato sia tra esperti in contesti formali e informali, che da soggetti, i quali pur non possedendo competenze specialistiche precise, vi si accostano per una molteplicità di motivazioni, cercando di recuperare informazioni di interesse in ragione di un dato bisogno informativo.

Nel presente capitolo sarà descritto l'approccio metodologico impiegato per definire i *mapping* esterni tra i termini del lessico INNOVance e i termini provenienti dal Nuovo Soggettario e da DBpediaItalia e illustrate le fasi che hanno portato alla definizione dei *mapping* esterni.

1. Il trasferimento del Lessico INNOVance in Protégé

Il Lessico INNOVance, come descritto nel precedente capitolo, fornisce la terminologia italiana in uso nel settore dell'edilizia, inerente ai prodotti da costruzione e ne propone contestualmente una rappresentazione strutturata, in cui i termini grazie alla

²⁷³ <<http://protege.stanford.edu/>> (ultima consultazione 16/12/2014).

specificazione di relazioni gerarchiche, di uguaglianza e associative sono inseriti in una rete semantica, dalla quale si evince il contesto in cui essi si collocano

A tal proposito, ciascuna entrata del lessico presuppone una scheda terminologica in cui possono essere riportate diverse informazioni: i sinonimi, gli iperonimi, gli iponimi, termini correlati, la fonte del termine, la traduzione in altre lingue, la definizione, ecc..

Date le finalità della presente ricerca, volte a sperimentare l'interoperabilità tra i termini appartenenti al lessico INNOVance e i termini provenienti da altre risorse terminologiche e semantiche, si è deciso di focalizzare l'attenzione sull'organizzazione sistematizzata dei termini, piuttosto che sull'apparato informativo predisposto per ciascuno di essi.



Figura 24. Scheda terminologica nel lessico INNOVance.

Gli elementi della scheda terminologica oggetto di interesse sono stati pertanto, i sinonimi, gli iperonimi, gli iponimi, i termini correlati e le definizioni laddove presenti, mentre non sono state al momento prese in considerazione (anche se non si esclude che in futuro queste informazioni possano essere coinvolte nel processo) le informazioni riguardanti, la fonte del termine, la fonte della definizione, la traduzione in lingua inglese e il dominio di appartenenza. Quest'ultimo pur ricoprendo un ruolo importante nella struttura gerarchica non è stato al momento considerato in quanto di recente introduzione e pertanto, si sta valutando a livello strutturale se sia opportuno o meno far confluire i termini in categorie generali mediante un approccio *bottom-up*, in

cui partendo dal singolo termine si costruisce una struttura classificatoria più generale. Il dominio infatti, non è un elemento consolidato all'interno dello schema previsto dal lessico INNOVance, il quale seppur completato nella parte inerente ai prodotti da costruzione da pochi mesi, in prospettiva si propone di andare avanti con l'implementazione, aggiungendo ulteriori termini inerenti ad altri aspetti del settore e arricchendo ulteriormente la sua struttura di modo da ampliare e superare i precipui scopi del progetto.

Prima di procedere alla definizione delle correlazioni esterne, le voci del lessico ritenute prioritarie, in quanto interessate da una maggiore articolazione strutturale, sono state trasferite sull'*editor Protégé* versione 4.3 impiegato anche in fase di specificazione dei *mapping* esterni. *Protégé* è una piattaforma *open source* principalmente impiegata nella realizzazione di ontologie, ma che ben si presta a questa tipologia di sperimentazioni. *Protégé* è stato sviluppato dall'Università di Stanford presso lo *Stanford Center for Biomedical Informatics* e fornisce un *framework* comune per la realizzazione di basi di conoscenza.

La scelta è ricaduta su questo *editor* in considerazione dell'interfaccia amichevole di cui dispone che lo rende adatto a qualsiasi fascia di utenza.

La prima fase ha riguardato quindi, il caricamento in *Protégé* dei termini del lessico INNOVance e delle relazioni che intercorrono tra di essi. Il trasferimento è stato realizzato manualmente innanzitutto perché il numero dei termini intervenuti in questo processo costituiscono un sotto-insieme della più vasta collezione terminologica prevista dal lessico INNOVance non ancora resa completamente disponibile poiché, la terminologia e la relativa struttura organizzativa è tuttora sottoposta a revisione e validazione da parte degli esperti di dominio; successivamente perché procedendo al trasferimento manuale delle relazioni si è acquisita una maggiore consapevolezza in merito ai criteri e alle regole impiegate per la loro definizione e si è potuto intervenire direttamente in taluni casi in cui sono state riscontrate alcune incongruenze nella struttura classificatoria.

L'*editor Protégé* prevede di specificare delle classi all'interno delle quali collocare gli oggetti di interesse.²⁷⁴ La prima operazione ha infatti riguardato, l'indicazione della classe *Lessico_INNOVance* a cui sono stati collegati i diversi termini che lo compongono e che nella logica di *Protégé* sono intesi come *Individuals*, ovvero insieme di unità che appartengono ad una certa classe.

²⁷⁴ «cfr.» MATTHEW HORRIDGE, *A practical guide to building OWL Ontologies using Protégé 4 and CO-ODE tools*, edition 1.3, University of Manchester, 2011.

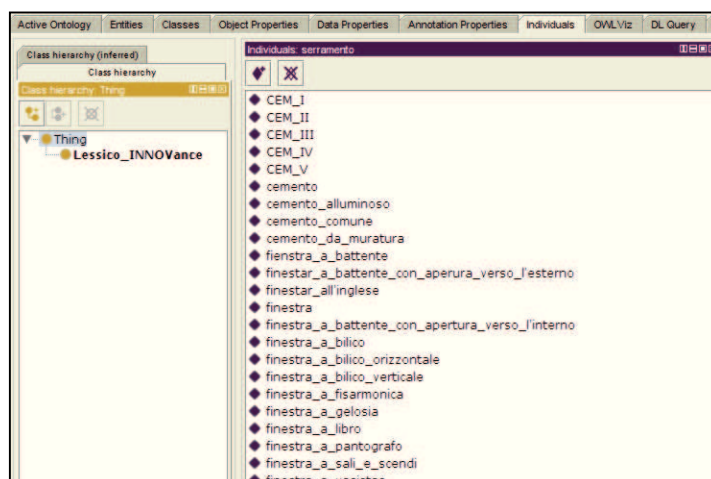


Figura 25. Definizione della classe Lessico_INNOVance.

Sono stati caricati dapprima i termini preferiti (TP), mentre i termini non preferiti (TNP), sinonimi e quasi sinonimi e infine i termini correlati che nella fattispecie richiamano le caratteristiche prestazionali che ciascun prodotto da costruzione soggetto a marcatura CE, deve garantire e rispettare prima di essere immesso sul mercato.

Successivamente, si è proceduto all'esplicitazione delle relazioni concettuali tra i termini, richiamando il linguaggio SKOS, lo standard del W3C che consente di esprimere la struttura interna di un qualsiasi sistema di organizzazione della conoscenza (sistemi di classificazione, thesauri, liste controllate, ecc.). SKOS non definisce le regole per lo sviluppo di KOS ma è un meta-modello che riflette i principi della più recente normativa in materia di thesauri e interoperabilità con altri vocabolari: la ISO 25964 *Information and Documentation – Thesauri and interoperability with other vocabularies*, la quale si compone delle seguenti parti:

- la Parte 1 - *Thesauri for information retrieval*: pubblicata nel 2011, fornisce le raccomandazioni in merito alla realizzazione e al mantenimento di thesauri monolingue e multilingue, nell'ottica di garantire l'interoperabilità in ambienti digitali;
- la Parte 2 - *Interoperability with other vocabularies*: pubblicata nel 2013, estende la sua competenza al *mapping* tra concetti appartenenti a vocabolari

differenti per favorire il recupero efficace ed efficiente di informazioni anche se indicizzate con vocabolari differenti.²⁷⁵

Tale norma che sostituisce i precedenti standard: la ISO 5964:1985 *Documentation - Guidelines for the establishment and development of multilingual thesauri* e la ISO 2788:1986 *Documentation - Guidelines for the establishment and development of monolingual thesauri*, introduce rispetto alle precedenti la differenza tra termine e concetto. Il concetto diviene l'entità principale da cui partire, il quale a sua volta può essere espresso ricorrendo ad uno o più termini e in una o più lingue.

Ritornando alle relazioni specificate esse provengono direttamente da *SKOS Core Vocabulary*,²⁷⁶ il vocabolario in cui sono specificati gli elementi attraverso i quali è possibile esprimere la struttura di un qualsiasi KOS.

Nello specifico, ci si è concentrati sulle proprietà riguardanti le relazioni semantiche di equivalenza, di gerarchia e di associazione e la proprietà per la documentazione (*documentation property*) inerente alla definizione.

Più dettagliatamente le etichette impiegate per specificare le relazioni semantiche tra i termini sono le seguenti:

- *skos:prefLabel*: introduce il termine preferito assegnato da un dato concetto:

ad es. *skos:prefLabel* "cemento comune";

- *skos:altLabel*: introduce l'etichetta non preferita, è assimilabile al termine non preferito (TNP) previsto dai thesauri e si propone di tenere traccia dei sinonimi, dei quasi-sinonimi e delle varianti lessicali di un dato concetto, quali acronimi o abbreviazioni:

ad es. *skos:prefLabel* "cemento comune"; *skos:altLabel* "CEM";

- *skos:broader*: introduce una relazione di tipo gerarchica, ovvero un concetto più generale e può essere esplicitato solo tra *skos:prefLabel*:

²⁷⁵ <<http://www.niso.org/schemas/iso25964/>> (ultima consultazione 16/12/2014).

²⁷⁶ <<http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20050510/>> (ultima consultazione 16/12/2014).

ad es. *skos:broader*: cemento;

- *skos:narrower*: introduce anch'esso una relazione gerarchica, ovvero un concetto più specifico e come *skos:broader* può coinvolgere solo *skos:prefLabel*:

ad es. *skos:broader*: cemento; *skos:narrower* cemento comune;

- *skos:related*: introduce la relazione di associazione e parimenti alle etichette che introducono una relazione di tipo gerarchico (*skos:broader*, *skos:narrower*) può essere stabilita esclusivamente tra *skos:prefLabel*:

ad. es. *skos:related*: durabilità.

