



UNIVERSITÀ
DEGLI STUDI
FIRENZE

**Dottorato Di Ricerca In Statistica, Informatica,
Applicazioni
Ciclo XXVII**

**Il ruolo di approcci e tecnologie semantiche nella
predisposizione di sistemi informativi statistici a supporto delle
decisioni in ambienti complessi**

Coordinatore:

Chiar.mo Prof. Corradi Fabio

Tutore:

Chiar.ma Prof.ssa Martelli Cristina

Candidato:

Kahlawi Adham

Anno Accademico 2015

Ringraziamenti

Desidero ricordare tutti coloro che mi hanno aiutato nella stesura della tesi con suggerimenti, critiche ed osservazioni: a loro va la mia gratitudine, anche se a me spetta la responsabilità per ogni errore contenuto in questa tesi.

Ringrazio anzitutto la professoressa Cristina Martelli, il professore Fabio Corradi, senza il loro supporto e la loro guida sapiente questa tesi non esisterebbe.

Un ringraziamento particolare va ai colleghi che mi hanno incoraggiato o che hanno speso parte del proprio tempo per leggere e discutere con me le bozze del lavoro. Un grazie speciale a Maria Flora Salvatori.

Vorrei ringraziare le persone a me più care:

Miei genitori Omar e Kawkab, senza il loro sostegno io non sarei riuscito a continuare il mio studio.

Miei fratelli e mie sorelle... Samira, Muhammad, Samir, Jihan, Fahid, Siba.

Mia moglie... Fatemah.

I miei nipoti

I miei amici...

Questo lavoro è dedicato al mio primo insegnante Profeta Muhammad (Maometto).



<< Signor mio, accresci la mia scienza>>

Quran, Surat Ta-Ha, Aia 114.

Indici	
1. Introduzione	1
2. Il problema del governo di contesti complessi	2
2.1. Costruire sistemi di fonti con strutture a rete complessa Evolutiva	2
2.2. La complessità: struttura dei problemi, del linguaggio	3
e dei dati	
2.3. Semantica memoria di sistema e comportamento a rete ...	5
2.4. Fare rete attraverso l'integrazione: il riuso statistico	7
delle fonti amministrative.	
3. Le fonti amministrative per l'analisi statistica	9
3.1. Definizioni e accezioni	9
3.2. I vantaggi di utilizzare fonti amministrative	10
3.3. Criticità delle fonti amministrative	11
3.4. Registri statistici come risultato della linkage tra fonti	13
amministrative diverse	
3.5. Fonti amministrative, interoperabilità e costituzione di	15
sistemi informativi statistici a rete	
3.6. Ruolo e importanza dei metadati	16
3.7. Qualità delle fonti amministrative per l'utilizzo statistico ..	17
4. Nuove metodologie per la modellazione e l'integrazione di	20
fonti e semantiche eterogenee: le ontologie	
4.1. Il semantic web: una evoluzione del web,	20
una prospettiva per la statistica.	
4.2. Le ontologie	21
4.3. Il web nella prospettiva di un database integrato:	22
il modello a grafo	
4.4. Dal graph database al sistema informativo semantico:	24
la tripla RDF	
4.4.1. Come funziona il RDF	25
4.4.2. RDF Schema	26
4.5. Web Ontology Language (OWL)	26
4.6. Utilizzare i sistemi semantici: SPARQL Protocol and	27
RDF Query Language (SPARQL) per la costruzione degli statistical data sets	
4.6.1. Protégé editor	28
4.7. Integrare le ontologie	28

4.7.1. L'integrazione tra le strutture e i dati delle ontologie	30
4.7.2. L'integrazione tra i triple store delle ontologie	30
4.8. Ruolo delle ontologie per la costituzione di archivi	30
amministrativi interoperabili e giacimenti di dati riusabili per a costituzione di registri statistici	
4.9. Ontologie e condivisione della conoscenza	31
4.10. Ontologie e database	34
4.10.1. Comunicazione tra un'ontologia e un database	35
4.10.2. Ottenere una ontologia da un database	35
4.10.2.1. Il metodo	36
4.10.2.2. Gli strumenti di conversione	36
4.10.3. Ottenere un database da un'ontologia	37
4.10.3.1. Il metodo	37
4.10.3.2. Gli strumenti di conversione	38
4.11. Ruolo delle ontologie per la costituzione di fonti	38
amministrative e registri statistici	
4.11.1. Ontologie e qualità della fonte statistica	39
4.11.2. Ruolo delle ontologie per la costituzione di sistemi ...	39
gestionali interoperabili e dei registri	
4.11.3. Il ruolo delle ontologie per la costruzione di fonti	41
integrabili e riusabili	
4.11.4. Ontologie e classificazioni statistiche.....	42
4.11.5. I sistemi gestionali a supportati delle ontologie	43
5. Caso di studio - Le grandi opere pubbliche	45
5.1. Database Monitor	45
5.2. La metodologia del trattamento delle tabelle	47
5.3. Tabella Lavoratori	48
5.3.1. Ontologie esterne	49
5.3.2. I concetti standard	50
5.4. La modulazione ontologica dei Lavoratori	54
5.5. I concetti standard di ontologia di grandi opera	62
5.6. La forma finale dell'ontologia di Grande Opera	66
5.7. La visualizzazione dell'ontologia	68
5.7.1. Interrogare graficamente una ontologia	70
5.8. Il collegamento tra due individui non sono collegati	71
Direttamente	
5.9. Estrarre e combinare le informazioni da due	74

ontologie diverie	
5.10. La creazione di un nuovo database Monitor sulla	75
basi dell'ontologia della Grande Opera	
5.11. Il funzionamento congiunto di una ontologia	79
e di un database	
6. Conclusioni	81
Bibliografia	83
Appendice 1	92
Appendice 2	97
Appendice 3	116
Appendice 4	129
Appendice 5	135
Appendice 6	137
Appendice 7	138

1. Introduzione

I sistemi informativi statistici ufficiali sono sempre più dipendenti dalle fonti e dai registri amministrativi che rappresentano sovente una alternativa economica, tempestiva e di grande dettaglio descrittivo.

