



Consiglio Nazionale delle Ricerche
Istituto di Fisiologia Clinica
Area della Ricerca S.Cataldo Pisa
Sez. Epidemiologia e Ricerca sui Servizi Sanitari

Technical Report / Jul. 2013

Ontologie e modellazione di dati sanitari
Attività di ricerca nell'ambito del progetto ODINET

Versione 1.0

Autori

S. Pieroni, M. Franchini, F. Mariani, L. Fortunato, S. Molinaro

Sommario

Introduzione	3
Ontologie e web semantico	4
Costruire un'ontologia	6
Fondamenti teorici	9
La scelta del linguaggio	12
La scelta dello strumento di editing e gestione automatica della ontologia	15
Principali ontologie in ambito sanitario	17
DATI OPEN	21
BIBLIOGRAFIA DI RIFERIMENTO	25

Introduzione

L'acquisizione, la gestione, la rappresentazione e fruibilità della conoscenza tramite l'utilizzo di strumenti informatici risulta una problematica ampiamente discussa e per la quale, nel corso degli anni, sono stati proposti diversi approcci e soluzioni.

La grande diffusione del web ha cambiato il modo tradizionale di intendere, reperire e gestire le informazioni, rendendo sempre più importanti gli aspetti legati all'interoperabilità mediante lo scambio di metadati.

Le tecnologie legate al mondo del Web Semantico rappresentano una possibile soluzione adatta al perseguimento di tale obiettivo, individuando in particolare nel modello basato sui principi del Linked Data, ovvero quell'insieme di buone pratiche atte a pubblicare e collegare dati sul web *a uso di una macchina*, le caratteristiche più valide per l'ottenimento di una modellazione dei dati completa e rispondente alle esigenze richieste dal dominio di applicazione.

In questo documento ci occuperemo in particolare del dominio della Sanità, dominio in cui la ricerca risente di un annoso problema di comunicazione e rappresentazione della conoscenza dovuto soprattutto al fatto che comunità distinte di ricercatori si servono di tecnologie diverse e spesso incompatibili fra loro. Diventa pertanto essenziale l'utilizzo di un'infrastruttura tecnologica che permetta di identificare i concetti, le equivalenze ed i sinonimi in modo univoco e il più possibile automatico, basandosi su una rappresentazione della conoscenza strettamente formale che permetta di mettere in relazione i concetti fra loro, eseguire ragionamenti e fare inferenza.

Il resto del documento è dedicato alle ontologie e alla loro contestualizzazione nello scenario dei linguaggi di comunicazione per il web. Gli argomenti verranno trattati in forma introduttiva secondo la scaletta visibile in Figura 1



Figura 1: Schema dei contenuti

Ontologie e web semantico

Le ontologie, i metadati e i Linked Open Data combinati insieme, rappresentano un punto di partenza imprescindibile per metodologie efficaci di ricerca sul web semantico.

Il concetto di linked data, è fortemente connesso al web semantico, anche se il web semantico non si risolve nel solo tecnicismo dei linked data, ma necessita del rispetto di alcune regole per la creazione di uno strato di contenuti accessibili a processi automatizzati. Produrre linked data significa, esprimere i significati delle informazioni, renderle condivisibili fra più applicazioni e utilizzabili da applicazioni diverse da quelle per cui erano state inizialmente create.

Mentre il web tradizionale è costituito da documenti e oggetti HTML connessi tramite hyperlink, il web semantico è costituito da oggetti reali descritti tramite dati: la caratteristica principale del web semantico è il *metadato*, ovvero quel dato che ha il compito specifico di *descrivere il dato*



I metadati si riferiscono al dominio che vogliono descrivere ovvero sono “informazioni relative al dato”.

L'informazione presente sul web può apparire sotto forma di 1) dati strutturati, quali ad esempio i database relazionali, 2) semi-strutturati come pagine scritte in XML, 3) non strutturati come i file testuali.

The screenshot shows a web interface for "Clinical Knowledge On Action" (Clio). The user is logged in as "sabrina molinaro". The page displays search results for "SMITH HENRY, Male, 77". There are tabs for "Vital Parameters", "Diagnosis", "Drug Treatment", and "Neuroendocrine Profile". The "Diagnosis" tab is active, showing a search for "Diagnosis 02/09/2012". A table lists two diagnoses:

Dates	Diagnosis code	Description	Date	Institute/Update	Copy	Delete
01/10/2012 Show exams	25330	Diabete con altri tipi di coma, tipo II o non specificato, non definito se scompensato	02/09/2012	FC Pisa		
02/09/2012 Show exams	4257	Cardiomiopatie nutrive e metaboliche	02/09/2012	FC Pisa		

Dati Strutturati

The screenshot shows a website with various XML-related links. The links are organized into several categories:

- Java / XML:** Validation Java and XML, Add XML to J2EE Applications, XPath-JSP Tutorial, Carthago ICT, Java/XML FAQ, Javatechnology and XML.
- PHP / XML:** TWNet, XML CMS, WPS : php-XML_publishing.
- Forms + XML=XFDL:** eforms.nl.
- SXML/XML:** CWI SXML_page, SXML W3C Specifications, SXML players.
- XML 101.com:** XML.org (tip), XML.com, XML introduction, Oasis.
- XML voor beginners:** Webmeas (nl) (tip), XML for the absolute beginner, XML cursus Genias, XML in 10 points.
- XML voor gevorderden:** XMLPitstop.com, XML School, XML 101 tutorial.
- XML FAQ:** The ?XML FAQ, What's XML?
- Prikbord XML (tip):** Vacatures/Personeel, Whitehorses, Piuflex, Oracle Vacatures, njobs, IT Detaching.
- Software XML editors:** xmlBlueprint XML Editor, Xemacs, Xmetal.com, XMLwriter, Extensibility, XMLspy.
- XML Newsfeed:** XML News Feeds for Your Site.

There is also a section titled "Extensible Markup Language (XML) is een standaard voor het definiëren van formele markuptalen voor de representatie van gestructureerde gegevens in de vorm van platte tekst. Deze representatie is zowel machineleesbaar als leesbaar voor de mens."

Dati Semi-Strutturati

The screenshot shows a webpage titled "Epidemiologia della salute". The text describes epidemiological research conducted within the Section, monitoring specific phenomena related to the health status of the Italian population. It mentions collaborations with non-European countries and the application of studies in public health. The text also lists specific activities:

- pianificazione e conduzione di vari tipi di studi epidemiologici (descrittivi e analitici, indagini di popolazione, indagini sul campo, indagini campionate, modelli matematici, ...);
- descrizione delle misure di prevalenza, incidenza e mortalità e implementazione di sistemi di sorveglianza epidemiologica di popolazione (bambini, adulti e anziani con particolare riferimento agli stili di vita rilevanti per le malattie cronico-degenerative);
- utilizzazione di metodi statistici, informatici ed informativi complessi per l'approfondimento delle conoscenze sullo stato di salute della popolazione;
- sviluppo di studi orientati alla medicina personalizzata;
- creazione di banche dati;
- promozione della formazione degli operatori del Servizio Sanitario Nazionale e di strutture internazionali (es. in Senegal) in epidemiologia e biostatistica, uso dei package di analisi statistica dei dati;
- divulgazione delle conoscenze attraverso la pubblicazione di articoli, rapporti e monografie a livello nazionale ed internazionale.

Dati Non Strutturati

Figura 2 Tipologie di dati sul Web

Qualunque informazione rappresentata in modo da poter essere trattata da un computer costituisce in pratica un dato. Diverso è il concetto di metadato, caratteristica principale del web semantico che ha che il compito di descrivere il dato: i metadati e le annotazioni si riferiscono al dominio specifico che vogliono descrivere e sono associati ai dati. Per database strutturati, il più comune esempio di metadato è lo schema del database. I metadati possono essere usati per memorizzare proprietà relative al formato, essi possono descrivere il contenuto informativo oppure rappresentarne un sommario quando i dati sono descritti in maniera molto analitica. Sono fondamentali per rappresentare proprietà o relazioni tra oggetti appartenenti a tipologie eterogenee.

Uno dei metodi ad oggi più efficiente per rappresentare formalmente un dominio è la rappresentazione mediante ontologie ed il passaggio dai dati alle ontologie avviene proprio attraverso i metadati:

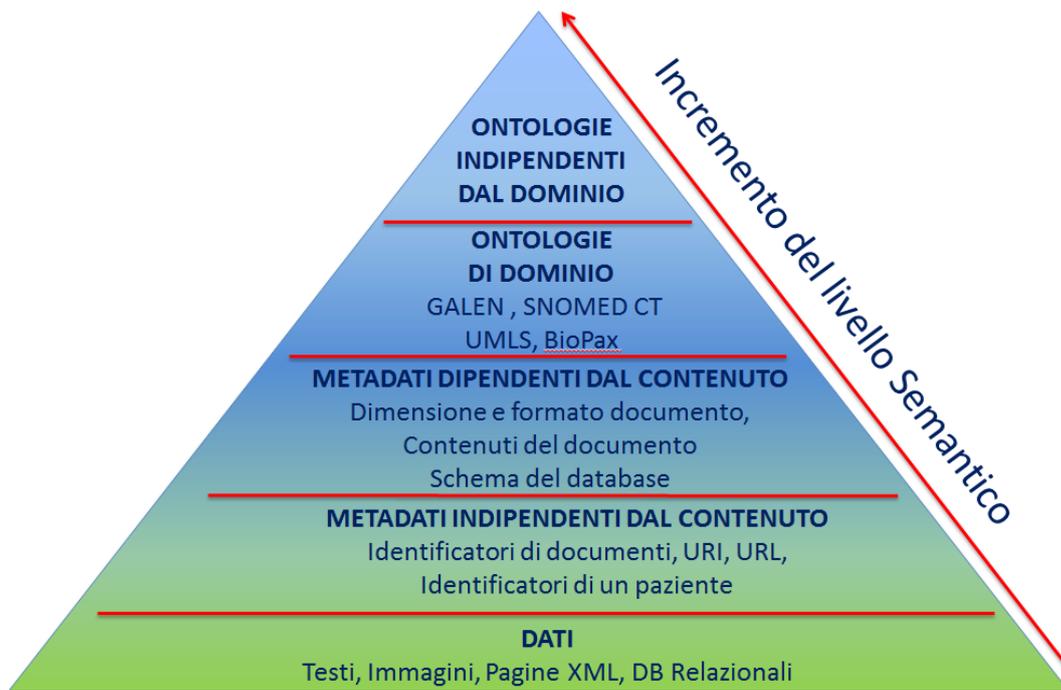


Figura 3: dai dati alle ontologie attraverso i metadati

L'ontologia fornisce la concettualizzazione che descrive la semantica del dato con un linguaggio sintatticamente e semanticamente ricco.