- *skos:definition*: fornisce una descrizione completa del concetto:

ad. es. *skos:prefLabel* cemento

skos:definition "Legante idraulico, cioè un materiale inorganico finemente macinato che, quando mescolato con acqua, forma una pasta che fa presa e indurisce a seguito di reazioni e processi di idratazione e che, una volta indurita, mantiene la sua resistenza e la sua stabilità anche sott'acqua".

Accanto a queste etichette attraverso le quali è possibile esplicitare relazioni di natura semantica, SKOS ne prevede delle altre, grazie alle quali possono essere specificate ulteriori informazioni utili a circostanziare meglio un dato concetto:

- *skos:hiddenLabel*: è un'etichetta nascosta, utilizzata nel caso in cui i KOS debbano essere sottoposti ad un processo computazionale:

ad es. *skos:hiddenlabel* "durabilita" invece di "durabilità";

- *skos:scopeNote*: definisce i confini semantici di un dato concetto o specifica l'uso del concetto all'interno del KOS e svolge la medesima funzione della nota d'ambito prevista dai thesauri;

ad. es. *skos:prefLabel* cemento

skos:scopeNote “Materiali da costruzione a base di argilla e calcare”

- *skos:historyNote*: analogamente alla nota storica prevista dai thesauri, descrive eventuali rivisitazioni apportate al concetto in relazione al significato o alla forma:

ad. es. *skos:prefLabel* cemento

skos:historyNote Cemento.

Prima di specificare le relazioni in *Protégé*, è stato necessario effettuare una comparazione tra la struttura interna del lessico INNOVance e il modello previsto da SKOS. Tale confronto ha permesso di comprendere e valutare la fattibilità dell'allineamento tra i due modelli e identificare i punti di contatto.

Lessico INNOVance	SKOS
Sinonimi	<i>skos:altLabel</i>
Iperonimi	<i>skos:broader</i>
Iponimi	<i>skos:narrower</i>
Termini correlati	<i>skos:related</i>
Definizione	<i>skos:definition</i>

Figura 26. Confronto tra la struttura del lessico INNOVance e SKOS.

Da questa comparazione, è emerso che il linguaggio SKOS non fa altro che fornire in un formato *machine-understandable* le relazioni thesaurali tipiche. Le relazioni che collegano i termini sono state da considerarsi come annotazioni, in quanto consentono per l'appunto di specificare informazioni in merito ad una classe, agli individui e agli

oggetti in generale.²⁷⁷ Le relazioni, come mostrato nella figura seguente sono state definite a partire dal pannello *Object properties*.

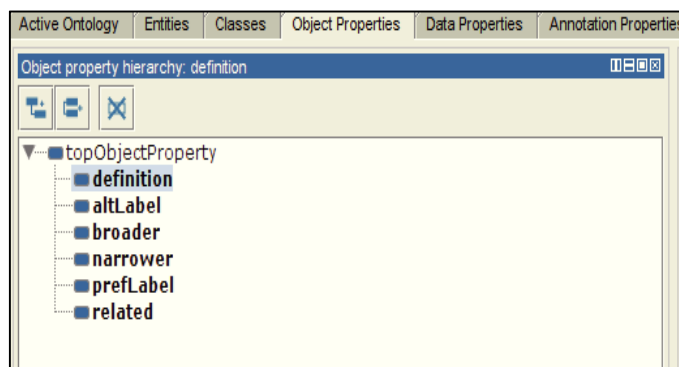


Figura 27. *Object properties* definite.

Terminato il caricamento dei termini e identificate le tipologie di relazioni che è possibile definire tra di essi, si è proceduto alla definizione vera e propria delle relazioni paradigmatiche e alla conseguente specificazione, laddove presente, della definizione che chiarisce il significato del concetto.

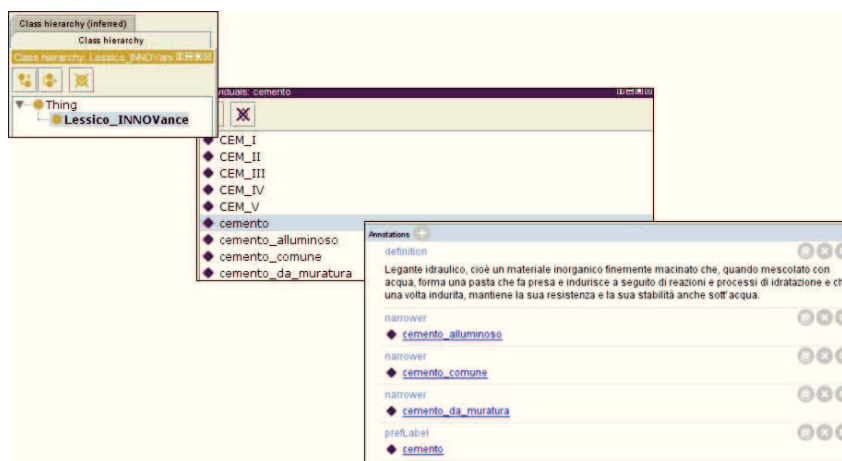


Figura 28. Specificazione delle relazioni per il termine *cemento*.

²⁷⁷ *Protege* prevede di specificare *Annotation properties*, ulteriori proprietà quali *Datatype properties* per definire una correlazione tra un individuo e il suo valore e *Object properties* che specificano relazioni tra due individui.

Al termine di questa fase i termini hanno riacquisito la conformazione strutturale originale prevista dal lessico INNOVance di partenza.

2. Identificazione delle risorse terminologiche e semantiche da mettere in comunicazione

L'identificazione dei vocabolari da allineare al Lessico INNOVance, ha tenuto conto della rassegna riguardante i sistemi di organizzazione e gestione della conoscenza presentata nel capitolo 3. A seguito della quale sono stati selezionati due KOS, il thesaurus del Nuovo Soggettario e la base di conoscenza DBpediaItalia, in ragione della loro ampia diffusione, del loro largo utilizzo del carattere multidisciplinare che li contraddistingue, della loro disponibilità in lingua italiana e della loro predisposizione nei formati previsti dal web semantico.

La decisione di impiegare questi sistemi di carattere multi-disciplinare e multi-dominio, è dovuta anzitutto all'assenza in ambito nazionale di una risorsa specifica per il dominio delle costruzioni, che funga da riferimento nella scelta dei termini da impiegare, non solo in contesti d'uso comune, tra i non addetti ai lavori, ma anche nella redazione dei documenti tecnici (capitolati speciali d'appalto, computi metrici estimativi, progetti esecutivi, ecc.) da parte degli esperti di dominio.

Di fatto, la mancata disposizione di una tale risorsa, determina all'interno del settore lo sviluppo di un linguaggio poco controllato, ambiguo e incoerente, in cui molto spesso i termini sono impiegati in maniera impropria.

A questa constatazione di fatto, si aggiungono ulteriori fattori quali: i continui aggiornamenti normativi, l'avvento di nuove tecnologie e materiali, l'eterogeneità dei soggetti coinvolti, la frammentarietà delle fasi del processo edilizio (progettazione, programmazione dei lavori, gestione, manutenzione e dismissione).

Pertanto, il linguaggio in uso nel settore delle costruzioni seppur di carattere specialistico è nato per rispondere alle esigenze comunicative di una comunità di pratica ristretta. Si tratta di un sistema dinamico che si evolve rapidamente, per cui è caratterizzato da un continuo mutamento che si manifesta attraverso la creazione di nuovi termini (neologismi); il superamento di alcuni di essi caduti in disuso perché obsoleti; la reinterpretazione di termini già esistenti e impiegati in altri contesti (terminologizzazione).

Dati questi presupposti, è evidente come riuscire a tenere traccia dei diversi usi del linguaggio sia una priorità anche in ambiti specialistici, seppur come ripreso in (Foppa Pedretti, 2010)²⁷⁸ essi si caratterizzano per esattezza, semplicità e chiarezza.

I due sistemi prescelti sono infatti, rappresentativi di due diversi registri linguistici: il Nuovo Soggettario è un linguaggio controllato, autorevole e impiegato in ambito bibliotecario e non solo, per indicizzare il contenuto delle risorse sia esse digitali che analogiche, DBpediaItalia invece è più propriamente una base di conoscenza che riflette il linguaggio d'uso comune, quindi è maggiormente indirizzata ad un utente medio. Entrambi, unitamente alla dimensione specialistica del lessico INNOVance se messi in comunicazione delineano i diversi aspetti di questo linguaggio, insieme infatti, possono rivestire un ruolo strategico nella gestione della terminologia di dominio poiché garantiscono la correlazione tra i diversi usi legati ad un medesimo linguaggio.

Ne consegue quindi un dialogo tra i diversi operatori del settore, il lessico INNOVance di carattere specialistico è prioritariamente di supporto agli esperti della materia, il Nuovo Soggettario in qualità di fonte autorevole è impiegato in contesti d'uso più formali mentre DBpediaItalia esplicita il linguaggio d'uso comune.

In aggiunta a questi fattori, è importante sottolineare che entrambe le risorse, seppur disponibili in italiano garantiscono il colloquio con altre lingue, il Nuovo Soggettario come già detto nel Capitolo 3 rinvia alla LCSH, a AGROVOC, a EUROVOC questi ultimi multilingue, mentre DBpediaItalia rinviando alle pagine Wikipedia attraverso il tag *"is foaf:primaryTopic of"* garantisce il colloquio con le altre lingue in cui l'entrata della base di conoscenza è predisposta, ad esempio la voce inerente a calcestruzzo in DBpediaItalia è disponibile in Wikipedia in più di dieci lingue. Questa caratteristica è molto importante ai fini del lessico INNOVance in quanto predispone, se attestata in un fonte autorevole, la traduzione in lingua inglese.

Il Nuovo Soggettario e la base di conoscenza DBpediaItalia, sono strumenti diversi tra loro, accomunati dal fatto che entrambi si propongono di tenere traccia dei diversi ambiti del sapere, sono flessibili e tentano di specializzare il loro patrimonio lessicale sulla base delle richieste provenienti dalle reali esigenze espresse dagli utenti. Il Nuovo Soggettario, è un linguaggio controllato che esemplifica un linguaggio, tiene traccia dei mutamenti culturali e dei termini ammessi e non ammessi attraverso rinvii reciproci ed è collegato ad altre risorse, elemento questo che conferisce alla sua struttura una maggiore potenzialità conoscitiva dei termini in esso ospitati.