Un limite importante al pieno sfruttamento di tali fonti da parte della statistica è rappresentato dalla qualità e dalla eterogeneità dei dati che spesso sono espressione di dialetti amministrativi e semantiche molto eterogenee.

Questa tesi intende esplorare le potenzialità offerte dalle nuove tecnologie semantiche sia sul piano dello sviluppo che della committenza di un sistema gestionale riutilizzabile dal punto di vista statistico. Non si sono affrontate applicazioni come i linked data, gli open data ed i big data. In questo lavoro si è riflettuto sulle possibilità di ottenere, grazie alle metodologie semantiche, delle fonti amministrative in grado di aderire profondamente a concettualizzazioni ufficiali e condivise: fonti amministrative buone, facilmente integrabili, non settoriali, in grado di rappresentare la complessità.

Il lavoro è così strutturato: innanzi tutto si sono discusse quelle caratteristiche dei sistemi complessi che devono essere soddisfatte anche dagli archivi amministrativi per poter rappresentare la complessità. Le fonti amministrative, come oggetti informativi generati lungo i processi di gestione, sono componenti informativi adeguati purchè in grado di integrarsi profondamente. Le metodologie semantiche sono state dunque richiamate in questa prospettiva: come nuovi strumenti per un classico obiettivo della statistica ufficiale, quello di costituire sistemi di fonti e registri in grado di sostenere misure su contesti complessi.

È stata infine proposta un'applicazione su un caso di studio e si è prodotta l'ontologia descrittiva delle grandi opere pubbliche.

2. Il problema del governo di contesti complessi

2.1. Costruire sistemi di fonti con strutture a rete complessa evolutiva

La complessità è una delle cifre caratteristiche della modernità che è chiamata a descriverla e a interpretarla a fini di governo e di decisione (Colander, Kupers 2014).

La costruzione di sistemi di indicatori e di misura in sistemi complessi esige dei sistemi di memoria avanzati, che siano in grado di rappresentare con i loro dati tale complessità.

Costruire questi sistemi di fonti e di archivi non è solo una questione tecnologica, ma innanzi tutto una questione di corretta concettualizzazione dell'area problematica (Wing, 2010).

L'importanza della fase di concettualizzazione è sempre stata affermata nella progettazione delle basi di dati, i cui modelli concettuali (tipicamente espressi attraverso il modello entità e relazioni) altro non sono se non una rappresentazione in logica del primo ordine della descrizione del contesto fornita dagli esperti di dominio.

Qualsiasi progettazione di fonti e archivi ha nella narrazione degli esperti del dominio il suo indispensabile punto di partenza.

Il linguaggio assume dunque un ruolo centrale, a maggior ragione nel caso della rappresentazione dei sistemi complessi.

Le nuove tecnologie semantiche offrono strumenti inediti nella traduzione delle narrazioni degli esperti di dominio in sistemi di informazione e di conoscenza, ma prima di approfondirne le potenzialità conviene valutare cosa implica adottare il paradigma della complessità nella produzione di un sistema informativo (Martelli, Bellini, 2012).

2.2. *La complessità: struttura dei problemi, del linguaggio e dei dati*

In un contesto complesso la struttura del problema è percepita generalmente come una rete di fattori interconnessi.

La connotazione a rete, comunque, non basta a descrivere un sistema complesso (Vega-Redundo, 2007) che è spesso caratterizzato da alcune tipicità che si dovrebbero ritrovare anche nei sistemi delle fonti che descrivono il dominio problematico (Martelli, 2011). Uno dei motivi per cui la rappresentazione della complessità non è adeguata a tale obiettivo sta proprio nel fatto che le fonti non riescono ad assumere tale struttura.

Queste caratteristiche sono:

a) Indipendenza da fattori di scala

Molti studi empirici focalizzati su ambiti diversi mostrano che le architetture di rete complesse sono in genere caratterizzate da pattern in cui molti nodi hanno poche connessioni, mentre altri ne presentano molte di più; le osservazioni empiriche mostrano che questo comportamento si riscontra a prescindere dalla dimensione della rete oggetto di analisi. In termini più rigorosi, le osservazioni empiriche mostrano che la connessione dei nodi è distribuita in accordo alla distribuzione di Pareto, secondo una distribuzione a invarianza di scala (Barabasi et al. 2000). La particolarità di questo tipo di distribuzione sta proprio nell'assenza di una scala caratteristica dei fenomeni. L'assenza di una scala definita è infatti uno dei fattori chiave che spiegano la complessità. Si può affermare che, volendo indurre un comportamento a rete, tutte le scale del fenomeno sono importanti e che nessun livello deve essere ignorato o trascurato. Detto in altri termini: un sistema che si comporta in modalità di rete complessa lo fa indipendentemente dalla scala in cui opera e che nessun contesto è troppo piccolo per non giustificare l'investimento di sforzi per indurre un comportamento a rete.

b) La distanza tra nodi: small world property

La teoria dei sei gradi di separazione formulata da Frigyes Karinthy nel 1929 è un'ipotesi secondo cui qualunque persona può essere collegata a qualunque altra attraverso una catena di conoscenze con un numero molto limitato di intermediari (Franceschetti, Meester 2006).