Ad esempio una applicazione di cartella clinica può essere corredata dei metadati di dominio relativi a : esami eseguiti durante il ricovero, farmaci somministrati, medicazioni effettuate, allergie, interventi chirurgici

Surname Name, Sex, Age

Vital Parameters Biomoral Exams ECG Holter

1/5/2011 2/5/2011 **5/5/2011** History New Date

Date	Time	Exam	Meas Unit	Value
05/05/11	07:00	Body Temp	C°	38.0*
05/05/11	07:00	Pulse Rate	Beat/M	70

New Exam



```
<owl:Class rdf:about="&Ontology12;Allergie">
<owl:Class rdf:about="&Ontology12;Esami"/>
<owl:Class rdf:about="&Ontology12;Farmaci"/>
<owl:Class rdf:about="&Ontology12;InterventiChirurgici"/>
<owl:Class rdf:about="&Ontology12;Medicazioni"/>
```

Figura 4: esempio di metadati ed ontologia di dominio

Definizione formale di ontologia

Ontologia è un termine che deriva dalla filosofia: esso appare per la prima volta negli scritti di Parmenide (circa 504 a.C.) e deriva quindi dal greco *eon logos*, in Italiano “discorso sull’ente”.

L’ontologia si occupa, infatti, dello studio dell’essere, ovvero di ciò che è e delle sue categorie fondamentali.

In informatica, un’ontologia, usata in particolar modo negli studi sull’intelligenza artificiale e nella classificazione dei dati, è il tentativo di formulare una classificazione di concetti, quindi una gerarchizzazione, nell’ambito di un dominio.

In letteratura si trovano diverse definizioni:

Neches ('91) “L'insieme dei termini basilari e delle relazioni, che costituiscono il vocabolario di un'area specifica, e delle regole per combinare termini e relazioni per determinare estensioni del vocabolario”

Gruber ('93) “L'ontologia è una specificazione esplicita di una concettualizzazione”

Borst, Gruber ('97) “L'ontologia è una specificazione formale ed esplicita di una concettualizzazione condivisa”

Guarino ('97) “Una teoria logica che spiega il significato inteso di un vocabolario formale”

Swartout ('97) “Un'ontologia è un insieme di termini descrittivi un dominio strutturato in maniera gerarchica che può essere usato come fondamento di una base di conoscenza”

Un’ontologia è dunque un tentativo di formulare uno schema concettuale esaustivo e rigoroso nell’ambito di un dato dominio. Essa permette la formalizzazione di un **modello** che include concetti e relazioni aggregate in gruppi di attori semplici e/o attori esperti. Lo schema concettuale può assumere forme diverse, dalle semplici tassonomie fino ad arrivare ai modelli di sistemi complessi con relativi assiomi logici, passando attraverso forme di classificazione intermedia come thesauri.

L’interazione tra persone e sistemi software porta alla ricerca di un sistema comune e condiviso per la comunicazione e comprensione dell’informazione, pertanto l’idea è quella di usare un vocabolario condiviso per descrivere il contenuto delle risorse, la cui semantica è descritta in un formato ragionevolmente non ambiguo e processabile da una macchina, l’ontologia.

Ogni informazione del dominio di pertinenza sarà mappata, attraverso la definizione dei metadati, dalla propria ontologia ed inserita in un contesto che la relazioni ad altre ontologie, in modo da creare delle relazioni logiche che permettano, per esempio, di distinguere il significato della parola “campione” in ambito clinico rispetto a “campione” in un contesto di sociale o statistico-economico. Ogni informazione dovrebbe quindi avere un significato compiuto in un certo dominio secondo il meccanismo di associazione delle informazioni, proprio della mente umana.

Costruire un’ontologia

Lo sviluppo di un’ontologia è un processo dinamico che parte con una ontologia iniziale approssimata, rivista successivamente, raffinata e quindi definita nei dettagli. [N. F. Noy and D. McGuinness. *Ontology development: A guide to creating your first ontology*].

Non esiste dunque un’unica metodologia di creazione corretta, la soluzione migliore dipende sempre dall’uso che si deve fare dell’ontologia.

Il processo prevede sempre più passi:

- Definire con precisione lo scopo dell’ontologia.
- Effettuare una ricerca bibliografica
- Riutilizzare quanto più possibile ontologie esistenti
- Considerare le fonti di dati (non solo OPEN) pertinenti
- Coinvolgere gli esperti di dominio prevedendo un team multidisciplinare



Figura 5: il processo di costruzione della ontologia

Come visualizzato in Figura 5, il processo è tipicamente ciclico e solitamente l'utilizzo pratico dell'ontologia fa emergere la necessità di modifiche che comportano la rivisitazione dell'intero processo. Nel dettaglio vediamo alcune attività connesse agli step fondamentali di costruzione dell'ontologia.

a. Esame del dominio ed acquisizione della conoscenza.

In questa prima fase, si cerca di raccogliere quante più informazioni possibili sul dominio di interesse e di comprendere i termini usati formalmente per descriverne le entità in maniera consistente.

Va considerato che si possono distinguere diverse tipologie di ontologie:

- ontologie top-level che descrivono concetti molto generali, quali spazio, tempo, materia, eventi che sono indipendenti da un particolare problema o dominio.
- Ontologie di dominio che descrivono un dominio (ad es. la medicina) o un generico task o attività (ad esempio l'imaging cardiovascolare) specializzando i concetti/relazioni introdotti nell'ontologia top-level
- Ontologie applicative che descrivono concetti dipendenti sia da un dominio particolare che da un task e che sono specializzazioni di entrambe le ontologie correlate fra loro (Ad es la *Risonanza Magnetica Cardiaca*)

In sintesi questa prima fase ha l'obiettivo di dare risposta a domande del tipo:

- *quale del dominio di interesse coprirà l'ontologia?* Ad esempio nell'ambito sanitario il dominio clinico e quello dei servizi sanitari sono concettualmente molto diversi.
- *Qual è lo scopo dell'ontologia?*
- *A quali tipi di domande l'informazione espressa dall'ontologia può fornire risposte?*
- E non ultimo *chi userà l'ontologia?*

b. Considerare il riuso di risorse esistenti.

Una volta effettuata la scelta del dominio, è fondamentale controllare l'esistenza di risorse già presenti da poter riutilizzare, questo perché, l'idea di rifinire ed estendere risorse esistenti, quali glossari, dizionari dei termini e dei sinonimi, tassonomie e altre ontologie è un vantaggio in termini di sviluppo dell'ontologia stessa e di tempo nella sua realizzazione.

c. Pianificare lo sviluppo dell'ontologia.

In questa fase si inizia a progettare la struttura concettuale complessiva del dominio, identificando i principali concetti del dominio e le loro proprietà, e stabilendo le relazioni tra i vari concetti. Si sviluppa quindi un *flat glossary* per documentare ciascun termine con una definizione in linguaggio naturale:

i nomi diventano oggetti (o attori) e i verbi si trasformano in relazioni o processi;

di seguito si elabora uno *structured glossary* ossia si specializzano i termini e si individuano gli attributi. Infine si identificano tutte le relazioni concettuali fra gli oggetti.

d. Organizzare ed integrare l'ontologia.

Per individuare nuovi concetti è possibile adottare tre tipi di approcci:

- a) **top-down**: prevede l'identificazione dei concetti generali e attraverso un affinamento successivo si procede verso i concetti particolari (es. da patologia a malattia cronica);
- b) **bottom-up**: si procede per livelli di astrazione, partendo dalle entità particolari del dominio per astrarre i concetti generali che racchiudono o fanno uso di quelli particolari (da cardiopatia a malattia cronica);
- c) **middle-out** (o combinato) che prevede di individuare prima i concetti salienti e poi generalizzare e specializzare.

L'approccio combinato risulta spesso quello preferito perché i concetti salienti sono tendenzialmente quelli che descrivono meglio il dominio.

I concetti da soli non forniscono informazioni sufficienti, pertanto è fondamentale definire anche le relazioni tra gli oggetti del dominio.

e. Verificare l'ontologia.