²⁷⁸ «cfr.» CHIARA FOPPA PEDRETTI, *Linguaggi specialistici o LSP*, 2010, <<http://www.proz.com/translation-articles/articles/2831/1/Linguaggi-specialistici-o-LSP>>.

DBpediaItalia, d'altro canto pur non potendo essere annoverato tra i sistemi di organizzazione e gestione della conoscenza veri e propri, in quanto derivato da Wikipedia e quindi non concepito con il precipuo scopo di "mettere ordine", riflette la percezione di un utente medio, che nel caso del settore delle costruzioni potrebbe riferirsi ai clienti, ai consulenti che non sono esperti di dominio ma che per motivi professionali vi si accostano e necessitano di avere informazioni di tipo specialistico.

3. La scelta del modello da impiegare per il mapping tra le risorse

Terminata la fase di selezione dei vocabolari e prima di procedere alla specificazione dei *mapping* esterni, è stato importante definire la modalità attraverso la quale esplicitare le correlazioni semantiche. È stato cioè necessario stabilire il *source vocabulary* ovvero il vocabolario di partenza, e il *target vocabulary* ovvero i vocabolari verso i quali puntare e definire la natura delle relazioni.

In merito alla definizione e al mantenimento di *mapping* tra più risorse, interviene la ISO 25964-2:2013 summenzionata, la quale riporta le raccomandazioni per formulare collegamenti di questo tipo mostrando contestualmente gli elementi da considerare in sede di definizione dei collegamenti.

La norma inoltre, fa riferimento al suo interno a taluni modelli che possono essere impiegati nel caso in cui ci si appresti all'allineamento di concetti provenienti da più fonti:

- *structural unit model*: prevede che i vocabolari coinvolti nel processo di *mapping* siano simmetrici; questo modello è appropriato per i thesauri multilingue, in cui un dato concetto è identificato per mezzo di un termine preferito in tutte le lingue del thesaurus e la struttura delle relazioni gerarchiche e associative è la medesima;
- *direct-linked model*: può essere implementato nel caso in cui ad essere coinvolti nel *mapping* ci siano due o più vocabolari eterogenei non solo nella struttura (sistemi di classificazione, thesauri, liste controllate, ecc.) ma nella copertura semantica, nella lingua, negli obiettivi che si propongono. In tal caso è possibile specificare relazioni dirette (*direct-linked*) tra coppie di concetti provenienti da più vocabolari;

- *hub structure model*: propone di candidare a preferito, per l'appunto *hub*, uno dei vocabolari prescelti e mappare i concetti di questo verso i concetti presenti nei vocabolari che assumono la posizione di satelliti. Così facendo, tutti i concetti appartenenti all'*hub vocabulary* saranno correlati a quelli degli altri vocabolari, e inoltre, qualora le relazioni venissero stabilite anche in senso opposto, cioè dai vocabolari satelliti verso l'*hub vocabulary*, tutti i concetti saranno relazionati tra di loro creando una vera e propria rete di relazioni;
- *selective mapping model*: interviene qualora non sia necessario mappare completamente l'intera collezione dei concetti presenti nei vocabolari ma solo quelli ritenuti utili agli scopi che si intendono perseguire. Questo tipo di modello è previsto qualora la sovrapposizione tra due o più vocabolari sia limitata ad alcuni concetti o nel caso in cui il *mapping* sia definito a partire da un indice o un catalogo.

In considerazione della natura e della struttura proposta dalle tre risorse coinvolte nel processo di *mapping*, si è deciso di ricorrere alla definizione di un modello combinato basato sulla compresenza di quelli sopraelencati: *direct-linked model*, *hub structure model*, *selective mapping model*, al fine di identificare un modello che tenga conto dei rispettivi punti di forza e che consenta di superarne i limiti.

Di fatto, la norma non impone l'uso di un modello piuttosto che un altro, anzi, incentiva la loro combinazione in considerazione degli obiettivi precisi che ci si propone, della tipologia di vocabolari coinvolti, della natura delle relazioni che si intende esplicitare, della direzione che i collegamenti intendono seguire.

Il modello che ne è scaturito, ha tenuto principalmente conto delle differenze tra le tre risorse coinvolte e che riguardano il contesto e le finalità per cui sono stati concepiti e la non coincidenza della copertura semantica. Come già detto, il lessico INNOVance è di tipo specialistico, in quanto legato ad un dominio del sapere ben preciso, mentre il Nuovo Soggettario e DBpediaItalia seppur condividono il carattere multidisciplinare, non possono essere equiparati dal punto di vista della struttura né nelle funzioni che sono chiamati ad assolvere.

Questi fattori, hanno portato alla definizione di un modello combinato che partendo dalla concezione proposta dall'*hub structure model*, integra le potenzialità degli altri due: il *direct-linked model* è infatti maggiormente appropriato qualora si voglia stabilire un *mapping* diretto tra coppie di concetti mentre il *selective mapping model*, è una

conseguenza dei primi due in quanto interviene nel caso in cui i vocabolari non condividano la medesima copertura semantica.

L'approccio metodologico adottato, conferisce quindi lo status di *hub* (centro) al lessico INNOVance mentre il Nuovo Soggettario e DBpediaItalia assumono la posizione di vocabolari satelliti. I *mapping* sono stati a loro volta stabiliti tra coppie di concetti in maniera diretta e secondo associazioni di tipo unidirezionale, prevedendo quindi che ciascun concetto del lessico INNOVance possa esser associato ad uno o più concetti provenienti dagli altri due vocabolari considerati, di modo da identificare negli altri vocabolari almeno un concetto a cui collegarsi.

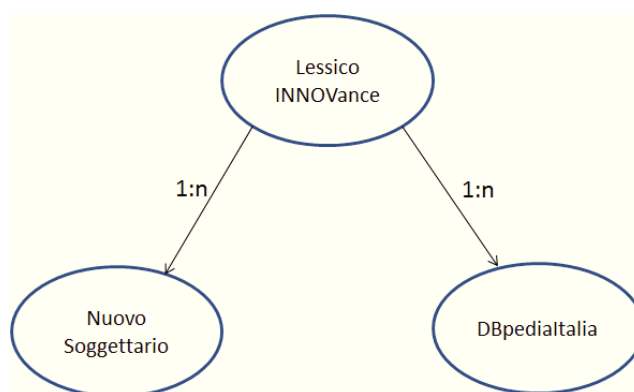


Figura 29. Modello prescelto.

La Figura 29 mette in evidenza la centralità del lessico INNOVance poiché è la risorsa da cui partire e specifica che le relazioni possono essere del tipo uno-a-molti e cioè che dato un termine del lessico INNOVance esso può essere correlato ad uno o più termini degli altri vocabolari coinvolti nel processo di *mapping*, in realtà si tratta solo di una condizione ideale in quanto per ciascun termine è stato identificato un solo termine di arrivo nel Nuovo Soggettario e in DBpediaItalia.

La fase successiva, ha interessato la specificazione della natura dei *mapping* da stabilire, i quali, data l'eterogeneità e il diverso grado di specializzazione con cui i concetti inerenti ai prodotti da costruzione sono trattati nei tre vocabolari, sono stati definiti in conformità alla normativa vigente ISO 25964-2:2013, la quale prevede, analogamente a quanto avviene all'interno di un medesimo thesaurus, *mapping* di vario tipo:

- *equivalence mapping*: stabilito contrariamente a quanto avviene all'interno di uno stesso thesaurus non tra un termine preferito (TP) e un termine non preferito (TNP) ma tra concetti presenti in vocabolari diversi. Questo tipo di *mapping* che si esprime tramite EQ, può essere a sua volta specializzato sulla base del grado di equivalenza in:

- equivalenza esatta (=EQ): si specifica quando i concetti provenienti dai diversi vocabolari possono essere impiegati in maniera interscambiabile, in quanto veicolano esattamente lo stesso concetto:

ad. es. Voc1:Finestra all'inglese **=EQ** Voc2:Finestra a battente con apertura verso l'esterno

- equivalenza inesatta (~EQ): quando i concetti non sono completamente sovrapponibili e sono quindi equivalenti in alcuni contesti e non in altri:

ad es. Es. Voc1:Canaletta di drenaggio **~EQ** Voc2: Cunetta

- *hierarchical mapping*: stabilita tra concetti legati tra loro sulla base di criteri di classificazione e collocati in una rete di sovra e sotto ordinati gerarchici. La struttura gerarchica lega tra loro concetti in cui il sovra-ordinato rappresenta la classe o il tutto mentre il sotto-ordinato ne costituisce un elemento o una parte. La sigla che definisce la classe generale è BM (*Broader Mapping*), mentre gli elementi in essa contenuti sono preceduti da NM (*Narrower Mapping*). Il *mapping* di tipo gerarchico analogamente alle relazioni gerarchiche definite all'interno di un medesimo thesaurus, possono essere generiche (genere-specie) esplicitate dalle sigle BTG (*Broader Term Generic*) e NTG (*Narrower Term Generic*), partitive (parte-tutto) espressa tramite le sigle BTP (*Broader Term Partitive*) e NTP (*Narrower Term Partitive*) ed esemplificative (classe-istanza) le cui sigle sono BTI (*Broader Term Instantial*) e NTI (*Narrower Term Instantial*):

Ad es. Es. Voc1: Finestra **BM** Voc2:Serramenti
Voc2:Serramenti **NM** Voc1: Finestra

- *related mapping*: stabilita qualora non sia possibile specificare relazioni gerarchiche e/o preferenziali mette in comunicazione concetti che hanno una forte implicazione reciproca. Tale relazione è espressa con la sigla RM (*Related Mapping*):

ad es. Voc1:Clinker **RM** Voc2:Cemento comune

Tali relazioni predisposte dalla ISO 25964-2:2013 trovano una parziale rispondenza nel linguaggio SKOS, il quale predispone un set di *tag* per definire *mapping* tra concetti appartenenti a sistemi di organizzazione e gestione della conoscenza differenti favorendo l'allineamento e incoraggiando il colloquio tra KOS eterogenei.