Questa intuizione sta reggendo alle verifiche più severe: ammettendo che la distanza tra due nodi di una rete possa essere misurata attraverso il numero di connessioni che occorre attivare per metterli in contatto, si è potuto verificarla anche nel caso di Internet. Infatti, nonostante i milioni di nodi che formano la rete WWW, è stato dimostrato che la loro distanza media è molto bassa: in media una coppia di pagine web prese casualmente sono distanti non più di 19 click (Barabasi et al. 2000), seguendo la catena di hyperlink che le connettono.

Con la diffusione dei social network la media è diventato più bassa, raggiungendo quattro (Backstrom, Ugander, 2012), quest'anno la ricerca del Facebook ha dichiarato che la media è diventata tre e mezzo (Lawrence, 2015).

Le reti come WWW che godono della proprietà della distanza breve tra i loro nodi è spesso detta godere della small-world property.

Riuscire ad indurre in un sistema un comportamento a rete vuol dire diminuire la distanza tra nodi. Tali distanze possono essere molto lunghe in sistemi con strutture non reticolari, come ad esempio quelli gerarchici. Le procedure gestionali ed amministrative possono esigere una connotazione organizzativa gerarchica; la Figura 1 rappresenta una situazione in cui si decide di collegare in modo diretto le informazioni generate dai nodi alla base di due distinte catene gerarchiche, decidendo di indurre la small world property, per generare un contesto informativo parallelo a quello di gestione.

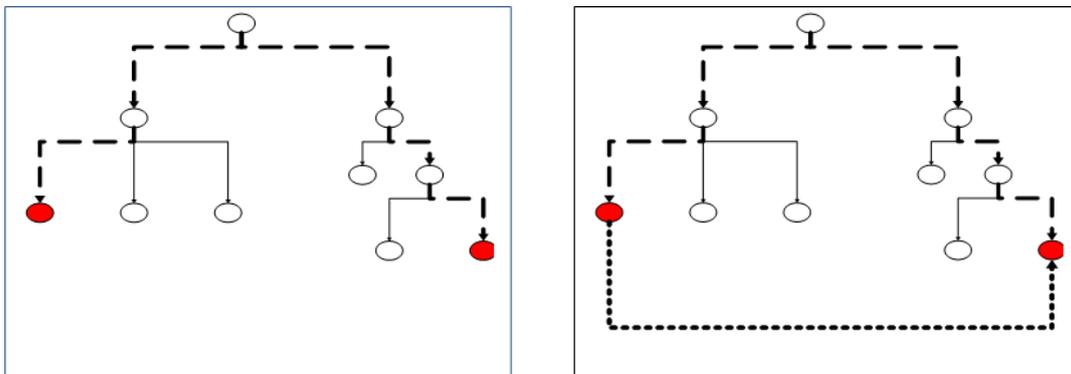


Figura 1 Struttura gestionale e organizzativa e struttura informativa supporto delle decisioni

2.3. Semantica memoria di sistema e comportamento a rete.

È assunto di questa tesi l'ipotesi che l'adozione di una politica di omogeneità semantica nei processi di generazione delle fonti amministrative sia la condizione indispensabile perché si diano le caratteristiche che consentano la strutturazione della memoria pubblica come sistema evolutivo complesso.

Si considerino gli schemi dell'esercizio rappresentato in Figura 2. Il primo grafo esemplifica una tipica struttura organizzativa con una gerarchia di gestione verticale: una delle due strutture potrebbe rappresentare un servizio e l'altra la sezione economica e finanziaria che si occupa del suo finanziamento e mantenimento. Tale modello è stato, per molti anni, l'unico riferimento organizzativo per aziende pubbliche e private ed istituzioni di governo.