Una volta sviluppata l'ontologia, occorre analizzarla per individuare eventuali inconsistenze sintattiche, logiche e semantiche tra i suoi elementi. Vanno previste diverse tipologie di verifica:

- verifica di completezza rispetto al dominio da modellare
- verifica di concisione per evitare di fare assunzioni implicite e ridondanti
- verifica di consistenza per evitare definizioni in contraddizione
- verifica di coerenza affinché tutte le relazioni siano consistenti con le definizioni dei concetti

Infine si può verificare anche la riusabilità e scalabilità che implica poter estendere l'ontologia senza che l'estensione interessi i concetti e le relazioni già presenti.

f. Rilasciare l'ontologia.

Al termine dello sviluppo dell'ontologia, è fondamentale corredarla di annotazioni attraverso la scrittura di 'etichette' da abbinare ai concetti:

- Etichetta di default, ovvero l'etichetta preferita per il concetto
- Una o più etichette alternative utilizzate per i sinonimi e/o varianti con loro scopo di consentire l'individuazione di tutti i dati semanticamente relazionati

Se l'ontologia deve essere multilingua, ogni etichetta dovrà avere specificata una **lingua** (es. @it).

Occorre inoltre corredare l'ontologia di una documentazione di supporto all'utenza.

Sarà infine necessaria una verifica da parte degli esperti del dominio, possibilmente attraverso una applicazione software che permetta una vera e propria validazione della ontologia.

Fondamenti teorici

L'ontologia si struttura principalmente in:

- una lista finita di termini che denotano i **concetti** (o **classi di oggetti**) del dominio
- le **relazioni** ovvero le gerarchie di concetti, loro **proprietà**, restrizioni di valori e relazioni logiche e di disgiunzione

1. Concetti

Un concetto può essere una qualsiasi entità astratta o concreta. Può rappresentare un oggetto, una nozione, o un'idea, la descrizione di un processo, una funzione, una strategia o un flusso di ragionamento.

Ad esempio nel settore medico un concetto può essere:

- l' «apparato cardiovascolare»
- il processo di vascolarizzazione
- il flusso informativo ospedaliero

Più formalmente un concetto è sempre caratterizzato da:

- un **termine** (o un simbolo)
- una **ESTENSIONE**, ovvero gli **oggetti** a cui il concetto può essere applicato
- un' **intensione** un insieme di **proprietà** e **attributi** usati per specificarne la semantica (l'insieme di caratteristiche che gli oggetti che compongono il concetto hanno in comune)

Ad esempio il concetto formalizzato con il termine di «patologia» ha come possibile estensione la tipologia: «cardiovasculopatia», «neoplasia», «gastroenteropatia», «diabete», «broncopneumopatia», ecc. Lo stesso concetto ha come possibile intensione le caratteristiche che ne specificano la semantica: «sintomi», «diagnosi», «trattamenti farmacologici», «percorsi terapeutici» ecc.

2. Relazioni

Le relazioni rappresentano **tipologie di interazione** fra i concetti del dominio e sono formalmente rappresentate come un qualsiasi sottoinsieme del prodotto di N insiemi.

Ogni relazione è caratterizzata da:

- un **termine**
- una **ESTENSIONE**, ovvero l'insieme delle possibili tuple degli oggetti manipolati con questa relazione. Per esempio per la relazione "affetto_da" l'estensione include tutte le patologie di cui si può essere affetti
- una **INTENSIONE**, ovvero il **TIPO** di oggetti manipolati. Per esempio nella relazione "affetto_da" l'intensione specifica pazienti e patologie

La Figura 6 mostra un esempio dove 7 oggetti (o individui) *Paz1*, *Paz2*, *Paz5*, *Cardiovasculopatia*, *Diabete*, *Neoplasia* e *Neuropatia* sono raggruppati in 2 concetti (o classi), *Pazienti* e *Patologie*, e relazionati attraverso la proprietà *affetto_da*

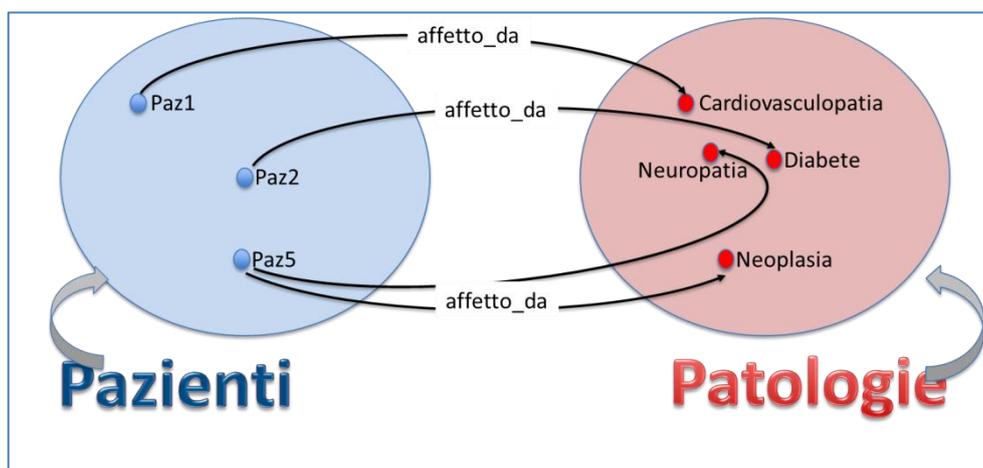


Figura 6: Concetti e Relazioni

3. Gerarchie

I concetti (ovvero le classi) possono essere organizzate in gerarchie di superclassi e sottoclassi. Dalla classe Patologie è possibile derivare le sotto-classi *Croniche* e *NonCroniche* per separare ad esempio le malattie cardiovascolari o neurologiche tipicamente croniche, dalle patologie correlate ai traumatismi o alla gravidanza tipicamente non croniche.

Dire che Croniche è sotto-classe di Patologie significa affermare che tutte le Croniche sono Patologie.

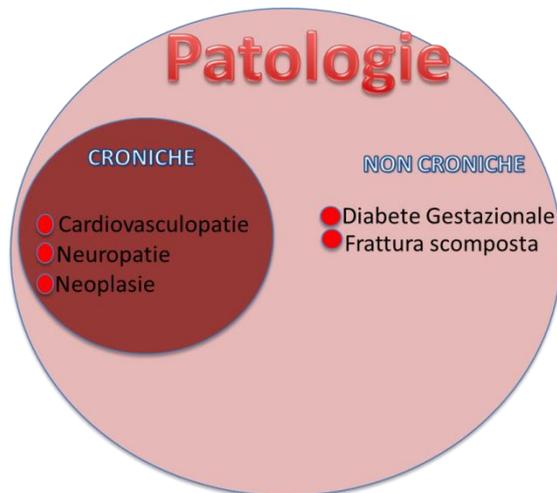


Figura 7: Gerarchie

4. Altri elementi costitutivi: funzioni, assiomi, istanze

Le **funzioni** sono un tipo speciale di relazioni in cui il valore dell'ultimo elemento è unico per la lista degli argomenti. Formalmente sono definite attraverso la seguente notazione:

$F: C1x C2x \dots x Cn-1 \rightarrow Cn$

Un'esempio di funzione ternaria (a tre argomenti) è quello che calcola il prezzo di un prodotto di seconda mano in funzione del modello, dell'anno di fabbricazione e del numero di anni che è stato utilizzato.

Un esempio di funzione a più argomenti nel dominio medico è quella che stabilisce che un paziente è scompensato in base ai criteri di inclusione (diagnosi ricevute durante il ricovero, esenzione specifica, farmaci assunti, prestazioni di specialistica, ecc)

Gli **assiomi** sono affermazioni sempre vere sul dominio definito dell'ontologia. Esempio *AIDS conclamato è HIV positivo*.

Gli assiomi sono inclusi in un'ontologia per scopi differenti:
vincolare la sua informazione, verificare la correttezza, dedurre nuova conoscenza.

Le **istanze**: a partire dalle classi dell'ontologia, si definiscono le istanze, che rappresentano specifici oggetti del mondo reale. Le istanze ereditano attributi e relazioni dalle classi.

Classe = Antitrombotici

Istanza = *Dipiridamolo*

Una ontologia popolata con istanze e regole di inferenza pone i fondamenti di una vera e propria base di conoscenza (Figura 8). La base di conoscenza è uno strumento dinamico e riusabile. Spesso questo termine viene confuso con il termine database, ma il database si concentra sul dato ed è specifico di una realtà, non ha caratteristiche di riusabilità (Figura 9).