I tag predisposti da SKOS per esprimere le relazioni summenzionate sono i seguenti:

- *skos:closeMatch*: stabilisce che tra due concetti appartenenti ad altrettanti KOS sussiste dal punto di vista del significato una similarità molto forte;
- *skos:exactMatch*: designa una specializzazione del primo e indica che dati due concetti provenienti da altrettanti KOS, essi possano essere impiegati in maniera interscambiabile indipendentemente dal contesto in cui sono calati;
- *skos:broadMatch*: introduce un rapporto di sovra-ordinazione tra concetti che esprimono una classe;
- *skos:narrowerMatch*: introduce concetti più specifici rispetto alla sua controparte *skos:broadMatch*;
- *skos:relatedMatch*: permette di gestire le altre tipologie di relazioni che non rientrano in quelle gerarchiche e di equivalenza.

Queste relazioni hanno supportato il processo di *mapping* tra i concetti appartenenti a fonti eterogenee, con l'unica eccezione che ha riguardato la scelta di utilizzare per

stabilire un *equivalence mapping*, il solo tag *skos:closeMatch* perché concettualmente più ampio del tag *skos:exactMatch*.

4. Definizione dei mapping tra il lessico INNOVance, il Nuovo Soggettario e DBpediaItalia

Una volta selezionati i vocabolari da impiegare nel processo di *mapping*, identificato il modello da implementare e le relazioni che si intendono specificare, si è proceduto alla comparazione e all'analisi dei concetti dei tre vocabolari considerati, al fine di identificare i punti di contatto e stabile tra di essi dei *mapping* su base semantica in cui ci si è concentrati sul concetto piuttosto che sul termine.

L'identificazione e la verifica di questi collegamenti è stata effettuata manualmente per due ordini di motivi: il primo è che i concetti e i relativi termini del lessico INNOVance afferiscono ad un dominio della conoscenza altamente specifico e granulare, ragion per cui necessitano di essere attentamente vagliati prima di essere collegati a concetti ospitati in altri vocabolari; il secondo invece è direttamente riconducibile alla consistenza dei concetti lessico INNOVance considerati, attestata intorno alle 500 voci per quanto concerne la parte inerente ai prodotti da costruzione,

Una siffatta operazione, ha garantito una maggiore precisione nella scelta dei concetti da correlare in considerazione del fatto che qualora non sia stato rintracciato nel *target vocabulary* un'equivalenza esatta si è ricorso all'individuazione di *hierarchical mapping* o nei casi più estremi *related mapping*.

A proposito dei *hierarchical mapping*, (Doerr, 2001)²⁷⁹ sostiene come questa specificazione debba essere attentamente valutata nel senso che, la *broader equivalence* dovrebbe essere minimale per cui il termine attestato nel *target vocabulary* dovrebbe essere il meno ampio possibile mentre la *narrower equivalence* dovrebbe essere massimale, per cui il termine del *target vocabulary* dovrebbe essere il più ampio possibile.²⁸⁰

Il processo di *mapping* si è proposto di stabilire per ciascun concetto del lessico INNOVance almeno una correlazione rispetto al Nuovo Soggettario e a DBpediaItalia, in virtù del principio proposto dal modello *direct-linked* di cui sopra, il quale prevede

²⁷⁹ «cfr.» MARTIN DOERR, *Semantic Problems of Thesaurus Mapping*, «Journal of Digital Information», vol. 1, n. 8, 2001, <<https://journals.tdl.org/jodi/index.php/jodi/article/view/31/32>> (ultima consultazione 16/12/2014).

²⁸⁰ ELISABETTA VITI, *Op. cit.*, p. 175.

che l'esplicitazione dei *mapping* coinvolga tutti i concetti di un dato vocabolario. A questa specificazione si aggiunge il carattere selettivo della metodologia che non ha preso come riferimento l'intera collezione dei concetti espressi nel Nuovo Soggettario e in DBpediaItalia ma li ha considerati in parte, ovvero solo per le voci che via via sono state individuate come di interesse.

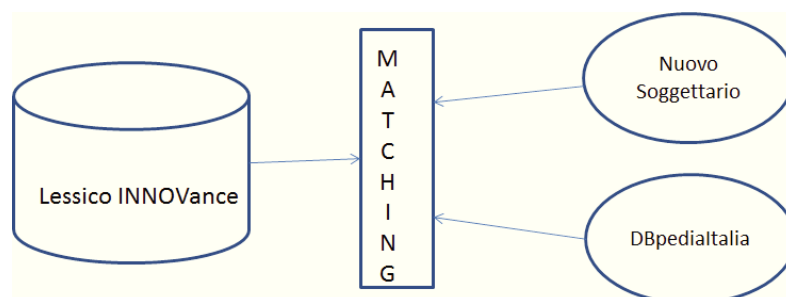


Figura 30. Esempificazione della modalità di *matching* adottata.

La specificazione dei *mapping* tra i concetti del lessico INNOVance e quelli ospitati negli altri vocabolari considerati, ha richiesto un'attenta analisi non solo del concetto in quanto tale, ma anche del contesto in cui ciascuno di essi è calato, di modo da realizzare un sistema di allineamento quanto più coerente.

Lo studio della sistematizzazione che interessa ciascun concetto è avvenuta contestualmente alla specificazione dei *mapping* al fine di individuare e comprendere nel dettaglio i principi e i criteri di classificazione che sono stati privilegiati. Da questa verifica è emerso come ad esempio il lessico INNOVance nella sua struttura organizzativa abbia la tendenza a prediligere un solo criterio, ad esempio nel caso delle *finestre* è stato privilegiato il senso di apertura per cui avremo *finestra a vasistas*, *finestra a battente*, ecc, e di come al contrario il Nuovo Soggettario per lo stesso termine impieghi due diversi criteri: il senso di apertura *finestra a sporto* e il materiale *finestra di legno* che di per sé non è discriminate rispetto al concetto di finestra.

Il *mapping* può essere in questa sede inteso come il risultato di un processo intellettuale, che partendo dall'analisi dei singoli concetti si è proposto di selezionare le correlazioni maggiormente coerenti e significative. Tale processo è stato supportato dal già citato editor *Protégé* che ha consentito di tener traccia di questi collegamenti.

La definizione delle correlazioni in *Protégé* è stata possibile introducendo a fianco della classe `Lessico_INNOVance`, due nuove classi, rispettivamente inerenti al Nuovo

Soggettario e a DBpediaItalia, e all'interno di queste classi sono stati ospitati solo i concetti che sono intervenuti nel *mapping*.



Figura 31. Le tre classi rappresentative dei tre vocabolari.

Una volta definite le classi, sono state predisposte secondo il formalismo proposto da SKOS le diverse tipologie di *mapping* che è possibile specificare: *skos:closeMatch*, *skos:broadMatch*, *skos:narrowerMatch* e *skos:relatedMatch*, che sono andate ad affiancare quelle precedentemente specificate per correlare i termini all'interno del lessico INNOVance.

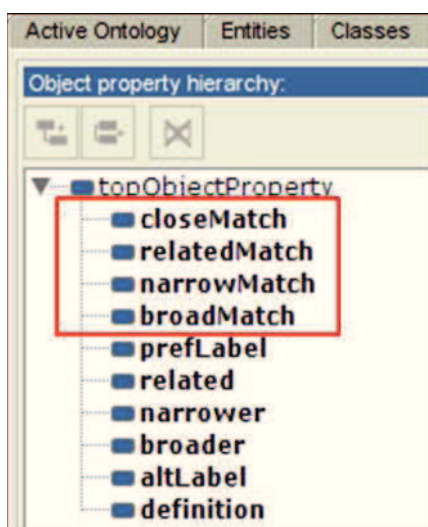


Figura 32. I tag per esprimere i *mapping* esterni.

A seguire è stata avviata la procedura di esplicitazione dei *mapping* sulla base dei criteri summenzionati e che ha portato ad arricchimento dei concetti del lessico INNOVance.



Figura 33. Mapping esterni e relazioni interne.

La figura mette in evidenza come il concetto di cemento espresso nel lessico INNOVance, contempra non soltanto le relazioni interne che lo specializzano in *cemento comune*, *cemento da muratura* e *cemento speciale* ma anche i *mapping* esterni dove *Cementi* rinvia al Nuovo Soggettario e *Cemento* rinvia a DBpediaItalia.

In merito a questo concetto, il Nuovo Soggettario fornisce una visione maggiormente strutturata e ricca evidenziata dai numerosi termini che vi sono stati correlati per mezzo della relazione associativa mentre DBpedia, data la sua natura fornisce un abstract di tipo divulgativo e rinvia alla relativa pagina Wikipedia maggiormente informativa. Un estratto della presentazione delle correlazioni semantiche stabilite in SKOS e inerenti il termine Cemento, è disponibile in (Appendice).

CONCLUSIONI

L'attività di ricerca condotta, si è proposta di mettere in evidenza il ruolo strategico che l'interoperabilità, soprattutto nella sua dimensione semantica, riveste nella gestione e nell'organizzazione di dati, informazioni e conoscenza.

All'interno di questo contesto si è cercato di delineare uno schema di corrispondenze su base semantica tra termini provenienti da fonti eterogenee, sfruttando le buone pratiche dell'approccio linked data e utilizzando SKOS, in qualità di standard *de facto* per il web semantico, in materia di sviluppo e allineamento interoperabile di vocabolari.

La metodologia adottata per l'esplicitazione di *mapping*, seppur basata su approcci definiti in altri contesti, trova il suo carattere innovativo nel dominio a cui è stata applicata e nella natura delle correlazioni che sono state specificate.

Nel panorama nazionale infatti, non sono state riscontrate iniziative condotte in tal senso, che coinvolgano il linguaggio in uso nel settore delle costruzioni e né tanto meno che si preoccupino di indagare il diverso grado di specializzazione della conoscenza di dominio via via che ci sposta dal livello specialistico più dettagliato e preciso, a quello generale che risulta essere maggiormente condiviso, soprattutto tra i non esperti di dominio.

Accanto a questa considerazione, vi è quella concernente la tipologia delle relazioni che sono state specificate. Non si è trattato unicamente di predisporre relazioni di equivalenza (*skos:closeMatch*), ma di prevedere e all'occorrenza specificare ulteriori relazioni di tipo gerarchico (*skos:broadMatch*, *skos:narrowerMatch*) e associativo (*skos:relatedMatch*), grazie alle quali è stato possibile esplicitare per ciascuna entrata del lessico INNOVance almeno un corrispettivo nel Nuovo Soggettario e in DBpediaItalia.