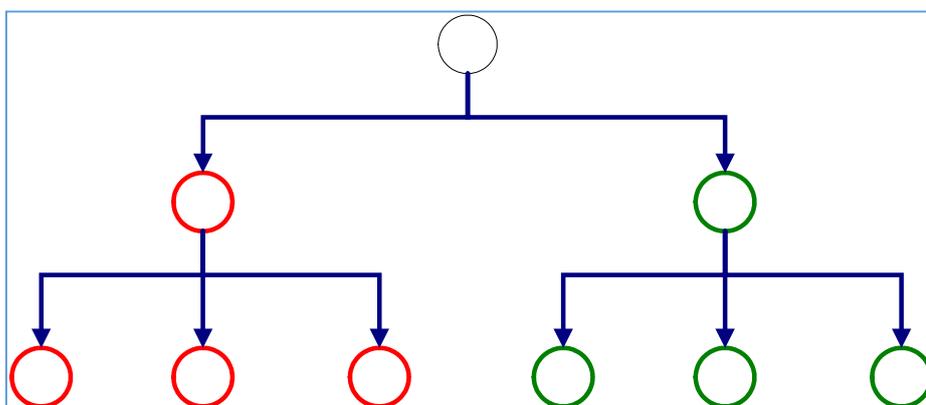


Figura 2 Struttura organizzativa verticale

Nell'organizzazione gerarchica ogni attore ha ben chiaro chi sia il proprio superiore. Normalmente, ogni elemento fa capo ad un soggetto personalmente responsabile della gestione della squadra e dei progetti su cui è impegnata. I progetti sono sviluppati in maniera sostanzialmente indipendente gli uni dagli altri. Ognuno di essi richiede un team assegnato che non è in genere coinvolto in altre attività esterne al progetto stesso. Questa verticalizzazione operativa e la relativa impermeabilità tra settori verticali è uno dei motivi che spiegano la nascita ed il consolidarsi di linguaggi diversi all'interno della struttura: anche se oggi è possibile definire modelli organizzativi non gerarchici che si discostano dalla visione tradizionale, una più o meno forte verticalizzazione organizzativa è pur sempre presente.

La Figura 3 seguente mostra una situazione in cui gli agenti situati nei diversi nodi hanno bisogno di riferirsi a informazione presenti in nodi situati su linee gestionali e decisionali parallele: questa è una situazione che tipicamente si genera in due eventualità, peraltro non alternative: (i) quando si decide di realizzare nuovi processi di servizio che possono godere della interoperabilità, tra i due sistemi; (ii) quando per poter decidere in maniera pienamente informata si producono sistemi informativi statistici che cercano informazioni sul contesto generale o in ambiti comunque influenzati dalle decisioni prese.

In questa tesi si esplorerà una via di sintesi: l'uso delle metodologie e degli strumenti semantici per governare la generazione di fonti amministrative, rispettose della struttura gerarchica di responsabilità ma in grado di integrarsi sia a livello gestionale che statistico.

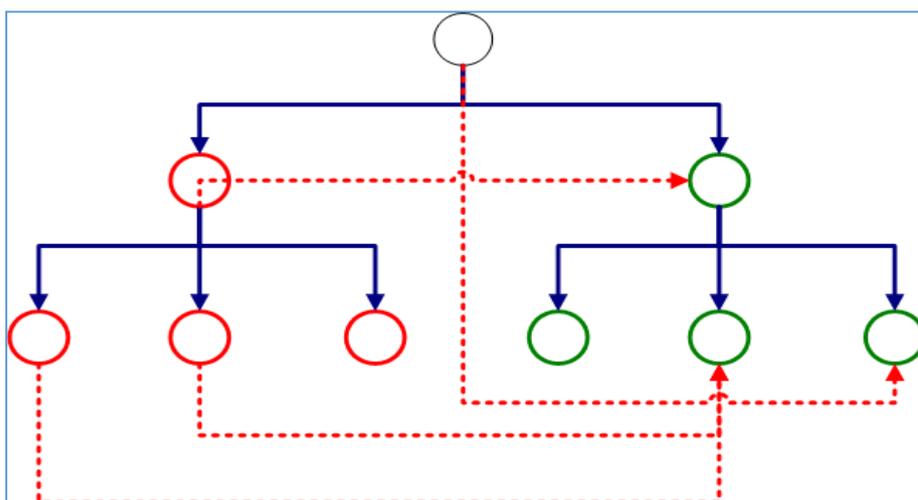


Figura 3 Della struttura di responsabilità gerarchica alla rete della informazione

Una simile evoluzione del sistema in una modalità collaborativa (Bititci et al. 2004) (Calzaroni, Contini 2004) (Rethemeyer, 2005) (Singh, 2005) è però possibile solo quando il linguaggio parlato dai diversi attori e, in generale dalla due strutture originarie è reciprocamente comprensibile.

2.4. Fare rete attraverso l'integrazione: il riuso statistico delle fonti amministrative.

Qualsiasi sistema che voglia acquisire capacità di tipo evolutivo ed adattativo, deve avere la capacità di leggere e valutare i propri comportamenti e deve poter influire sulle proprie modalità di funzionamento a partire da ciò che ha appreso su sé stesso. La crescente richiesta di informazioni statistiche nuove, più tempestive e a maggior dettaglio, ha spinto a promuovere l'utilizzazione sistematica delle fonti amministrative a fini statistici (Tronti, 2007) (Quintano, 2008) (Zavanella, 2004) (Martini, 2004) (Brancato, 2015), in ciò favorita dallo sviluppo

dell'informatica, che ha reso disponibile una grande quantità di informazioni, strutturate e facilmente utilizzabili. L'uso degli archivi amministrativi a fini di produzione statistica (Baldacci, 2014) non è un'operazione a costo zero, richiede che i dati contenuti in tali archivi siano confrontabili tra loro e con i dati "tradizionali" dei sistemi informativi statistici. È necessario, che l'omogeneità semantica sia estesa anche a concetti, definizioni e classificazioni il più possibile omogenei con quelli della statistica ufficiale.