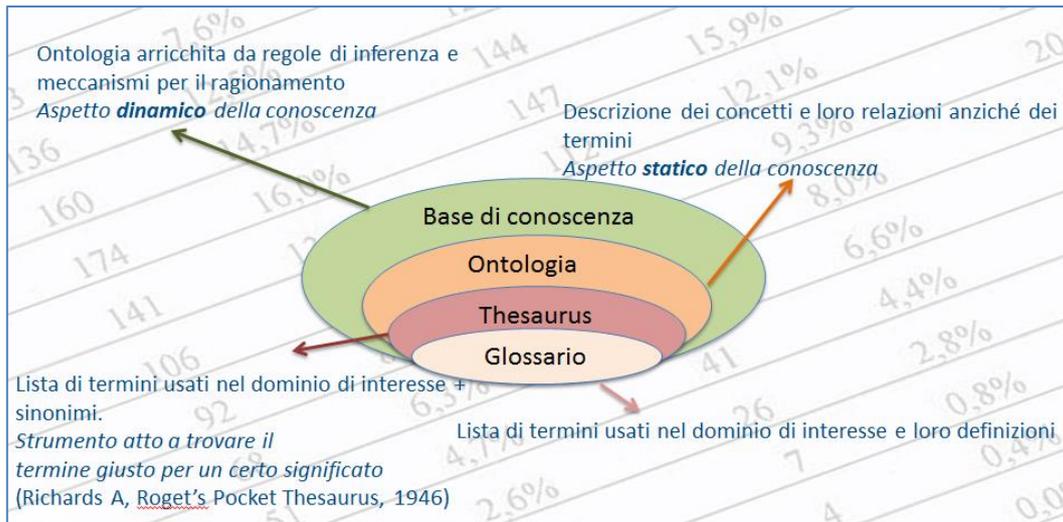


Figura 8

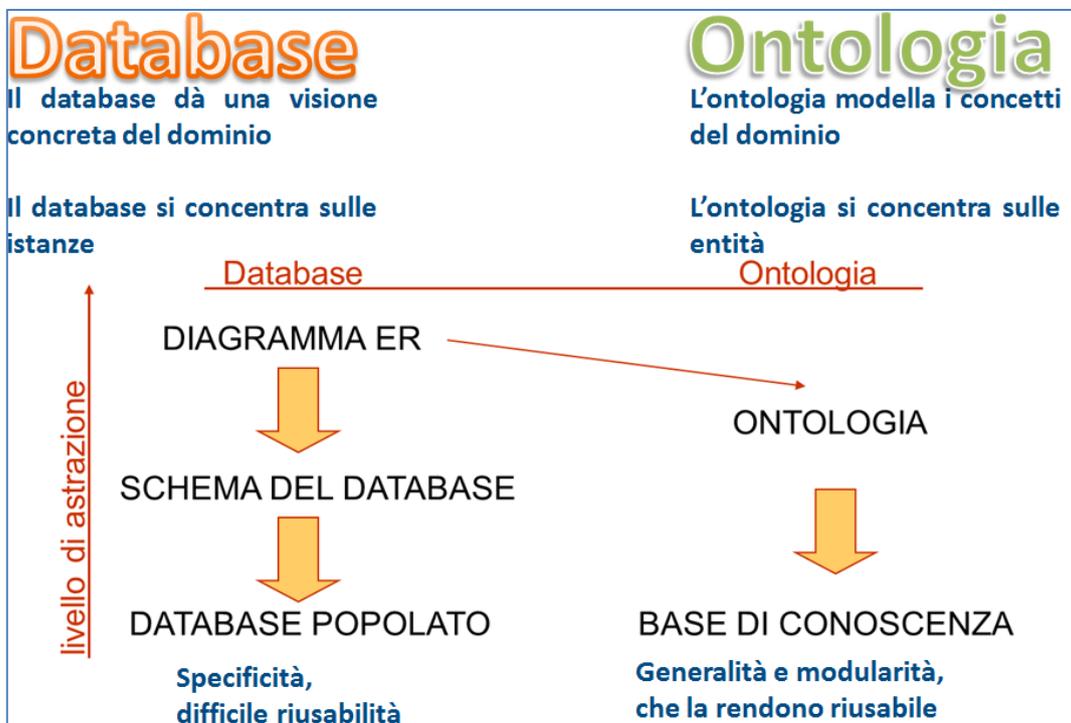


Figura 9

Da questa breve introduzione ai fondamenti teorici delle ontologie e da quanto descritto in precedenza, appare evidente che è soprattutto l'esperienza e la conoscenza del dominio di interesse, che aiuta nel processo di costruzione di un'ontologia.

Inoltre, il lavoro di modellazione ed editing deve essere supportato da uno strumento che fornisca una visualizzazione grafica dell'ontologia e meccanismi automatici di gestione e controllo che nascondano la complessità del linguaggio formale con cui esprimere l'ontologia .

La scelta del linguaggio

Una volta selezionati i componenti dell'ontologia, occorre definirli in maniera esplicita, mediante un opportuno linguaggio, condivisibile sul web.

XML Extensible Markup Language

Pur dando la possibilità di descrivere adeguatamente i contenuti di una risorsa web, non definisce meccanismi per qualificare le relazioni tra documenti. Quindi per esprimere in maniera formale le ontologie sono stati proposti diversi linguaggi, ciascuno con diverse capacità espressive

- RDF (Resource Description Framework)
- OIL (Ontology Interchange Language, Ontology Inference Layer)
- DAML (DARPA Agent Markup Language)
- OWL (Ontology Web Language)

RDF Resource Description Framework

RDF, creato da W3C come strumento per la modellazione di metadati, ispirato a quello entità-relazioni storicamente usato per la descrizione dei DataBase relazionali è divenuto uno standard di modellazione di risorse web sfruttando l'idea di descriverle attraverso la logica dei predicati: le informazioni sono **asserzioni** della forma soggetto-predicato-oggetto, le **TRIPLE**.

Ad esempio, 'Il paziente è affetto da cardiopatia' è rappresentato dalla tripla:

(Il paziente, è affetto da, cardiopatia)

RDF è un modello astratto con diversi formati di serializzazione ovvero di traduzione delle strutture dati in documenti digitali (ad esempio file di testo) e quindi il modo particolare con cui ogni tripla è codificata varia da formato a formato.

Il formato sintattico più frequentemente usato per le applicazioni RDF è l'XML

RDF lavora con tre tipi di elementi fondamentali, 1) le risorse ovvero le entità riferite attraverso URI, 2) le proprietà ovvero le relazioni binarie tra risorse e/o valori atomici primitivi, 3) le affermazioni che specificano il valore di una certa proprietà relativa a una risorsa. Per alcuni di questi elementi è possibile reperire sul Web le URI (risorse) che li identificano univocamente. Ad esempio:

Il paziente	https://www.ifc.cnr.it.cartellaclinicapaziente.anagrafe.jsp
È affetto da	http://it.wiktionary.org/wiki/affetto
cardiopatia	http://it.wikipedia.org/wiki/Cardiopatia

La formalizzazione XML di questa semplice tripla è

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:wikipedia="http://it.wikipedia.org/wiki/"
  xmlns:wikidizionario="http://it.wiktionary.org/wiki">

  <rdf:Description
    rdf:about="https://www.ifc.cnr.it.cartellaclinicapaziente.anagrafe.jsp">
    < wikidizionario:affetto
      rdf:resource=" http://it.wikipedia.org/wiki/Cardiopatia/>
    </rdf:Description>