L'attività di *mapping* ha dato luogo ad uno schema all'interno del quale, similmente a quanto avviene all'interno del thesaurus, i termini delle tre risorse sono organizzati sia sul piano orizzontale che verticale. Uno schema così concepito, può essere impiegato per mettere in relazione l'insieme di risorse indicizzate e collegate ai termini oppure essere impiegato come *training vocabulary* per scoprire nuove relazioni di dominio e

limitare l'intervento dell'esperto umano in fase di validazione dei collegamenti automaticamente estratti.

La definizione dei *mapping* ha richiesto l'analisi del linguaggio di dominio e lo studio dei criteri e delle modalità attraverso cui esso viene strutturato e definito nel Nuovo Soggettario e in DBpediaItalia. La comparazione tra il lessico INNOVance e le altre risorse considerate, ha messo in evidenza talune differenze riguardanti tanto il piano puramente lessicale (differenze morfologiche) quanto quello concettuale (articolazione classificatoria). A tali difformità è stato possibile sopperire predisponendo lo schema di allineamento, che in quanto tale ha lasciato inalterata, la struttura e la consistenza delle tre risorse messe a confronto prima, e allineate poi.

In molti casi l'allineamento è stato diretto, per cui la stringa lessicale è stata sufficiente a definire una correlazione del tipo *skos:closeMatch*, come nel caso del termine *cemento* nel lessico INNOVance e *Cementi* del Nuovo Soggettario. Questa condizione non è stata tuttavia sempre possibile, ragion per cui, si è ritenuto necessario procedere ad un'analisi quanto più accurata possibile del termine e del concetto da esso veicolato, di modo da identificare nelle altre risorse almeno un corrispettivo che designasse un concetto più generale o in alternativa un concetto più specifico. A titolo esemplificativo si riporta il caso del termine *cemento composito* del lessico INNOVance per il quale non è stato possibile specificare un rapporto di equivalenza né nel Nuovo Soggettario, né in DBpediaItalia, ma è stata stabilita una relazione del tipo *skos:broadMatch*. In più casi si è dovuto ricorrere a questa strategia giustificata dal fatto che le tre risorse messe a confronto prevedono un variegato grado di specializzazione della conoscenza inerente al settore oggetto di indagine.

In taluni casi estremi ovvero laddove non sia stato possibile definire un rapporto di equivalenza o di gerarchia si è optato per la specificazione di una relazione associativa, designata tramite *skos:relatedMatch*, ad esempio il termine complesso *sistema di ritenuta stradale* del lessico INNOVance è stato correlato tramite questa relazione a *Strade* nel Nuovo Soggettario.

La terminologia dei prodotti da costruzione è stata allineata ai termini del Nuovo Soggettario e a quelli di DBpediaItalia, in prospettiva ci si propone di proseguire con l'attività e di estendere l'allineamento ai termini riguardanti le caratteristiche prestazionali attualmente presenti nel lessico INNOVance e a tutti gli altri termini che via via andranno ad incrementare la base terminologica ad oggi disponibile, con l'obiettivo ultimo di renderlo fruibile in modalità linked data.

La sperimentazione dell'allineamento condotta, malgrado le dissimilitudini tra le tre risorse coinvolte - riguardanti sia la struttura che la copertura semantica - e le difficoltà incontrate nella definizione di siffatte correlazioni, ha dato dei risultati incoraggianti, pertanto nel prosieguo ci si propone di raffinare la metodologia impiegata, di potenziare le strategie di aggiornamento e di allargare il panorama delle risorse verso cui mappare.

La combinazione della conoscenza di dominio e la sua rappresentazione secondo regole e principi comuni fanno sì che essa possa essere interpretata in maniera condivisa tra gli esperti di dominio e i non addetti ai lavori.

BIBLIOGRAFIA

AGOSTI MARISTELLA, FERRARI NICOLA, *Interoperabilità tra sistemi di biblioteche digitali*, «Digitalia» vol. 1, n. 1, 2010. ISSN: 1972-621X <<http://digitalia.sbn.it/article/view/258/169>>.

ANTONIOU GRIGORIS, VAN HARMELEN FRANK, *A Semantic Web Primer*, Cambridge, The MIT Press, 2004. ISBN 0-262-01210-3.

ARANDA-MENA GUILLERMO, WAKEFIELD RON, *Interoperability of building information: myth or reality?*, in *eWork and eBusiness, Engineering and Construction*, a cura di Manuel Martinez, Raimar Scherer, London, Taylor & Francis Group, 2006, p. 127-133, <<http://mams.rmit.edu.au/lrssi7jd7nd4.pdf>>.

ASSOCIAZIONE ITALIANA PER L'OPEN GOVERNMENT, *Come di fa Open Data? Istruzioni per l'uso per Enti e Amministrazioni Pubbliche*, a cura di Ernesto Belisario, Gianluigi Cogo, Stefano Epifani, Claudio Forghieri, 2011, <<http://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.retecamere.it%2Fdownload%2F180.html&ei=ZlXoVN35E6vcywOmgYFA&usg=AFQjCNFCvoRf2SEbsnmNoMRBFXgw8LLn6w&bvm=bv.79142246,d.bGQ>>.

BACIGALUPI VINCENZO, *Dizionario ragionato di termini tecnici dell'edilizia*, Roma, Kappa, 1998. ISBN 8878903108.

BERNERS-LEE TIM, et alii, *The Semantic Web*, «Scientific American», May 2001, p. 29-37.

BIANCHINI CARLO, *Dalla professione all'etica. Il servizio di reference come esperienza esistenziale*, in *Il servizio di reference*, a cura di Carlo Bianchini, Firenze, Le Lettere, p. XIX-XLII. ISBN 9788860873095 <<http://eprints.rclis.org/16294/1/Dalla%20professione%20all%20etica%20Prefaz%20a%20Il%20servizio%20di%20reference.pdf>>.

BIANCHINI CARLO, GUERRINI MAURO, *Introduzione a RDA*, Milano, Editrice Bibliografica, 2014, (Collana: Biblioteconomia e scienza dell'informazione, 3). ISBN/EAN 9788870757804.

BIZER CHRISTIAN, et alii, *Linked Data - The Story So Far*, «International Journal on Semantic Web and Information Systems», vol. 5, n. 3, 2009, p. 1-22, DOI 10.4018/jswis.2009081901, <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>.

BIZER CHRISTIAN, et alii, *Linked Data on the Web (LDOW2008)*, Workshop Summary at the 17th International World Wide Web Conference, Beijing, China, April 21-25, 2008, p. 1265-1266.

BONANNI LAURA, *L'Authority file*, ICCU, 2011, <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/sbn_notizie_1998_2001/pagina_185.html;jsessionid=1809496D6B6E95FE5B32ABF089034FDD>.

BROUGHTON VANDA, *Costruire thesauri: Strumenti per indicizzazione e metadati semantici*, edizione italiana a cura di Piero Cavalieri, traduzione di Laura Balestra e Luisa Venuti, Milano, Editrice Bibliografica, 2006. ISBN 978-88-7075-674-6.

CARACCILO MARINA, et alii, *Thesaurus maintenance, alignment and publication as linked data: the AGROVOC use case*, «Int. J. Metadata, Semantics and Ontologies», vol. 7, n. 1, 2012, <http://eprints.rclis.org/17735/1/IJMSO%207_1_Paper%206_PublishedVersion%5B1%5D.pdf>.

CHAKRABARTI SOUMEN, *Mining the Web: Discovering Knowledge from Hypertext Data, Parte 2*, San Francisco, Morgan Kaufmann, 2003. ISBN: 1-55860-754-4.

CHETI ALBERTO, *Manuale ipertestuale di analisi concettuale*, 1996, <http://biblioteche.unibo.it/manuals/html_1/HOME.HTML>.

CHITI EDOARDO, GULDO RICCARDO, *Il regime linguistico dei sistemi comuni europei*, numero monografico della «Rivista di Diritto Pubblico», Milano, Giuffrè, 2008. ISBN: 88-14-14191-6 - EAN: 9788814141911

COGNETTI GAETANA, Recensione a *The LCSH Century: One Hundred Years with the Library of Congress Subject Headings System*, « Bollettino AIB», vol .42, n. 1, 2002. URN:NBN:IT:AIB-5257 <<http://bollettino.aib.it/article/view/7155>>.

COMMISSIONE DI COORDINAMENTO SPC, *Linee guida per l'interoperabilità semantica attraverso i linked open data*, 2012, <http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>

COYLE KAREN, *Library Linked data: un'evoluzione*, «JLIS.it», vol. 4, n. 2, Gennaio 2014, DOI:10.4403/jlis.it-5443, <http://kcoyle.net/presentations/FI2012_it.pdf>.

DI DONATO FRANCESCA, *Lo stato trasparente. Linked open data e cittadinanza attiva*, Pisa, ETS, 2010. ISBN: 9788846728876 <<http://www.linkedopendata.it/wp-content/uploads/statotrasparente.pdf>>.

DI NOIA TOMMASO, et alii, *Semantic web. Tra ontologie e Open Data*, Milano, Apogeo, 2013. ISBN: 8850332149 - ISBN-13: 9788850332144.

DOERR MARTIN, *Semantic Problems of Thesaurus Mapping*, «Journal of Digital Information», vol. 1, n. 8, 2001, <<https://journals.tdl.org/jodi/index.php/jodi/article/view/31/32>>.

DORATI ANTONELLA, COSTANTINI STEFANIA, *Approcci al Web Semantico*, <<http://www.websemantico.org/articoli/approcciwebsemantico.php>>.

ESCOLANO RODRIGUEZ ELENA, *L'adattamento di ISBD al web semantico dei dati bibliografici espressi in linked data*, «JLIS.it 3.2», DOI: 10.4403/jlis.it-5484, 2012.

EUROPEAN COMMISSION, *European Interoperability Framework for pan-European egovernment services*, Belgium, 2004. ISBN 92-894-8389-X.

EUZENAT JÉRÔME, et alii, *Ontology Alignment Evaluation Initiative: Six Years of Experience*, «Journal on Data Semantics XV», Berlin, Springer Verlag, 2011 (Collana: Lecture Notes in Computer Science, 6720), p. 158-192, DOI: 10.1007/978-3-642-22630-4_6, <<http://disi.unitn.it/~p2p/RelatedWork/Matching/OAEI-JODS.pdf>>.