Sono noti i vantaggi (Calzaroni, 2008) che l'uso degli archivi amministrativi in possesso della Pubblica amministrazione centrale e locale e di altri enti pubblici e privati consentono di ottenere nella produzione di statistiche. Vale forse la pena ricordare i più rilevanti:

- il significativo ampliamento dei contenuti informativi della produzione statistica corrente, attraverso la diffusione di nuove informazioni sui fenomeni e sulle popolazioni già oggetto di indagine o di dati relativi a fenomeni e realtà non ancora analizzati dal punto di vista statistico;
- una consistente riduzione dei costi di produzione degli istituti pubblici di statistica, nella misura in cui gli archivi amministrativi si pongano in alternativa alle indagini correnti;
- la riduzione del disturbo statistico, derivante dalla eliminazione o dal ridimensionamento delle indagini correnti in tutto o in parte sostituibili con archivi amministrativi e conseguente aumento della qualità delle informazioni acquisite, per effetto della riduzione dei rischi di saturazione e di rifiuto dei rispondenti;
- la possibilità di ottenere una copertura totale delle popolazioni di riferimento delle statistiche, che consente di ampliare in modo rilevante il dettaglio territoriale al quale vengono diffusi i dati (anche al di sotto del livello comunale).

3. Le fonti amministrative per l'analisi statistica

Nel capitolo precedente si è introdotto l'importanza e la criticità del ruolo delle fonti amministrative per costruire sistemi di fonti per descrivere contesti complessi.

In questo lavoro si affronterà il supporto che le tecnologie semantiche possono fornire alle fonti amministrative, ma prima di approfondire questo aspetto si rileggono alcune delle caratteristiche di tali fonti che le hanno rese indispensabili nella messa a punto dei sistemi informativi statistici.

3.1. *Definizioni e accezioni*

Le fonti amministrative in ambito statistico sono tradizionalmente definite come collezioni di dati detenute da specifiche istituzioni e amministrazioni, che vengono raccolte e trattate per finalità amministrative. In generale, vengono identificate quattro caratteristiche distintive di dati amministrativi:

- a) L'agente che fornisce i dati per l'agenzia statistica e l'unità a cui si riferiscono sono differenti (contrariamente alla maggior parte indagini statistiche);
- b) I dati sono stati originariamente raccolti per uno scopo non statistico. Tale caratteristica potrebbe pertanto influenzare il trattamento del gruppo di origine;
- c) L'obiettivo della raccolta di dati è quello della copertura completa della popolazione bersaglio;
- d) Il controllo dei metodi con cui i dati amministrativi vengono raccolti e trattati spetta l'agenzia amministrativa.

Questa definizione è sostanzialmente in linea con quello proposto dalla Iniziativa dati statistici e Metadata Exchange¹:

¹ Guarda <http://ec.europa.eu/eurostat/data/sdmx-data-metadata-exchange>

In generale, la definizione di fonte amministrativa è ampia e non dovrebbe porre limitazioni sulle statistiche; in questo lavoro ci si rifà alla seguente definizione (Vale, 2006):

“Administrative sources are data holdings containing information which is not primarily collected for statistical purposes.”

3.2. *I vantaggi di utilizzare fonti amministrative*

Molti sono i vantaggi (Wesley, Lavallée, Julie, 2012) che vengono dall'utilizzo delle fonti amministrative nelle statistiche ufficiali.

- Costo

Le rilevazioni tradizionali sono generalmente molto costose; anche i costi di set-up per l'utilizzo statistico delle fonti amministrative può essere alto ma i costi di gestione sono di solito significativamente più bassi.

- Onere di risposta

L'utilizzo dei dati di origine amministrativa contribuisce a ridurre l'onere di risposta per i fornitori di dati. Si tratta di una forte motivazione, soprattutto se gli intervistati sono imprese.

- Frequenza

Oltre alla riduzione dei costi e all'abbassamento dell'onere di risposta, un ulteriore vantaggio presentato dall'uso di fonti amministrative è dato dalla possibilità di mantenere le informazioni aggiornate, senza oneri aggiuntivi in termini di risposte, e con un contenuto costo aggiuntivo.

- Copertura

Le fonti amministrative spesso danno una copertura totale, o quasi totale della loro popolazione target. L'utilizzo di fonti amministrative può contribuire a eliminare reerrori di indagine, a rimuovere (o ridurre significativamente) mancate risposte, e fornisce stime più accurate e dettagliate per le varie sotto-popolazioni.

- Timeliness

L'utilizzo di fonti amministrative può aumentare la timeliness dei risultati statistici, consentendo l'accesso a più informazioni aggiornate riguardo a certe variabili. Questo perché indagini statistiche generalmente richiedono tempo per pianificare, progettare e forme pilota, per analizzare la popolazione e ottimizzare il campione.

- Immagine pubblica

L'opinione pubblica in materia di condivisione dei dati, in particolare tra i diversi dipartimenti governativi, varia notevolmente da paese a paese. Dove l'opinione pubblica in generale accetta, o è a favore della condivisione dei dati, un maggiore uso delle fonti di dati esistenti può contribuire ad accrescere il prestigio di un istituto nazionale di statistica, rendendolo più efficiente e conveniente.

3.3. *Criticità delle fonti amministrativi*

Accanto ai vantaggi, presentati nel paragrafo precedente, ci sono anche alcune criticità (The Center for Survey Statistics and Methodology, 2015).

- Opinione pubblica e questioni di privacy

Nel paragrafo precedente si era considerata l'ipotesi in cui l'opinione pubblica di un paese possa favorire la condivisione dei dati. In molte culture nazionali, tuttavia, esiste un forte sentimento di disagio dell'opinione pubblica al pensiero dei dati condivisi e della loro piena visibilità da parte del governo che potrebbe contrastare il riuso delle fonti amministrative a fini statistici.