  </rdf:RDF>
</xml>
```

OIL Ontology Inference Layer

OIL è un linguaggio standard per le ontologie che fornisce, a partire dagli Schemi RDF, un approccio multilivello, basato su una gerarchia stratificata. Permette di supportare la definizione di ontologie, il loro cambiamento, la loro verifica e integrazione per mezzo del sistema FaCT, un reasoner che esamina la consistenza di tutte le definizioni ontologiche per scoprire i legami fra le sottoclassi. Presenta tuttavia alcune limitazioni perché prevede un numero limitato e immutabile di proprietà senza la possibilità di avere definizioni di relazioni composte. Inoltre non supporta domini concreti come interi e stringhe, e questo costituisce un grande limite per la rappresentazione della conoscenza ontologica.

DAML DARPA Agent Markup Language

È un linguaggio di markup ontologico promosso negli Stati Uniti dall'agenzia Darpa che combina diverse caratteristiche di altri linguaggi basati sul Semantic Web. È basato su RDF e lo integra con elementi di logica descrittiva. Offre inoltre un mezzo per l'espressione ed interpretazione della classificazione, delle asserzioni e delle proprietà dei metadati rappresentati da RDF e RDFS

OWL Web Ontology Language

OWL è uno standard W3C. È un linguaggio a marcatori realizzato appositamente dal W3C per lo sviluppo delle ontologie e per la loro diffusione su WWW. Attualmente il linguaggio è distribuito in versione 2, informalmente Owl2. Le ontologie scritte in OWL 2 mettono a disposizione classi, proprietà, oggetti e valori e sono memorizzate come documenti del Web semantico. Le ontologie scritte in OWL 2 vengono scambiate come documenti RDF.

In particolare OWL consente di specificare meglio le relazioni tra le classi:

DISGIUNZIONE (maschi / femmine)

EQUIVALENZA (malattia / patologia)

COMBINAZIONI BOOLEANE:

- **Unione** (patologie = croniche e non croniche)
- **Intersezione** (Diabete Gestazionale = Diabete Mellito durante la gravidanza)

Il linguaggio inoltre permette di definire restrizioni di valore e cardinalità sulle proprietà.

Il quadrante sinistro in Figura 10 mostra uno stralcio di ontologia, creata attraverso un editor grafico che rappresenta il primo livello di classificazione della Cardiovascolopatia. Il quadrante di destra mostra la relativa sintassi OWL

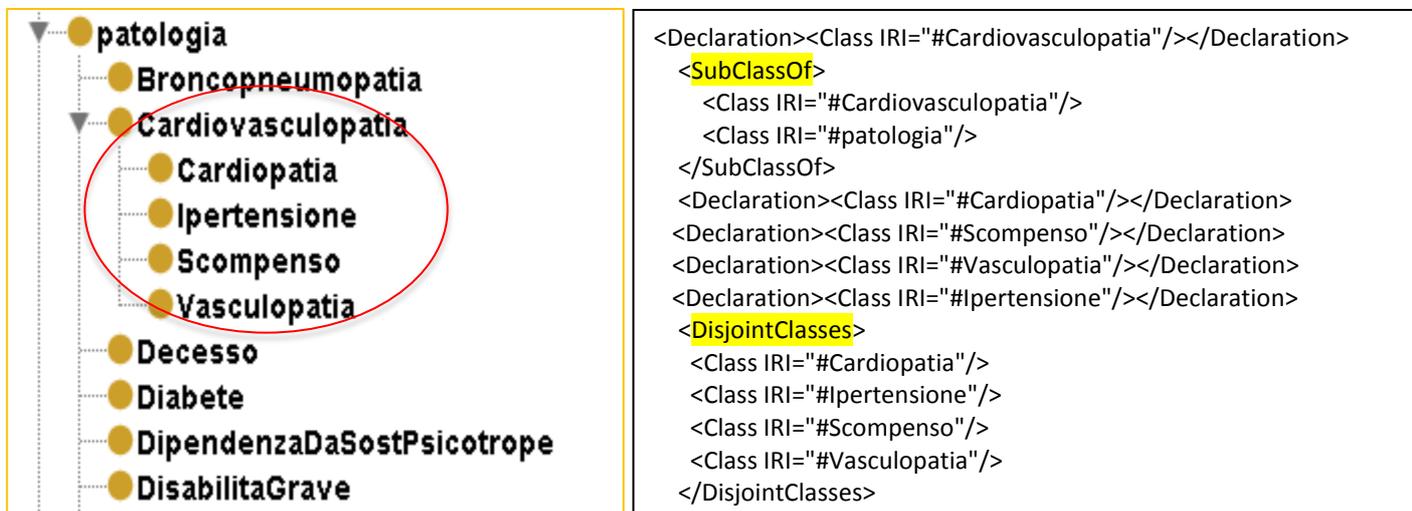


Figura 10: esempio di ontologia in owl

Di fatto, nell'ambito del Web Semantico, si è giunti ad un insieme di linguaggi standard (W3C), ciascuno dei quali espande ed integra le potenzialità del livello precedente:

HTML	Non distingue tra struttura del contenuto e rappresentazione del documento
XML	Permette di definire i <u>tag</u> e di strutturare le informazioni indipendentemente dalla loro resa grafica
RDF	Permette di esprimere relazioni esistenti tra istanze di concetti. Ad esempio <u>AffettoDa</u> (Mario Rossi, Diabete Mellito)
RDFS	Permette di definire relazioni e concetti. Ad esempio <u>AffettoDa</u> è una relazione, <u>Patologia</u> è un concetto
OWL	Poggia su RDFS e permette di esprimere pienamente classi, relazioni, istanze, vincoli, ecc.

Figura 11: dall'HTML all'OWL

SKOS Simple Knowledge Organisation System

SKOS, nasce come area di lavoro che sviluppa specifiche e standard per supportare l'uso dei sistemi d'organizzazione della conoscenza (KOS) nel quadro del web semantico. Sviluppato inizialmente nel 2003 come progetto open-source all'interno del progetto SWAD-E (Semantic Web Advanced Development for Europe) intendeva definire un modello per thesaurus compatibile con gli standard ISO più importanti (ISO 2788, ISO 5964). Dal Settembre 2004 la responsabilità per i successivi sviluppi è passata al W3C e nel 2005 è stato pubblicato il primo Working Draft. Oggi è uno standard del W3C basato su altri standard del Semantic Web (RDF e OWL), è a sua volta una ontologia OWL che definisce come formalizzare glossari, tassonomie e thesauri.

SKOS Core Vocabulary consiste in una serie di classi RDFS e proprietà di RDF che sono usate per rappresentare il contenuto e la struttura di base dei cosiddetti concept-schema cioè thesaurus, schemi di classificazione, elenchi di titoli d'argomento, tassonomie, terminologie, glossari e altri tipi di vocabolario. Poiché sono in formato RDF, questi possono essere manipolati, trasformati, archiviati, e vi si possono fare ricerche usando gli strumenti RDF. Pertanto SKOS può essere usato come formato d'interscambio attraverso le librerie digitali e per il recupero dati. Inoltre è possibile l'interazione con altri strumenti e rappresentazioni usati nel web semantico.

La scelta dello strumento di editing e gestione automatica della ontologia

Data la complessità di scrittura della ontologia mediante la scrittura esplicita in un linguaggio formale, risulta indispensabile un strumento avanzato di Ontology Editing che sia di supporto alla creazione e mantenimento dell'ontologia e che generi la risorsa digitale per renderla fruibile via WEB.

La rapidità di sviluppo tecnologico degli ultimi decenni ha prodotto un'ampia serie di strumenti validi, alcuni dei quali sono mostrati in tabella

Chimaera	COE
Enterprise Architect	FluentEditor for OWL
HOZO	KAON
Neologism	OBO-Edit
OntoStudio	Ontolingua
Protégé	Semantic Turkey
Swoop	Semaphore Ontology Manager
Synaptica	WebODE
OWLGrEd	Thesaurus Master
TODE	WebODE

Nella scelta dello strumento vanno considerate alcune caratteristiche quali ad esempio la modalità di lavoro (Online o Stand-alone), l'usabilità, la velocità di apprendimento, la possibilità di attivare funzionalità di reasoning, la possibilità di fare inferenza, di effettuare ricerca testuale sull'intera ontologia, di impostare link fra concetti, di export in diversi formati, la facilità di gestione delle modifiche all'ontologia o di effettuare merge fra ontologie diverse.

Un numero di MohaByte di qualche anno fa metteva a disposizione una tabella di valutazione in base alle caratteristiche:

	Protégé 3.4 beta	SWOOP	OWLSight	Ontology Browser	Ontology Explorer
Online/ Stand-alone	Stand-alone	Stand-alone	Online	Online	Online
Editing	SI	SI	NO	NO	NO
Reasoning /Inferenza	SI	SI	SI (usa Pellet)	NO	NO
Uso Label	SI	NO	NO	NO	SI
Istanze nella tassonomia	NO	NO	NO	NO	SI
Ricerca	SI	SI	NO	SI	SI
Ricerca Interattiva	SI	NO	NO	SI	SI
Ricerca su label	NO	SI	NO	NO	SI
Ricerca su comment	NO	SI	NO	NO	NO
Mostra ontology meta info	SI	SI	SI	SI	SI
Scelta della lingua	NO	SI	NO	NO	NO
Gerarchia proprietà	SI	SI	SI	SI	NO
Show properties inherited	SI	NO	NO	SI	SI
Valori "hasValue"	SI	SI	NO	NO	SI
Altre Restrizioni OWL	SI	SI	NO	SI	NO
Visualizzazione multipla concetti	NO	NO	NO	NO	SI

Figura 12: confrontare gli strumenti

Nell'ambito del progetto ODINET, abbiamo scelto di utilizzare Protege-OWL. Nato come progetto di modellazione ontologica del dominio biomedico presso l'università di Stanford, è attualmente uno strumento completo per la gestione di ontologie in OWL2. E' uno strumento open source scritto in linguaggio Java, scaricabile gratuitamente all'indirizzo <http://protege.stanford.edu/> e fornisce un ambiente grafico e interattivo per la progettazione delle ontologie e un ambiente di sviluppo concettuale.

La sua architettura, basata su componenti, semplifica l'aggiunta di nuove funzionalità attraverso la Protégé Plugin library che raccoglie plug-in creati dalla community di sviluppatori di tutto il mondo.

I plug-in più usati sono quelli che forniscono capacità di visualizzazione avanzata, come OntoGraf che visualizza un'ontologia come un grafico le classi e le istanze sono rappresentate come nodi, mentre le relazioni sono visualizzate come archi orientati.

Menzioniamo inoltre SKOSEd, un plugin che fornisce una finestra di lavoro Protege (chiamata Skos VIEW) per la modifica di entità SKOS e fornisce diverse prospettive sulla tassonomia SKOS (ad esempio filtrando i concetti a seconda delle skos:ConceptSchemes cui appartengono)

Principali ontologie in ambito sanitario

L'ambito biomedico vanta un numero elevato di risorse ontologiche.

Il portale BioPortal (<http://bioportal.bioontology.org/>) sviluppato dal National Center for Biomedical Ontology, funge da repository on line delle principali ontologie biomediche e permette di esplorare e visualizzare contemporaneamente più ontologie collegate fra loro, di navigare all'interno di una specifica ontologia, oppure di recuperare risorse ontologiche inerenti un termine specifico.

In questo paragrafo citeremo alcune delle risorse ontologiche più significative del dominio sanitario.

a. MeSH – Medical Subject Headings