FEIGENBAUM LEE, et alii, *The Semantic Web in Action*, «Scientific American», vol. 297, December 2007, p. 90-97, <<http://thefigtrees.net/lee/sw/sciam/semantic-web-in-action>>.

FERRAGINA PAOLO, *Sui motori di ricerca*, in *Documenti Digitali*, a cura di Roberto Guarasci, Antonietta Folino, Milano, ITER, 2013, p. 331-370. ISBN 9788890341939.

FOLINO ANTONIETTA, *Tassonomie e thesauri*, in *Documenti Digitali*, a cura di Roberto Guarasci, Antonietta Folino, Milano, ITER, 2013, p. 387-444. ISBN 9788890341939.

FOPPA PEDRETTI CHIARA, *Linguaggi specialistici o LSP*, 2010, <<http://www.proz.com/translation-articles/articles/2831/1/Linguaggi-specialistici-o-LSP>>.

GILCHRIST ALAN, VERNAU JUDI, *Facets of Knowledge Organization: Proceedings of the ISKO UK Second Biennial Conference, 4th - 5th July, London, 2011*, Emerald Group Publishing Limited, 2012. ISBN-10: 1780526148. ISBN-13: 978-1780526140.

GNOLI CLAUDIO, *Connettere la conoscenza: quattro livelli di interoperabilità*, «Biblioteche oggi», vol. 32, n. 5, 2014, p. 9-16, <http://biblioteche.provincia.bergamo.it/portali/c/document_library/get_file?p_l_id=22502&folderId=271699&name=DLFE-9361.pdf>.

GNOLI CLAUDIO, *Glossario di organizzazione della conoscenza*, 2011, <<http://www.iskoi.org/doc/glossario.htm>>.

GNOLI CLAUDIO, *Il tavolino di Ranganathan*, «Bibliotime», a. III, n. 3, novembre 2000, <<http://www.aib.it/aib/sezioni/emr/bibtime/num-iii-3/gnoli.htm#5leggi>>.

GUAGLIANONE MARIA TERESA, *MNEMO. Una metodologia combinata per l'acquisizione e la modellizzazione della conoscenza delle organizzazioni*, Tesi di Dottorato, Università della Calabria, 2012.

GUERRINI MAURO, *La biblioteca spiegata agli studenti universitari*, Milano, Editrice Bibliografica, 2012 (Collana: Conoscere la biblioteca, 8). ISBN/EAN 9788870757231.

GUERRINI MAURO, POSSEMATO TIZIANA, *Linked data: un nuovo alfabeto del web semantico*, «Biblioteche oggi», vol. 30, n.3, 2012, p. 7-15; <<http://www.bibliotecheoggi.it/content/201200300701.pdf>>.

HASSANZADEH OKTIE, et alii, *Data Management Issues on the Semantic Web*, in *ICDE '12 Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*, Washington, DC, USA, IEEE Computer Society, 2012, p. 1204-1206, DOI 10.1109/ICDE.2012.141, <<http://disi.unitn.it/~velgias/docs/HassanzadehKV12.pdf>>.

HEATH TOM, BIZER CHRISTIAN, *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool, 2011, DOI 10.2200/S00334ED1V01Y201102WBE001.

HERSH WILLIAM, *Information Retrieval. A health and biomedical perspective*, «Journal of the Medical Library Association», vol. 93, n. 1, New York, NY: Springer, 2005² (Collana: Health Informatics Series), p. 134-135. ISBN 0-387-95522-4.

HODGE GAIL, *Systems of knowledge organization for digital libraries: Beyond traditional authority files*, Washington, DC, The Digital Library Federation Council on Library and Information Resources, 2000. ISBN 1-887334-76-9, <<http://www.clir.org/pubs/reports/pub91/pub91.pdf>>.

HORRIDGE MATTHEW, *A practical guide to building OWL Ontologies using Protégé 4 and CO-ODE tools, edition 1.3*, University of Manchester, 2011

IFLA, *International Standard Bibliographic Description*, Edizione italiana a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche, Roma, 2012. ISBN 978-88-7107-131-2, <http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2012/ISBD_NOV2012_online.pdf>.

INTERNATIONAL ALLIANCE FOR INTEROPERABILITY, *An Introduction to the International Alliance for Interoperability and the Industry Foundation Classes*, 1999, <http://iaiweb.lbl.gov/Resources/IFC_Releases/IFC_Release_2.0/BETA_Docs_for_Review/IFC_R2_IntroToIAI_Beta_d2.PDF>.

ISO 25964-1:2011 *Information and documentation – Thesauri and interoperability with other vocabularies - Part1: Thesauri for information retrieval*.

ISO/IEC 2382-1:1993(EN) *Information technology – Vocabulary – Part 1: Fundamental terms*, 1993.

JANOWICZ KRZYSZTOF, et alii, *Five Stars of Linked Data Vocabulary Use*, «Semantic Web 0», IOS Press, 2014.

KOIVUNEN MARJA-RIITTA, MILLER ERIC, *W3C Semantic Web Activity*, in *Proceedings of the Semantic Web Kick-off Seminar*, Finland, 2nd November 2001, 2002, <<http://www.w3.org/2001/12/semweb-fin/w3csw>>.

LACASTA JAVIER, et alii, *Terminological Ontologies: Design, Management and Practical Applications*, «Semantic Web and Beyond», vol. 9, Springer, 2010. ISBN-13: 9781461426448 - ISBN-10: 1461426448

LEI ZENG MARCIA, SALABA ATHENA, *Toward an International Sharing and Use of Subject Authority Data*, FRBR Workshop, OCLC, 2005.

LENCI ALESSANDRO, et alii, *Testo e computer*, Roma, Carocci Editore, 2005, (Collana: Studi Superiori, 829). ISBN 9788843034253.

LUCARELLI ANNA, *Esperienze di linked data e web semantico in BNCF*, 6^a Incontro ISKO Italia Firenze 20 maggio 2013, <<http://www.iskoi.org/doc/firenze13/lucarelli.pdf>>.

MANNING CHRISTOPHER D., et alii, *Introduction to Information Retrieval*, Cambridge, Cambridge University Press, 2007. ISBN-10: 0521865719. ISBN-13: 978-0521865715.

MAZZOCCHI FULVIO, PAOLO PLINI, *Development of environmental application reference thesaurus (EARTH)*, ISKO. 7^o Capítulo español. Congreso, Barcelo, 2005, p.448-461, <<http://bd.ub.edu/isko2005/mazzocchi.pdf>>.

MELUCCI MASSIMO, *Information retrieval. Metodi e modelli per i motori di ricerca*, Franco Angeli Editore, 2013 (Collana: Informatica). ISBN: 8820421127. ISBN-13: 9788820421120.

MILES ALISTAIR, BECHHOFFER SEAN, *SKOS Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant*, 2009. <<http://www.w3.org/TR/skos-reference/skos-xl.html>>.

MORSHED ALAN, et. alii, *Thesaurus alignment for Linked Data publishing, Proceedings International Conference on Dublin Core and Metadata Applications*, The Hague, the Netherlands, 21-23 September, 2011, p. 65-75. DOI: 10.1504/IJMSO.2012.048511; ISSN: 1744-2621.

PEMBERTON STEVEN, et alii, *XHTML™ 1.0: Extensible HyperText Markup Language*, 2000, <<http://www.w3.org/TR/xhtml1/>>.

PRATI GIULIO, *Web 2.0. Internet è cambiato*, Trento, UNI Service, 2007. ISBN 9788861780187.

Proceedings of the ISKO UK Second Biennial Conference, 4th - 5th July, London, 2011, Emerald Group Publishing Limited, 2012. ISBN-10: 1780526148. ISBN-13: 978-1780526140.

RANGANATHAN SHIYALI RAMAMRITA, *Le cinque leggi della biblioteconomia*, a cura di Laura Toti, Firenze, Le Lettere, 2010 (Collana: Pinakes, 6). ISBN 9788860873309. Regolamento UE N. 305/2011 del Parlamento Europeo e del Consiglio, 9 marzo 2011, *Condizioni armonizzate per la commercializzazione dei prodotti da costruzione e che abroga la direttiva 89/106/CEE del Consiglio*, Gazzetta Ufficiale dell'Unione Europea 04/04/2011, n. L 85/5.

RICCARDO ALBERTONI, et. alii, *EARTH: an Environmental Application Reference Thesaurus in the Linked Open Data Cloud*, «Semantic Web», vol.5, n. 2, 2014, Amsterdam, IOS Press, 2014. ISSN 1570-0844 <http://www.semantic-web-journal.net/system/files/swj520_0.pdf>.

RIEDIGER HELMUT, *Cos'è la terminologia e come si fa un glossario*, 2010, <http://www.term-minator.it/corso/doc/mod3_termينو_glossa.pdf>.

ROSS ANN, HETREED JONATHAN, *Architect's Pocket Book*, Architectural Press, 2011. ISBN: 9780080969596.

SABOU MARTA, et alii, *Supporting Tourism Decision making with linked data*, in *8th Int. Conference on Semantic Systems I-SEMANTICS/I-CHALLENGE*, ACM, Graz, Austria, 2012, p. 201-204.

SIGNORE ORESTE, *RDF per la rappresentazione della conoscenza*, <<http://www.w3c.it/papers/RDF.pdf>>.

SINCLAIR JOHN, *Trust the text: language, corpus and discourse*, Londra, Routledge, 2004, p.14. ISBN 0-203-59407-X.

SUCHANEK FABIAN M., et alii, *YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia*, proceedings of the 16th International World Wide Web Conference (WWW 2007) Banff, Canada, 2007. ISBN:978-1-59593-654-7 - DOI:10.1145/1242572.1242667 <<http://suchanek.name/work/publications/www2007.pdf>>.

TUDHOPE DOUGLAS, NIELSEN MARIANNE LYKKE, *Introduction to knowledge organization systems and services*, «New Review of Hypermedia and multimedia», vol. 12, n. 1, June 2006, p.4. DOI: 10.1080/13614560600856433.

UNDERWOOD JASON , ISIKDAG UMIT, *Handbook of Research on Building Information Modeling and Construction Informatics: Concepts and Technologies*, Information Science Publishing, 2009, p. 114. ISBN 978-1-60566-928-1.

UNI 11337:2009 *Edilizia e opere di ingegneria civile - Criteri di codificazione di opere e prodotti da costruzione, attività e risorse - Identificazione, descrizione e interoperabilità*.