- Profilo Pubblico

Il contatto diretto con il pubblico tramite sondaggi e indagini contribuisce a mantenere viva la presenza di una agenzia statistica. L'utilizzo dei dati amministrativi, riducendo i contatti può contribuire a

ridurre anche la consapevolezza pubblica del lavoro dell'organizzazione statistica.

- Gestione del cambiamento

Le fonti amministrative del settore pubblico sono generalmente una emanazione delle attività di governo e quindi sono suscettibili a cambiamenti dovuti a decisioni politiche e organizzative: possono cambiare in termini di copertura, definizioni, soglie ecc, I certi casi certe fonti possono essere addirittura abolite completamente. Anche le modifiche ai sistemi informatici utilizzati per archiviare ed elaborare dati amministrativi possono anche avere un impatto sulla fornitura di dati a fini statistici. Anche le fonti del settore privato non sono immuni da questo genere di cambiamenti, anche se in questo caso, il cambiamento è più probabile che sia guidato da cambiando fattore mercato.

- Unità

Uno dei principali problemi spesso incontrati quando si utilizzano fonti amministrative consiste nel fatto che le unità utilizzate in tali fonti possono non essere esattamente in linea con la definizione delle unità statistiche richieste.

- Definizioni di variabili

Così come possono esserci differenze nelle definizioni delle unità trattate, analogamente possono esserci anche differenze nelle definizioni delle variabili tra i sistemi amministrativi e quelli statistici. I dati di fonte amministrativa vengono raccolti per specifici scopi amministrativi, e le esigenze e le priorità relative a tale scopo sono suscettibili di essere diverse da quelle del sistema statistico.

- Sistemi di classificazione

Anche i sistemi di classificazione utilizzati all'interno dei fonti amministrative possono essere diversi da quelli utilizzati nel mondo statistico. Anche nel caso in cui sono uguali, possono tuttavia essere

applicati in modo diverso, in funzione dello scopo primario della fonte amministrativa, forse concentrandosi su attributi specifici dell'unità.

In altri casi, le classificazioni delle fonti amministrative potrebbero non essere applicate al livello di dettaglio richiesto dai fini statistici, o la classificazione potrebbero non essere una priorità per la fonte amministrativa, con conseguente scarsa qualità della fonte, quando riusata a fini statistici.

- **Timeliness**

Ci sono tre questioni particolari relative alla timeliness che influenzano l'utilità dei dati amministrativi a fini statistici:

I dati amministrativi potrebbero non essere disponibili in tempo per rispondere alle esigenze statistiche.

I dati amministrativi possono riguardare un periodo che non corrisponde con il periodo di riferimento statistico.

I dati amministrativi possono essere misurati in un periodo, mentre il requisito statistico è per un determinato tempo (o viceversa).

- **Incoerenza tra i fonti**

L'utilizzo delle fonti amministrative può presentare problemi di incoerenza: i dati provenienti da una fonte possono cioè apparire in contraddizione con quelli di altre. Ciò può essere dovuto a diverse definizioni o classificazioni e differenze nei tempi di rilevazione, o semplicemente per un errore in una fonte. Il problema della incoerenza può essere rilevato sia nel confronto di dati amministrativi con quelli statistici, che quando si confrontano due fonti amministrative.

3.4. Registri statistici come risultato della linkage tra fonti amministrative diverse

Nel paragrafo precedente si sono brevemente ricordati vantaggi e criticità delle fonti amministrative nell'ambito dei sistemi informativi

statistici: tali archivi trovano la più ampia applicazione, in ambito statistico nel contesto dei registri statistici.

Tipicamente un registro è una sorta di lista strutturata di unità, contenente una serie di attributi per ciascuna di quelle unità, e che è caratterizzato da un protocollo di aggiornamento regolare (United Nations, 2000). In questa prospettiva, dunque, molti archivi amministrativi possono essere considerati registri, ma non è una condizione scontata. Ad esempio gli archivi che contengano dati raccolti una sola volta non lo sono.

Secondo la definizione UNECE² (Vale, 2013) *“A statistical register is a register that is constructed and maintained for statistical purposes, according to statistical concepts and definitions, and under the control of statisticians. Administrative registers can therefore be used as sources for statistical registers, but the reverse would normally be seen as contradicting the principle of the “one-way flow” of data “.*

Un registro statistico può svolgere il ruolo di strumento di coordinamento dei dati, attraverso l'organizzazione e la normalizzazione dei processi di integrazione dei dati provenienti da fonti diverse, (statistici e/o amministrativi).

È chiaro dunque come la possibilità di effettuare operazioni di integrazione di fonti in maniera qualitativamente accettabile sia fondamentale per la costituzione dei registri statistici. Si noti come l'utilizzo di più fonti integrate consenta, quando effettuato con alti standard qualitativi di avere un miglior livello di precisione dei dati. Purtroppo il lato negativo di questo è che diventa necessario avere una strategia per gestire dati provenienti da fonti diverse. Tuttavia, se le variabili nei registri statistici vengono memorizzati con i dati e codici fonti, gli algoritmi automatici possono essere utilizzati per dare priorità ai fonti e risolvere i contrasti tra i dati.

² United Nations Economic Commission for Europe

3.5. *Fonti amministrative, interoperabilità e costituzione di sistemi informativi statistici a rete*

I registri statistici, intesi anche come frutto di attività di integrazione tra archivi diversi possono essere avvantaggiati dalle procedure di interoperabilità: intesa come scambio e utilizzo cooperativo delle informazioni (Fisher, 2006).