Il dizionario MeSH è il componente principale dell'Unified Medical Language System (UMLS) un ambizioso progetto della National Library of Medicine (NLM) per lo sviluppo di sistemi di information retrieval.

MeSH è un dizionario standardizzato per consentire una ricerca efficace sul dominio della medicina, che utilizza molto frequentemente terminologie diverse per esprimere lo stesso concetto, anche al fine di dare consistenza e uniformità alla catalogazione della letteratura biomedica. MeSH consente la indicizzazione di Medline, un database di letteratura biomedica della United States National Library of Medicine, uno dei database più importanti e sicuramente il più diffuso grazie alla piattaforma internet Pubmed che lo ospita e lo diffonde gratuitamente a livello mondiale. I termini MeSH rappresentano le parole chiave attraverso le quali si possono ricercare le citazioni all'interno del database ed esprimono un concetto che gli autori potrebbero formulare con termini diversi. I concetti all'interno del MeSH sono organizzati secondo una struttura gerarchica ad albero (MeSH Tree Structure) in modo che effettuando una ricerca con un termine generico si possa arrivare anche ad articoli indicizzati con termini più specifici.

Le categorie principali sono visualizzate in tabella

Anatomy [A]	Anthropology, Education, Sociology and Social Phenomena [I]
Organisms [B]	Technology, Industry, Agriculture [J]
Diseases [C]	Humanities [K]
Chemicals and Drugs [D]	Information Science [L]
Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]	Named Groups [M]
Psychiatry and Psychology [F]	Health Care [N]
Phenomena and Processes [G]	Publication Characteristics [V]
Disciplines and Occupations [H]	Geographicals [Z]

b. UMLS Metathesaurus - Unified Medical Language System

UMLS è un repository di concetti biomedici correlati fra loro e ha l'obiettivo di unificare molte delle classificazioni mediche e dei dizionari esistenti, per cui Integra più di 2 milioni di nomi e 900.000 concetti.

Rappresenta un compendio di circa 100 vocabolari combinati attraverso un processo di mappatura retrospettiva basata sull'identificazione di relazioni di sinonimia fra i termini che li costituiscono. I vocabolari non hanno architettura comune e i concetti comuni vengono identificati con un unico codice (CUI – concept unique identifier).

E' possibile effettuare il download delle 32 ontologie UMLS gratuitamente attraverso il link

<http://bioportal.bioontology.org>

ONTOLOGY NAME	VISIBILITY	TERMS	NOTES	REVIEWS	PROJECTS	UPLOADED	CONTACT
AI/RHEUM (AIR)	Public	681	0	0	2	02/05/2010	May Cheh
Coding Symbols for a Thesaurus of Adverse Reaction Terms (CST)	Public	1,641	0	0	3	07/31/2009	U.S. Food and Center for Drug Research
CRISP Thesaurus, 2006 (CSP)	Public	8,912	0	0	2	08/27/2010	Anita Ghebeles
Current Procedural Terminology (CPT)	Public	9,442	0	0	2	05/26/2010	Dorith Brown
Foundational Model of Anatomy (FMA)	Public	83,281	0	1	15	09/14/2010	Onard Mejino
Gene Ontology (GO)	Public	39,248	0	1	56	04/25/2013	Jane Lomax
Gene Ontology Extension (GO-EXT)	Public	34,400	0	0	3	06/24/2011	Gene Ontology
Healthcare Common Procedure Coding System (HCPCS)	Public	5,660	0	0	3	09/01/2010	Cynthia Hake
ICD-10-PCS (ICD10PCS)	Public	177,782	0	0	2	09/29/2011	Pat Brooks
ICD10 (ICD10)	Public	12,318	0	0	5	07/21/2010	World Health C
ICD10CM (ICD10CM)	Public	91,590	0	0	2	09/29/2011	Patricia Brooks
ICPC-2 PLUS (ICPC2P)	Public	8,319	0	0	2	02/05/2010	Prof Helena Br
International Classification of Diseases (ICD9CM)	Public	22,400	0	1	8	04/06/2012	Patricia Brooks
International Classification of Primary Care (ICPC)	Public	747	0	0	3	07/31/2009	Henk Lambert Amsterdam
Logical Observation Identifier Names and Codes (LNC)	Public	171,399	0	0	7	06/08/2012	Ms. Kathy Mer
Master Drug Data Base (MDDB)	Public	12,467	0	0	2	02/04/2010	Karen Eckert
MedDRA (MDR)	Public	69,389	0	0	3	02/04/2010	MSSO Help Des
Medical Subject Headings (MeSH) (MSH)	Public	229,698	0	1	10	02/02/2012	Stuart Nelson,
MedlinePlus Health Topics (MEDLINEPLUS)	Public	970	0	0	3	07/31/2009	Ms. Naomi Mill
Metathesaurus CPT Hierarchical Terms (MTHCH)	Public	1,033	0	0	3	06/28/2010	Jan Willis
National Drug Data File (NDDF)	Public	26,866	0	0	2	02/04/2010	First Data Repl

Figura 12: download di ontologie UMLS

c. SNOMED CT - Systematized Nomenclature Of Medicine Clinical Terms

SNOMED CT è il risultato di uno sviluppo congiunto dell'NHS (National Health Service⁴) in Inghilterra e del CAP (College of American Pathologists).

Rappresenta la terminologia clinica sanitaria più completa e multilingue al mondo (utilizzata in circa 50 paesi, tra cui non compare l'Italia). Il suo obiettivo è rappresentare precisamente le informazioni cliniche in tutti i campi di applicazione dell'assistenza sanitaria: fornisce infatti contenuti clinici per la documentazione clinica e collabora con altri standard internazionali. È una risorsa essenziale per le cartelle cliniche elettroniche, con un contenuto completo e validato scientificamente.

Le sue categorie principali, sono visibili in Figura 14:

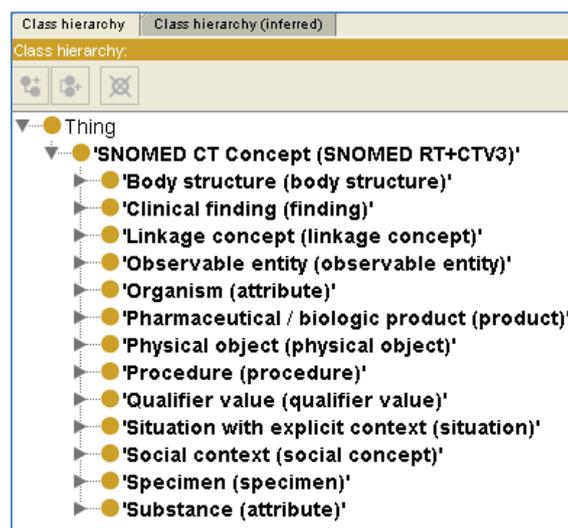


Figura 14: Snomed Concepts attraverso Protege

d. GALEN

Il progetto GALEN, che ha dato vita al più noto Open GALEN, nasce in ambito europeo e punta alla realizzazione di un modello di conoscenza relativo a termini medici. Il progetto GALEN ha sviluppato un cospicuo insieme di concetti che sono stati organizzati mediante l'utilizzo del linguaggio (definito in GALEN) GRAIL. Oggi è possibile tradurre le ontologie GRAIL in ontologie OWL-RDF (OWL 2.0).

L'obiettivo di GALEN è quello di rendere il più possibile machine-friendly lo scibile e le informazioni in campo (bio)medico. Il risultato è un sistema computerizzato di codifica multilingue per l'utilizzo in campo medico. Tale obiettivo punta a risolvere alcune questioni chiave quali il problema di avere sistemi multilingue che però siano in grado di preservare e condividere la conoscenza sottostante allo specifico linguaggio oppure la necessità di superare la barriera tra l'elevato livello di dettaglio informativo (prossimo al linguaggio naturale) necessario per la gestione dei dati clinici di un paziente e quello aggregato utile a fini statistici. I fronti su cui GALEN si propone sono molteplici e riguardano il superamento dei vincoli informativi legati alla semantica del dato.

Un primo risultato è il GALEN Common Reference Model, un insieme di concetti medici. L'ontologia di alto livello di GALEN è indipendente dalla lingua utilizzata ed è uno schema di che consente di stabilire le regole fondamentali per la collaborazione e la condivisione dell'informazione. GALEN si basa su un principio fondamentale che è quello della separazione tassonomica. In pratica, i concetti vengono descritti in termini di loro componenti che vengono raggruppati secondo gerarchie pure. Questi elementi di gerarchie pure vengono quindi ricombinati a formare tutti i concetti necessari. La classificazione di tali composizioni può essere completamente automatizzata.

E' possibile effettuare il download gratuito di GALEN all'indirizzo

<http://www.co-ode.org/galen/>

d. OBO - Open Biomedical Ontologies

Il progetto OBO Foundry nasce dalla collaborazione di numerosi sviluppatori di ontologie con il fine di stabilire un insieme di principi per la loro costruzione con l'obiettivo di creare un insieme di ontologie interoperabili in campo biologico e biomedico. Le principali ontologie della OBO Foundry sono: i processi biologici, le componenti cellulari, le entità chimiche di interesse biologico, le funzioni molecolari, qualità fenotipica, proteine e anatomia. Il formato OBO è molto ricco e può essere convertito in OWL. Allo scopo di agevolare la conversione bidirezionale tra i due linguaggi per la rappresentazione di ontologie è nato un progetto specifico che permette la traduzione di ontologie scritte per uno nell'altro senza alcuna perdita. Le ontologie OBO possono essere scaricate gratuitamente ed in diversi formati all'indirizzo: <http://www.berkeleybop.org/ontologies/>

Ontologies		
Ontology	ID Prefix	Files
AERO:Adverse Event Reporting Ontology	AERO	pro[45 kb md5] -- obo[25 kb md5] -- obo_xml[34 kb md5] -- owl[77 kb md5] -- tbl[12 kb md5] -- prolog[45 kb md5] -- obo.html[101 kb md5] --
NIF_Cell:NIF Cell	NIF_Cell	pro[462 kb md5] -- obo[293 kb md5] -- obo_xml[536 kb md5] -- owl[460 kb md5] -- tbl[32 kb md5] -- prolog[462 kb md5] -- obo.html[1219 kb md5] --
NIF_Dysfunction:NIF Dysfunction	NIF_Dysfunction	pro[106 kb md5] -- obo[56 kb md5] -- obo_xml[67 kb md5] -- owl[826 kb md5] -- tbl[19 kb md5] -- prolog[106 kb md5] -- obo.html[269 kb md5] --
NIF_GrossAnatomy:NIF Gross Anatomy	NIF_GrossAnatomy	pro[754 kb md5] -- obo[496 kb md5] -- obo_xml[566 kb md5] -- owl[3030 kb md5] -- tbl[100 kb md5] -- prolog[754 kb md5] -- obo.html[2016 kb md5] --