VITI ELISABETTA, *Interoperabilità fra thesauri generali e thesauri specialistici in ambito economico finanziario. il caso del Nuovo Soggettario. Sintesi e risultati di un progetto di ricerca*, Seminario nazionale di Biblioteconomia. Didattica e ricerca nell'università italiana e confronti internazionali, <http://w3.uniroma1.it/seminario-biblioteconomia/wp-content/uploads/2013/05/Viti_Elisabetta.pdf>.

WEIMBARGER DAVID, *Elogio del disordine: le nuove regole del mondo digitale*, Milano, BUR Rizzoli, 2010. ISBN 978-88-17-03911-6.

WESTON PAUL GABRIELE, *Il catalogo elettronico. Dalla biblioteca cartacea alla biblioteca digitale*, (Collana: Beni culturali, 26), Roma, Carocci editore, 2002. ISBN 88-430-2174-5.

YUM KI HWAN, DROGEMULLER ROBIN, *How Much Interoperability Can Be Achieved For the Construction Industry Today*, INCITE 2000 Conference, Implementing IT To Obtain a Competitive Advantage in the 21st Century, The Hong Kong Polytechnic University, Hong Kong, January 17.

ZANOLA MARIA TERESA, *La terminologia dei sistemi fotovoltaici tra comunicazione istituzionale e saperi specialistici. Un'introduzione*, in *Costruire un glossario: la terminologia dei sistemi fotovoltaici*, a cura di Maria Teresa Zanola, Milano, Vita&Pensiero, 2012, p. 7-14. ISBN 9788834322970.

SITOGRAFIA

< <http://www.iskoi.org/doc/glossario.htm>>
<<http://agclass.nal.usda.gov>>
<<http://aims.fao.org>>
<<http://biblioteche.unibo.it>>
<<http://ckan.org>>
<<http://csc-dcc.ca>>
<<http://data.gov.uk>>
<<http://dati.senato.it>>
<<http://dbpedia.org>>
<<http://dewey.info>>
<<http://dublincore.org>>
<<http://ec.europa.eu>>
<<http://en.wikipedia.org>>
<<http://eurovoc.europa.eu>>
<<http://flarenet.eu>>
<<http://id.loc.gov>>
<<http://it.dbpedia.org>>
<<http://kcoyle.net>>
<<http://linkeddata.org>>
<<http://linkeddatacatalog.dws.informatik.uni-mannheim.de>>
<<http://lov.okfn.org>>
<<http://nkos.slis.kent.edu>>
<<http://oaei.ontologymatching.org>>
<<http://open-data.europa.eu>>
<<http://protege.stanford.edu>>

<<http://publications.europa.eu>>
<<http://store.uni.com>>
<<http://techwiki.umbel.org>>
<<http://thes.bncf.firenze.sbn.it>>
<<http://thesauronline.iaa.cnr.it>>
<<http://viaf.org>>
<<http://wifo5-03.informatik.uni-mannheim.de>>
<<http://wordnet.princeton.edu>>
<<http://www.bibliotecheoggi.it>>
<<http://www.bncf.firenze.sbn.it>>
<<http://www.bnf.fr>>
<<http://www.buildingsmart.org>>
<<http://www.buildingsmart-tech.org>>
<<http://www.cpic.org.uk>>
<<http://www.csinet.org>>
<<http://www.dati.gov.it>>
<<http://www.eea.europa.eu>>
<<http://www.eionet.europa.eu>>
<<http://www.europeana-libraries.eu>>
<<http://www.fbk.eu>>
<<http://www.geonames.org>>
<<http://www.gesis.org>>
<<http://www.ieee.org>>
<<http://www.ieee.org>>
<<http://www.ifd-library.org>>
<<http://www.ifla.org>>
<<http://www.innovance.it/lessico>>
<<http://www.isni.org>>
<<http://www.iso.org>>

<<http://www.loc.gov>>
<<http://www.meta-net.eu>>
<<http://www.minambiente.it>>
<<http://www.mpi-inf.mpg.de>>
<<http://www.niso.org>>
<<http://www.nist.gov>>
<<http://www.oclc.org>>
<<http://www.omniclass.org/>>
<<http://www.portalu.de>>
<<http://www.rda-jsc.org>>
<<http://www.sbn.it>>
<<http://www.spaziodati.eu>>
<<http://www.steptools.com>>
<<http://www.tecnologica.altervista.org>>
<<http://www.thenbs.com>>
<<http://www.umbel.org>>
<<http://www.uni.com>>
<<http://www.w3.org/>>
<<http://zbw.eu>>
<<https://creativecommons.org>>
<<https://it.wikipedia.org>>
<<https://lib-thesaurus.un.org>>
<<https://okfn.org>>
<<https://portal.dnb.de>>
<<https://www.buildingsmartcanada.ca>>
<<https://www.fsf.org>>
<<https://www.gnu.org/home.it.html>>
<<https://www.princeton.edu>>

APPENDICE

Presentazione di un estratto dello schema di relazioni semantiche interne e di corrispondenze esterne specificate ricorrendo all'ausilio del linguaggio SKOS e riguardanti il termine *Cemento*.

cemento

<skos:prefLabel> cemento </skos:prefLabel>

<skos:definition>

Legante idraulico, cioè un materiale inorganico finemente macinato che, quando mescolato con acqua, forma una pasta che fa presa e indurisce a seguito di reazioni e processi di idratazione e che, una volta indurita, mantiene la sua resistenza e la sua stabilità anche sott'acqua.

</skos:definition>

<skos:narrower> cemento comune </skos:narrower>

<skos:narrower> cemento da muratura </skos:narrower>

<skos:narrower> cemento speciale </skos:narrower>

<skos:exactMatch> <http://purl.org/bnct/tid/3571> </skos:exactMatch> "Cementi - Nuovo Soggettario"

<skos:exactMatch> <http://it.dbpedia.org/page/Cemento> </skos:exactMatch> "Cemento - DBpediaItalia"

cemento comune

<skos:prefLabel> cemento comune </skos:prefLabel>

<skos:altLabel> CEM </skos:altLabel>

<skos:broader> cemento </skos:broader>

<skos:narrower> CEM I </skos:narrower>

<skos:narrower> CEM II </skos:narrower>

<skos:narrower> CEM III </skos:narrower>

<skos:narrower> CEM IV </skos:narrower>

<skos:narrower> CEM V </skos:narrower>

<skos:related> durabilità </skos:related>

<skos:related> resistenza alla compressione </skos:related>

< skos:related> stabilità </skos:related>

< skos:related> tempo di presa </skos:related>

<skos:related> tenore di cloruro </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento da muratura

<skos:prefLabel> cemento da muratura </skos:prefLabel>

<skos:definition>

Factory made finely powdered hydraulic binder which relies essentially upon the presence of Portland cement clinker to develop strength. When mixed with sand and water only and without the addition of further materials it produces a workable mortar suitable for use in rendering, plastering and masonry work.

</skos:definition>

<skos:broader> cemento </skos:broader>

<skos:related> composizione </skos:related>

<skos:related> contenuto d'aria della malta fresca </skos:related>

<skos:related> contenuto di cloruro </skos:related>

<skos:related> durabilità </skos:related>

<skos:related> finezza </skos:related>

<skos:related> resistenza alla compressione </skos:related>

<skos:related> ritenzione idrica della malta fresca </skos:related>

<skos:related> solidità </skos:related>

<skos:related> tempo di presa </skos:related>

< skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

< kos:relatedMatch> <http://purl.org/bncf/tid/23872> </skos:relatedMatch> "Murature - Nuovo Soggettario"

<skos:retatedMatch> <http://it.dbpedia.org/page/Muratura> </skos:retatedMatch> "Muratura - DBpediaItalia"

cemento speciale

<skos:prefLabel> cemento speciale </skos:prefLabel>

<skos:broader> cemento </skos:broader>

<skos:narrower> cemento sovrasolfato </skos:narrower>

<skos:narrower> cemento speciale a calore di idratazione molto basso
</skos:narrower>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM I

<skos:prefLabel> CEM I </skos:prefLabel>

<skos:altLabel> cemento Portland </skos:altLabel>

<skos:broader> cemento comune </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> perdita al fuoco </skos:related>

<skos:related> residuo insolubile </skos:related>

<skos:closeMatch> "http://purl.org/bncf/tid/26991" </skos:closeMatch> "Cementi portland - Nuovo Soggettario"

<skos:closeMatch> "http://it.dbpedia.org/page/Cemento_Portland" </skos:closeMatch> "Cemento_Portland - DBpediaItalia"

CEM II

<skos:prefLabel> CEM II </skos:prefLabel>

<skos:broader> cemento comune </skos:broader>

<skos:narrower> CEM II/A-D </skos:narrower>

<skos:narrower> cemento portland al calcare </skos:narrower>

<skos:narrower> cemento portland alla loppa </skos:narrower>

<skos:narrower> cemento portland alla pozzolana </skos:narrower>

<skos:narrower> cemento portland alle ceneri volanti </skos:narrower>

<skos:narrower> cemento portland allo scisto calcinato </skos:narrower>

<skos:narrower> cemento portland composito </skos:narrower>

<skos:broadMatch> http://purl.org/bncf/tid/3571 </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> http://it.dbpedia.org/page/Cemento </skos:broadMatch> "Cemento - DBpediaItalia"

CEM III

<skos:prefLabel> CEM III </skos:prefLabel>

<skos:broader> cemento comune </skos:broader>

<skos:narrower> cemento d'altoforno </skos:narrower>

<skos:related> perdita al fuoco </skos:related>

<skos:related> residuo solubile </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento DBpediaItalia"

CEM IV

<skos:prefLabel> CEM IV </skos:prefLabel>

<skos:broader> cemento comune </skos:broader>

<skos:narrower> cemento pozzolanico </skos:narrower>

<skos:related> pozzolanicità </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

<skos:relatedMatch> <http://purl.org/bncf/tid/27001> </skos:relatedMatch> "Cementi pozzolanici - Nuovo Soggettario"

CEM V

<skos:prefLabel> CEM V </skos:prefLabel>

<skos:broader> cemento comune </skos:broader>

<skos:narrower> cemento composito </skos:narrower>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento
- DBpediaItalia"

Cemento sovrasolfato

<skos:prefLabel> cemento sovrasolfato </skos:prefLabel>

<skos:altLabel> SSC </skos:altLabel>

<skos:definition>

E' un legante idraulico, ossia un materiale inorganico finemente macinato che, miscelato con acqua forma una pasta che fa presa e indurisce per mezzo di processi e reazioni di idratazione e che, dopo l'indurimento mantiene la sua resistenza e stabilità persino sott'acqua.