L'interoperabilità si realizza quando, nell'ambito di ambienti cooperativi evolutivi, i sistemi (o più) informativi devono inter-operare tra loro al fine di realizzare l'obiettivo determinato dal contesto cooperazione. Tipicamente, questa cooperazione richiede che si coordinino sistemi informativi eterogenei condividendo informazioni (Laurini, Murgante, 2008). L'interoperabilità si verifica quando ciascuno di questi sistemi informativi è in grado di utilizzare con successo le informazioni scambiate con gli altri, effettuando anche in molti casi operazioni per conto di un altro sistema (Whitman, Santanu, Whitman, Santanu, 2006).

In questa prospettiva, dunque, la realizzazione della cooperazione dipende fortemente dall'efficacia della interoperabilità tra i sistemi partecipanti.

Un elemento importante di interoperabilità è rappresentato dalla interoperabilità semantica che avviene quando (Yahia, Aubry, Panetto, 2012).

Il tema della interoperabilità è stato affrontato anche a livello europeo attraverso l'iniziativa The European System of interoperable Business Registers (Eurostat, 2013).

Il sistema statistico europeo ha preso passi verso la creazione di un vero e proprio sistema di registri statistici delle imprese e ha lanciato il progetto SBR³ con i seguenti obiettivi:

- Rafforzare e razionalizzare SBR nazionali
- Integrazione in un sistema interoperabile;

³ Statistical business register

- Servire la produzione statistica nazionale ed europea;
- Supportare efficienza e qualità utili per l'intero sistema.
- Migliorare il Group Register europeo (cuore del sistema)
- Supportare accesso on-line degli istituti nazionali statistici, facilitando l'interazione con EGR⁴
- Integrazione delle informazioni di profiling in EGR
- NSIs⁵ accesso servizi condivisi per la gestione del registro
- servizi certificati da Eurostat sulla base degli standard ESS⁶

3.6. *Ruolo e importanza dei metadati*

Nella prospettiva della piena conoscenza ed integrabilità delle fonti, un ruolo importante è rivestito dai metadati.

I metadati (ISO/IEC, 2013) sono i dati che definiscono e descrivono altri dati e sono importanti per informare produttori e utilizzatori circa la qualità dei dati. I dati gestiti da un sistema informativo dovrebbero essere sempre accompagnati dai metadati sufficienti per comprenderli appieno, e per garantire che i valori siano assegnati correttamente alle variabili rilevanti. Anche la documentazione dettagliata su concetti, definizioni e finalità della fonte, nonché sulla raccolta e al trattamento dei metodi utilizzati, è importante: darà una migliore comprensione dei potenziali problemi di qualità, e dovrebbe costituire la base per le regole di modifica dei dati in fase di lavorazione.

Durante l'elaborazione dei dati è inoltre importante registrare anche tutte le informazioni relative al trattamento. Questo non solo fornisce informazioni vitali per le valutazioni di qualità della lavorazione, ma fornisce anche un meccanismo per indagare eventuali problemi nel processo e annullare eventuali errori.

⁴ EuroGroups Register

⁵ the national statistical institutes

⁶ The European Statistical System

Il concetto di metadato è particolarmente importante in ambito statistico (Rinaldi, 2002): le misure infatti devono essere accompagnati da metadati sufficienti per consentire agli utenti di recuperarli, interpretarli correttamente e formarsi un'opinione sulla loro qualità. La comunicazione di qualità può essere spesso difficile da ottenere, in quanto alcuni utenti vogliono tutti i dettagli, mentre ad altri sono sufficienti indicatori sintetici di alto livello. Un modello di metadati che consente agli utenti di vedere i diversi livelli di informazioni, a partire da una sintesi, ma con la possibilità di vedere più in dettaglio, è forse la più appropriata.

3.7. *Qualità delle fonti amministrative per l'utilizzo statistico*

Il punto di partenza per un simile quadro è la definizione di qualità. Ancora una volta, molto lavoro è stato fatto in questo settore da organismi statistici nazionali ed internazionali, la maggior parte delle quali si basa sullo standard internazionale ISO 9000/2005⁷, che definisce la qualità come:

"Il grado in cui un insieme di caratteristiche intrinseche soddisfa i requisiti".

a) Qualità del framework

La qualità delle fonti amministrative è un elemento di grande importanza e criticità: la qualità per l'uso statistico è diversa da quella progettata per l'utilizzo amministrativo. Per questo motivo non si può utilizzare, anche se disponibile, il manuale e la documentazione di qualità predisposta per la fonte amministrativa.

Il tema della qualità è oggetto di importanti approfondimenti da parte delle agenzie nazionali statistiche (Fazio, 2013). In questo lavoro il tema della qualità della fonte viene affrontato dal punto di vista della qualità del framework, ossia dei meccanismi attraverso i quali i dati vengono generati e trasferiti.

⁷ Guarda http://www.iso.org/iso/catalogue_detail?csnumber=42180

La qualità del framework per dati di fonti amministrativi è composto da diverse viste relative al concetto di hyperdimensions (Karr, Sanil, Banks, 2006). Gli aspetti di qualità in ogni hyperdimension influenzano l'usabilità di una sorgente di dati in un modo diverso. Ci sono tre hyperdimensions, fonte, metadati e dati: vengono utilizzati per determinare l'utilizzabilità statistica di una sorgente di dati amministrativi (Piet, Daas, Schouten, Kuivenhoven, 2008). Ogni hyperdimension è composta di varie dimensioni; ogni dimensione contiene una serie di indicatori di qualità. Un indicatore di qualità sono misurate o stimate da uno o più metodi sia qualitativi o quantitativi.