OGMS:Ontology for General Medical Science	OGMS	pro[25 kb md5] -- obo[17 kb md5] -- obo_xml[24 kb md5] -- owl[61 kb md5] -- tbl[11 kb md5] -- prolog[25 kb md5] -- obo.html[71 kb md5] --
adult_mouse_anatomy:Mouse adult gross anatomy	MA	obo_xml[661 kb md5] -- obo[318 kb md5] -- owl[1439 kb md5] -- tbl[105 kb md5] -- prolog[754 kb md5] -- owl2[636 kb md5] --
adverse_events:Ontology of Adverse Events	OAE	pro[698 kb md5] -- obo[382 kb md5] -- obo_xml[511 kb md5] -- owl[1036 kb md5] -- tbl[183 kb md5] -- prolog[698 kb md5] -- obo.html[1501 kb md5] --
aeo:Anatomical Entity Ontology	AEO	obo_xml[122 kb md5] -- obo[63 kb md5] -- owl[290 kb md5] -- tbl[31 kb md5] -- prolog[117 kb md5] -- owl2[120 kb md5] -- xref[6 kb md5] --
amphibian_anatomy:Amphibian gross anatomy	AAO	obo_xml[680 kb md5] -- obo[368 kb md5] -- owl[1549 kb md5] -- tbl[173 kb md5] -- prolog[678 kb md5] -- owl2[779 kb md5] -- xref[29 kb md5] --
amphibian_taxonomy:Amphibian taxonomy	ATO	Could not index: Source not available at specified URL as recognised format
animalhist:Animal natural history and life history	ADW	
ascmycete_phenotype:Ascomycete phenotype ontology	APO	obo_xml[138 kb md5] -- obo[83 kb md5] -- owl[317 kb md5] -- tbl[31 kb md5] -- prolog[173 kb md5] -- owl2[100 kb md5] -- xref[4 kb md5] --
bfo:Basic Formal Ontology	BFO	pro[63 kb md5] -- obo[52 kb md5] -- obo_xml[74 kb md5] -- owl[44 kb md5] -- tbl[8 kb md5] -- prolog[63 kb md5] -- obo.html[151 kb md5] -- imports-local.obo[0 kb md5] -- imports-local.owl[4 kb md5] -- owl2[119 kb md5] -- xref[10 kb md5] --

Figura 15: ontologie OBO

EUROVOC

EuroVoc è un thesaurus multilingue e pluridisciplinare che considera tutti i settori inerenti alle attività delle istituzioni europee:

- vita politica
- relazioni internazionali
- Comunità europee
- diritto
- vita economica
- scambi economici e commerciali
- finanze
- questioni sociali
- istruzione e comunicazione
- scienze
- impresa e concorrenza
- occupazione e lavoro
- trasporto
- ambiente
- agricoltura, silvicoltura e pesca
- agroalimentare
- produzione, tecnologia e ricerca
- energia
- industria
- geografia
- organizzazioni internazionali

E' disponibile in 23 lingue ufficiali dell'Unione europea, nonché nella lingua di un paese extra-UE (serbo). EuroVoc è curato dall'**Ufficio delle pubblicazioni**, che è passato ad una gestione del thesaurus basata sull'ontologia e alle tecnologie del web semantico, in particolare SKOS, in linea con le raccomandazioni del W3C e con gli ultimi sviluppi negli standard di classificazione.

Relativamente al dominio sanitario, il microtesauro 2841 appartenente al settore 28 (questioni sociali) è quello dedicato alla salute.

WORDNET

E' importante citare anche Wordnet, che pur non essendo una ontologia medica, rappresenta il database lessicale maggiormente utilizzato e riconosciuto: esso è stato sviluppato presso l'università di Princeton, sotto la direzione del professore G.A. Miller ed è disponibile gratuitamente all'indirizzo <https://wordnet.princeton.edu/wordnet/download/> a condizione che siano citati gli autori ed il sito ufficiale del progetto.

Wordnet non è un semplice dizionario di termini inglesi, ma un sistema lessicale di riferimento il cui disegno si basa sul meccanismo della memoria umana e sulle teorie psico-linguistiche: i vocaboli, infatti, non sono disposti in ordine alfabetico, ma per affinità di significato. WordNet comprende quattro categorie sintattiche: nomi, verbi, aggettivi e avverbi, ognuna suddivisa in diversi insiemi di sinonimi; questi ultimi sono associati ad un unico significato condiviso da tutti i termini che li compongono. Un termine può naturalmente possedere più di un significato ed essere presente in molti di questi insiemi, ed anche in più di una categoria sintattica.

DATI OPEN

L'Open Data è un fenomeno in forte ascesa, incentivato anche dalle ultime strategie comunitarie (ad esempio il programma dell'Agenda Digitale Europea di Dicembre 2011). Questo modello prevede di "liberare i dati pubblici", ovvero renderli accessibili online, favorendo la trasparenza e consentendo il loro ritorno alla comunità come un valore che crea sviluppo.

Un primo monitoraggio nazionale è stato effettuato nel 2013 e condotto da *Diritto Di Sapere*, associazione creata nel 2012 grazie al sostegno dell'[Open Society Foundations](#) con lo scopo di difendere ed espandere il diritto umano di accesso all'informazione.

Trentatré rappresentanti di organizzazioni della società civile, giornalisti ecc, hanno presentato 300 richieste di informazioni ad autorità della PA in nove aree tematiche: ambiente, diritti umani, giustizia, spesa pubblica, istruzione, servizi sociali, spettacoli, investimenti finanziari e salute.

Solo il 27% delle risposte ricevute è risultato soddisfacente mentre il 73% è risultato non adeguato agli standard internazionali previsti. Il livello di silenzio, ovvero la completa assenza di risposta dalla PA dopo 30 giorni rappresenta ben il 65% delle richieste, tant'è che il report è stato titolato '[The Silent State](#)'.

Sono disponibili comunque alcune fonti istituzionali di dati open che vengono citate di seguito.

a. FONTE Ministero della Salute

Indirizzo: <http://opendatasalute.cloudapp.net/DataCatalog/DataSetList>

E' basata sulla tecnologia Cloud PaaS Platform as a Service (Microsoft Azure) e sul toolkit, gratuito e Open Source, denominato OGD – Open Government Data Initiative: mediante tale soluzione sono resi disponibili in questo portale una serie di strumenti cloud-based aperti e interoperabili.

Alcuni dati possono includere informazioni geospaziali, restituite nel formato Keyhole Markup Language (KML). Questo formato è compatibile con le più popolari tecnologie per la gestione delle mappe sia desktop che web, incluse Microsoft Bing Maps, Google Maps, Yahoo! Maps, Google Earth.

Formato: Export diretto CSV, e API client per gli sviluppatori per la maggior parte dei dataset

Aggiornamento: Aprile 2013

Copertura: Nazionale

Fra i dati disponibili menzioniamo i seguenti come particolarmente rilevanti per il dominio sanitario.

Farmacie

Contiene l'elenco completo delle farmacie aperte al pubblico (includendo anche succursali, dispensari e dispensari stagionali), includendo Dati Anagrafici e di localizzazione delle Farmacie

Distributori di Farmaci

Contiene l'elenco completo dei Siti Logistici attivi autorizzati alla distribuzione di medicinali ad uso umano sul territorio nazionale

Aziende Sanitarie Locali

Elenco delle ASL: per ciascuna ASL è riportato il recapito della sede della ASL

Dispositivi medici

Il dataset contiene l'elenco completo dei dispositivi medici registrati presso la banca dati e il Repertorio del Ministero della Salute.

Farmaci più venduti nell'ultimo semestre

Il dataset contiene l'elenco dei 50 medicinali senza obbligo di prescrizione più venduti, nell'ultimo semestre

Posti letto per struttura ospedaliera

Il data-set contiene i posti letto accreditati e a carico del Servizio Sanitario Nazionale, presenti al 1° gennaio dell'anno di riferimento in ciascuna struttura ospedaliera pubblica, equiparata o casa di cura privata accreditata.

Posti letto per regione e disciplina

Per ciascuna Regione e per ciascuna disciplina, sono riportati, su righe distinte, i dati relativi alla dotazione di posti letto, distinti in posti letto per degenza ordinaria, day hospital, day surgery e totali.

Corrispondenze ASL-Comuni e popolazione residente

Per ciascuna ASL è riportato l'elenco dei comuni che afferiscono alla ASL; inoltre per ogni comune viene fornita la popolazione residente distinta per genere.

Per i comuni articolati su più ASL, sono riportati gli ambiti territoriali subcomunali (circostrizioni/municipi) che afferiscono alla ASL.

b. FONTE Portale Italiano degli Open Data

Indirizzo <http://www.datiopen.it/>

Formato: Export diretto CSV, XLS, e API client

Aggiornamento: 2013

Copertura: Nazionale

Le risorse sono descritte secondo le 15 proprietà standard definite dal Dublincore. Al 24 Aprile 2013 erano reperibili 81 dati selezionabili attraverso la parola chiave Sanità, di cui menzioniamo i seguenti come particolarmente rilevanti per il dominio sanitario.

Giornate in strutture semiresid. e resid. malati terminali ogni 1000 abitanti e regione

Distribuzione per Regione del valore dell'indicatore: Giornate in strutture semiresidenziali e residenziali per l'assistenza ai malati terminali ogni 1000 residenti

Fonti di finanziamento del Servizio Sanitario Nazionale