</skos:definition>

<skos:broader> cemento speciale </skos:broader>

<skos:related> additivo </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> loppa granulata d'altoforno </skos:related>

<skos:related> solfato di calcio </skos:related>

<skos:related> calore di idratazione </skos:related>

<skos:related> contenuto di sostanze pericolose </skos:related>

<skos:related> perdita al fuoco </skos:related>

<skos:related> residuo insolubile </skos:related>

<skos:related> resistenza a compressione </skos:related>

<skos:related> rilascio di sostanze pericolose </skos:related>

<skos:related> stabilità </skos:related>

<skos:related> tempo di presa </skos:related>

<skos:related> tenore di cloruro </skos:related>

<skos:related> tenore di so3 </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo
Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento DBpediaItalia"

cemento speciale a calore di idratazione molto basso

< skos:prefLabel> cemento speciale a calore di idratazione molto basso
</skos:prefLabel>

<skos:altLabel> SSC </skos:altLabel>

<skos:definition>

È un legante idraulico, ossia un materiale inorganico comunemente macinato che, quando mescolato con acqua, forma una pasta che rapprende e indurisce a seguito di reazioni e processi di idratazione e che, una volta indurita, mantiene la sua resistenza e la sua stabilità anche sott'acqua.

</skos:definition>

<skos:broader> cemento speciale </skos:broader>

<skos:narrower> cemento composito speciale a calore di idratazione molto basso
</skos:narrower>

<skos:narrower> cemento d'altoforno speciale a calore di idratazione molto basso
</skos:narrower>

<skos:narrower> cemento pozzolanico speciale a calore di idratazione molto basso
</skos:narrower>

<skos:related> calore di idratazione </skos:related>

<skos:related> durabilità </skos:related>

<skos:related> resistenza alla compressione </skos:related>

<skos:related> stabilità </skos:related>

<skos:related> tempo di presa </skos:related>

<skos:related> tenore di cloruro </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-D

< skos:prefLabel> CEM II/A-D </skos:prefLabel>

<skos:altLabel> cemento Portland ai fumi di silice</skos:altLabel>

<skos:broader> CEM II </skos:broader>

<skos:related> clinker di cemento portlan </skos:related>

<skos:related> fumi di silice </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento portland al calcare

< skos:prefLabel> cemento portland al calcare </skos:prefLabel>

<skos:broader> CEM II </skos:broader>

<skos:narrower> CEM II/A-L </skos:narrower>

<skos:narrower> CEM II/A-LL </skos:narrower>

<skos:narrower> CEM II/B-L </skos:narrower>

<skos:narrower> CEM II/B-LL </skos:narrower>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento portland alla loppa

<skos:prefLabel> cemento portland al calcare </skos:prefLabel>

<skos:broader> CEM II </skos:broader>

<skos:narrower> CEM II/A-S </skos:narrower>

<skos:narrower> CEM II/B-S </skos:narrower>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento portland alla pozzolana

<skos:prefLabel> cemento portland alla pozzolana</skos:prefLabel>

<skos:broader> CEM II </skos:broader>

<skos:narrower> CEM II/A-P </skos:narrower>

<skos:narrower> CEM II/A-Q </skos:narrower>

<skos:narrower> CEM II/B-P </skos:narrower>

<skos:narrower> CEM II/B-Q </skos:narrower>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento portland alle ceneri volanti

<skos:prefLabel> cemento portland alle ceneri volanti</skos:prefLabel>

<skos:broader> CEM II </skos:broader>

<skos:narrower> CEM II/A-V </skos:narrower>

<skos:narrower> CEM II/A-W </skos:narrower>

<skos:narrower> CEM II/B-V </skos:narrower>

<skos:narrower> CEM II/B-W </skos:narrower>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento portland allo scisto calcinato

<skos:prefLabel> cemento portland allo scisto calcinato </skos:prefLabel>

<skos:broader> CEM II </skos:broader>

<skos:narrower> CEM II/A-T </skos:narrower>

<skos:narrower> CEM II/B-T </skos:narrower>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento portland composito

<skos:prefLabel> cemento portland composito </skos:prefLabel>

<skos:broader> CEM II </skos:broader>

<skos:narrower> CEM II/A-M </skos:narrower>

<skos:narrower> CEM II/B-M </skos:narrower>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento d'altoforno

<skos:prefLabel> cemento d'altoforno </skos:prefLabel>

<skos:broader> CEM III </skos:broader>

<skos:narrower> CEM III/A </skos:narrower>

<skos:narrower> CEM III/B </skos:narrower>

<skos:narrower> CEM III/C </skos:narrower>

<skos:related> cemento d'altoforno resistente ai solfati </skos:related>

<skos:closeMatch> <http://purl.org/bncf/tid/27010> </closeMatch> "Cementi d'altoforno – Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

cemento pozzolanico

<skos:prefLabel> cemento pozzolanico </skos:prefLabel>

<skos:broader> CEM IV </skos:broader>

<skos:narrower> CEM IV/A </skos:narrower>

<skos:narrower> CEM IV/B </skos:narrower>

<skos:closeMatch> <http://purl.org/bncf/tid/27001> </closeMatch> "Cementi pozzolanici - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-L

<skos:prefLabel> CEM II/A-L </skos:prefLabel>

<skos:broader> cemento portland al calcare </skos:broader>

<skos:related> calcare </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-LL

<skos:prefLabel> CEM II/A-LL </skos:prefLabel>

<skos:broader> cemento Portland al calcare </skos:broader>

<skos:related> calcare </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/B-L

<skos:prefLabel> CEM II/B-L </skos:prefLabel>

<skos:broader> cemento portland al calcare </skos:broader>

<skos:related> calcare </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/B-LL

<skos:prefLabel> CEM II/B-LL </skos:prefLabel>

<skos:broader> cemento Portland al calcare </skos:broader>

<skos:related> calcare </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi - Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-S

<skos:prefLabel> CEM II/A-S </skos:prefLabel>

<skos:broader> cemento portland alla loppa </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> loppa granulata d'altoforno </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/B-S

<skos:prefLabel> CEM II/B-S </skos:prefLabel>

<skos:broader> cemento portland alla loppa </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> loppa granulata d'altoforno </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-P

<skos:prefLabel> CEM II/A-P </skos:prefLabel>

<skos:broader> cemento portland alla pozzolana </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> pozzolana naturale </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento
- DBpediaItalia"

CEM II/A-Q

<skos:prefLabel> CEM II/A-Q </skos:prefLabel>

<skos:broader> cemento portland alla pozzolana </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> pozzolana naturale calcinata </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi Nuovo -
Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento
- DBpediaItalia"

CEM II/B-P

<skos:prefLabel> CEM II/B-P </skos:prefLabel>

<skos:broader> cemento portland alla pozzolana </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> pozzolana naturale </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/3571> </skos:broadMatch> "Cementi Nuovo -
Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento
- DBpediaItalia"

CEM II/B-Q

<skos:prefLabel> CEM II/B-Q </skos:prefLabel>

<skos:broader> cemento portland alla pozzolana </skos:broader>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> pozzolana naturale calcinata </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-V

<skos:prefLabel> CEM II/A-V </skos:prefLabel>

<skos:broader> cemento portland alle ceneri volanti </skos:broader>

<skos:related> cenere volante silicea </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/A-W

<skos:prefLabel> CEM II/A-W </skos:prefLabel>

<skos:broader> cemento portland alle ceneri volanti </skos:broader>

<skos:related> cenere volante calcarea </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/B-V

<skos:prefLabel> CEM II/B-V </skos:prefLabel>

<skos:broader> cemento portland alle ceneri volanti </skos:broader>

<skos:related> cenere volante calcarea </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM II/B-W

<skos:prefLabel> CEM II/B-W</skos:prefLabel>

<skos:broader> cemento portland alle ceneri volanti </skos:broader>

<skos:related> cenere volante calcarea </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/3571> </skos:broadMatch> "Cementi Nuovo - Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM III/A

<skos:prefLabel> CEM III/A</skos:prefLabel>

<skos:broader> cemento d'altoforno </skos:broader>

<skos:related> loppa granulata d'altoforno </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/27010> </broadMatch> "Cementi d'altoforno – Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM III/B

<skos:prefLabel> CEM III/B</skos:prefLabel>

<skos:broader> cemento d'altoforno </skos:broader>

<skos:related> loppa granulata d'altoforno </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/27010> </broadMatch> "Cementi d'altoforno – Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM III/C

<skos:prefLabel> CEM III/C</skos:prefLabel>

<skos:broader> cemento d'altoforno </skos:broader>

<skos:related> loppa granulata d'altoforno </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:broadMatch> <http://purl.org/bnct/tid/27010> </broadMatch> "Cementi d'altoforno – Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento - DBpediaItalia"

CEM IV/A

<skos:prefLabel> CEM IV/A</skos:prefLabel>

<skos:broader> cemento pozzolanico </skos:broader>

<skos:related> cenere volante calcarea </skos:related>

<skos:related> cenere volante silicea </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> fumi di silice </skos:related>

<skos:related> pozzolana naturale </skos:related>

skos:related> pozzolana naturale calcinata </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/27001> </broadMatch> "Cementi pozzolanici
- Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento
- DBpediaItalia"

CEM IV/B

<skos:prefLabel> CEM IV/B</skos:prefLabel>

<skos:broader> cemento pozzolanico </skos:broader>

<skos:related> cenere volante calcarea </skos:related>

<skos:related> cenere volante silicea </skos:related>

<skos:related> clinker di cemento portland </skos:related>

<skos:related> fumi di silice </skos:related>

<skos:related> pozzolana naturale </skos:related>

skos:related> pozzolana naturale calcinata </skos:related>

<skos:broadMatch> <http://purl.org/bncf/tid/27001> </broadMatch> "Cementi pozzolanici
- Nuovo Soggettario"

<skos:broadMatch> <http://it.dbpedia.org/page/Cemento> </skos:broadMatch> "Cemento
- DBpediaItalia"