Riporta i valori assoluti fonti di finanziamento (Irap e addizionale IRPEF; Fabbisogno ex D.L.vo 56/2000; FSN e ulteriori integrazioni a carico dello Stato). Dati relativi all'anno 2009.

Numero di bambini residenti e di scelte per Pediatra di Libera Scelta

Distribuzione per Regione del valore dell'indicatore: nr. di bambini residenti e di scelte per Pediatra di Libera Scelta.

Degenza media per acuti, regime ordinario e regione

Distribuzione per Regione del valore dell'indicatore: Attività di degenza per acuti in regime ordinario: degenza media - Ospedali pubblici e Case di Cura private accreditate e non accreditate.

Ospedali per classe di posti letto e regione

Distribuzione per regione degli ospedali per classe di posti letto (fino a 120 pl; 121-400; 401-600, 601-800; 801-1500; oltre 1500).

Percentuale di anziani trattati in Assistenza Domiciliare Integrata (ADI) per regione, dal 2005 al 2010

Percentuale di anziani trattati in Assistenza Domiciliare Integrata (ADI) rispetto al totale della popolazione anziana (>=65 anni), per regione: Anni 2005 - 2010.

Numero di scelte per Medico di Medicina Generale per regione

Distribuzione per Regione del valore dell'indicatore: nr. adulti residenti e di scelte per Medico di Medicina Generale.

Percentuale di anziani trattati in Assistenza Domiciliare Integrata (ADI) per Regione

Distribuzione per Regione del valore dell'indicatore: percentuale di anziani (>= 65 anni) trattati in Assistenza Domiciliare Integrata (ADI) pubblicato il 31 luglio 2012.

Spesa pro-capite del Servizio Sanitario Nazionale per regione - Anno 2009

Consumo nazionale di farmaci a carico del Servizio Sanitario Nazionale per categorie terapeutiche

ATC al I° livello - Dati relativi all'anno 2009.

Giornate in strutture semiresid. e resid. per disabili fisici ogni 1000 residenti e regione

Distribuzione per Regione del valore dell'indicatore: Giornate in strutture semiresidenziali e residenziali per l'assistenza ai disabili fisici ogni 1000 abitanti - Anno 2008. Dati relativi all'anno

Posti residenziali in strutture riabilitazione per regione - Anno 2008

Distribuzione per Regione del valore dell'indicatore: posti residenziali nelle strutture di riabilitazione ex art.26 L.833/1978 per 1000 residenti. Dati relativi all'anno 2008.

Indice di attrazione, indice di fuga per acuti, regime ordinario per regione - Anno 2008

Distribuzione per Regione dei valori degli indicatori: indice di attrazione, indice di fuga per acuti, regime ordinario. Dati relativi all'anno 2008

Tasso di ospedalizzazione per acuti e regione- Anno 2008

Distribuzione per Regione del valore dell'indicatore: tasso di ospedalizzazione per acuti (ricoveri in Regime Ordinario e Day Hospital per acuti ogni 1000 residenti). Dati relativi all'anno 2008

Personale medico ospedaliero per 1000 residenti per regione - Anno 2008

Distribuzione per Regione del valore dell'indicatore: Personale medico che opera nelle Strutture di Ricovero Pubbliche per 1.000 residenti. Dati al 1° gennaio 2008.

Informazioni sui reparti delle strutture di ricovero pubbliche

Dati di struttura ed attività dei reparti presenti in ciascuna struttura di ricovero pubblica.

Informazioni sul personale delle case di cura private accreditate

Il dato contiene numerose informazioni quantitative sul personale delle case di cura private accreditate.

Informazioni su strutture e attività delle ASL

Il dato contiene numerose informazioni quantitative sulle strutture e attività ASL presenti nel territorio della ASL, classificate per comune ed estese al territorio nazionale.

Case di cura accreditate

E' la tabella delle case di cura accreditate presenti nel territorio della ASL e relativi dati statistici

Strutture di ricovero pubbliche

Il dato contiene numerose informazioni quantitative sulle strutture di ricovero pubbliche presenti nel territorio della ASL, classificate per regione ed estese al territorio nazionale.

c. FONTE ISTAT

Indirizzo <http://dati.istat.it/>

Formato: Export diretto CSV, XLS, e API client

Aggiornamento: 2013

Copertura: Nazionale

I dati rilevanti in ambito socio-sanitario sono sintetizzati in tabella

Stili di vita e fattori di rischio

Stile alimentare

Abitudine al fumo

Condizioni di salute

Stato di salute della popolazione - per 100 persone

Persone con presenza di alcune malattie croniche

Consumo di farmaci nei due giorni precedenti l'intervista

Cause di morte

Mortalità

Mortalità infantile

Incidenti stradali

Incidenti stradali con lesioni alle persone

Morti e feriti in incidenti stradali

Veicoli coinvolti in incidenti stradali

Indicatori di incidentalità stradale

Veicoli - Pubblico registro automobilistico

Salute riproduttiva della donna

Dimissioni dagli istituti di cura per aborto spontaneo rispetto al luogo di residenza

Dimissioni dagli istituti di cura per aborto spontaneo rispetto al luogo dell'evento

Interruzioni volontarie della gravidanza

Ricorso ai servizi sanitari

Ospedalizzazione per disturbi psichici rispetto al luogo di dimissione

Ospedalizzazione per disturbi psichici rispetto al luogo di residenza

Strutture sanitarie distrettuali

Assistenza sanitaria di base

BIBLIOGRAFIA DI RIFERIMENTO

Building an Ontology of Cardio-Vascular Diseases for Concept-Based Information Retrieval

S Gedzelman, M Simonet, D Bernhard, G Diallo, P Palmer
Université Joseph Fourier, TIMC Laboratory, Grenoble, France

Ontologies for Knowledge Representation in a Computer-Based Patient Record

Elisabeth Bayegan, Øystein Nytrø and Anders Grimsmo
*Norwegian University of Science and Technology, Department of Computer and Information Science
Department of Community Medicine and General Practice
NO-7491 Trondheim, Norway, bayegan,nytroe@idi.ntnu.no; grimsmo@medisin.ntnu.no*

The Unified Medical Language System (UMLS):integrating biomedical terminology

Olivier Bodenreider
*Lister Hill Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Building
38A, 8600 Rockville Pike, Bethesda, MD 20894, USA*

Spatial location and its relevance for terminological inferences in bio-ontologies

Stefan Schulz*^{1,3}, Kornél Markó^{1,2} and Udo Hahn²
*Medical Informatics Department, Freiburg University Hospital, Freiburg, Germany, Language and Information
Engineering (JULIE)
Lab, Jena University, Germany and 3Master Program in Health Technology, Pontifical Catholic University of Paraná,
Curtiba, Brazil*

LINEE GUIDA PER L'INTEROPERABILITÀ SEMANTICA ATTRAVERSO I LINKED OPEN DATA

COMMISSIONE DI COORDINAMENTO SPC
SPC Sistema Pubblico di Connettività e Cooperazione

Linked data: un nuovo alfabeto del web semantico

Mauro Guerrini, Tiziana Possemato

Congenital Heart Disease: An Ontology-Based Approach for the Examination of the Cardiovascular System

M. Esposito
*Institute for High-Performance Computing and Networking (ICAR)
National Research Council (CNR)*

LinKBase®, a Philosophically-inspired Ontology for NLP/NLU Applications

Maria van Gorp, PhD, Manuel Decoene, MD, Marnix Holvoet and Mariana Casella dos Santos, MD.
Language and Computing NV, Sint-Denijs-Westrem, Belgium

Heart failure ontology

Authors: Alan Jovica, Dragan Gambergerb, Goran Krstacic
*Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia; email:
alan.jovic@fer.hr
Rudjer Boskovic Institute, Bijenicka 54, 10000 Zagreb, Croatia; email: dragan.gamberger@irb.hr
Institute for Cardiovascular Diseases and Rehabilitation, Draskoviceva 13, 10000 Zagreb, Croatia;*

A SURVEY ON ONTOLOGY CONSTRUCTION METHODOLOGIES

R.Subhashini, Dr. J. Akilandeswari
*Department of CSE, Sona College of Technology, Salem, India
Department of IT, Sona College of Technology, Salem, India*

Ontology development 101: A guide to creating your first ontology.

N. F. Noy and D. McGuinness.
Technical Report KLS-01-05, Standford KSL, 2000.

Semantic Web Tra ontologie e open data

T. Di Noia, R. De Virgilio, E. Di Sciascio, F. M. Donini
2013 Apogeo – IF – Idee Editoriali Feltrinelli Srl

THE SILENT STATE Access to information in Italy

Results and recommendations from first national monitoring
Diritto Di Sapere, Access-Info Europe
April 2013

A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.3

Matthew Horridge
2011 University of Manchester

Progettazione ed implementazione di un'ontologia per il dominio commerciale-finanziario

Lavoro di Tesi M. Sartore A.A 2008-2009 Università degli Studi di Genova

Analisi comparata di due ontologie in ambito biomedico: snomed ct e rome

Lavoro di Tesi M. Gentile A.A 2009-2010 Università Politecnica delle Marche

Studio e sperimentazione dell'annotazione lessicale delle ontologie definizionali

Lavoro di Tesi Elisa Fusari A.A 2008-2009 Università degli Studi di Modena e Reggio Emilia