Le hyperdimensions sono ordinate anche in base a un crescente livello di dettaglio. Gli indicatori di qualità nel hyperdimension dati, per esempio, informano sugli aspetti di qualità ad un livello molto più dettagliato rispetto agli indicatori di qualità inclusi nella hyperdimension metadati. Lo stesso vale per i metadati e la fonte hyperdimensions. Un risultato importante di questa distinzione ordinato è il fatto che essa guida efficacemente l'utente nella studio della qualità di una fonte di dati.

b) Le fonte

Nella hyperdimension relativa alla fonte sono studiati (Piet, Daas, Schouten, Kuivenhoven, 2008) gli aspetti qualitativi legati alla fonte di dati. Il hyperdimension relativa alla fonte è composto da cinque dimensioni della qualità; queste sono: Fornitore, rilevanza, privacy e sicurezza, consegna, e procedure.

c) Metadata

Il hyperdimension metadati si concentra in particolare sugli aspetti relativi ai metadati della fonte dei dati. Chiarezza delle definizioni e la completezza delle meta informazioni sono alcuni degli aspetti qualitativi inclusi. Il hyperdimension metadati si compone di quattro dimensioni: chiarezza, comparabilità, chiavi uniche, e il trattamento dei dati. La dimensione trattamento dei dati è un caso speciale (Brancato, 2015). Consiste di indicatori di qualità utilizzati per determinare se il titolare della fonte dei dati esegue controlli su e/o modifica i dati nell'origine. Questa

meta informazione è molto importante per un Istituto Nazionale di statistica come certo influisce sulla qualità del prodotto fornito dal custode dei dati.

d) Data

Il hyperdimension data si concentra sugli aspetti di qualità dei dati nella fonte dati. Sebbene la maggior parte dei risultati descritti in questo paragrafo attenzione agli aspetti di qualità incluso nel hyperdimension Fonte e metadati, i dati hyperdimension è discusso qui per completezza.

Gli aspetti di qualità del hyperdimension data sono prevalentemente accuratezza legati ad eccezione di quelli inclusi nella dimensione tecnica controlli (Cerroni, Bella, Galiè, 2014). Questa dimensione contiene indicatori che verificano la leggibilità del file di dati e la conformità dei dati per la definizione dei metadati. Gli altri sono: sovracopertura, sottocopertura, linkabilità, non risposta, processo di misurazione precisione e sensibilità. La dimensione della sensibilità è principalmente utilizzato per determinare l'effetto sulla qualità dei dati delle variazioni dipendenti dal tempo nella composizione della popolazione.

4. Nuove metodologie per la modellazione e l'integrazione di fonti e semantiche eterogenee: le ontologie

Nei capitoli precedenti si è riflettuto sul fatto che l'analisi e la misura statistica dei sistemi complessi esige un sistema di memoria che sia in grado di rappresentare a sua volta un sistema a rete evolutivo e complesso: nel corso della riflessione si è sottolineato il fatto che gli archivi amministrativi sono in grado, quando sia possibile integrarli con un adeguato framework di qualità, di configurarsi come il sistema a rete cercato.

Si sono anche visti gli elementi di criticità che sottostanno a questo processo. Si vuole adesso valutare quale sia il ruolo delle nuove tecnologie semantiche per ottenere il massimo della integrabilità delle fonti amministrative in un sistema informativo statistico.

Prima di affrontare il tema dell'utilizzo di tali tecnologie nella prospettiva dei sistemi informativi statistici, si richiamano brevemente i più importanti elementi caratterizzanti per rileggerne l'uso nella prospettiva statistica.

4.1. Il semantic web: una evoluzione del web, una prospettiva per la statistica.

Il semantic web (Bernes Lee, 2001) è la più recente e importante evoluzione del web, ed è caratterizzato dall'utilizzo di metadati per aggiungere o estrarre il significato dei contenuti del web, con l'obiettivo di renderli comprensibili anche ad agenti automatici: si tratta di una evoluzione dell'attuale web, che nella sua versione originale (compresa anche la sua versione social, web2.0) è costituito da una rete di documenti collegati da hyperlinks.

A tutti gli effetti il web attuale è un web dei documenti pensati essenzialmente per un operatore umano che cerca, trova e aggrega l'informazione.

Il semantic web, invece, è una prospettiva che intende presentare i dati in modo che siano comprensibili da un agente automatico, mettendolo in condizioni di effettuare ricerche, aggregando e combinando le informazioni in autonomia: l'obiettivo di lavoro è quella di un collegamento tra dati presenti in rete in modo che sia possibile collegarli e navigarci.

4.2. *Le ontologie*

Si vuole, in altri termini, strutturare sistemi che rappresentino la conoscenza di un dato dominio e che ne consentano la fruizione anche attraverso strumenti automatici: questa visione è sintetizzata nel termine ontologia, che in informatica si riferisce al tentativo di formulare un quadro concettuale rigoroso di un dato dominio (Chandrasekaran, Josephson, Benjamins, 1999).

In un'ontologia il sistema di relazioni tra le entità viene definito e precisato: quando un oggetto è definito (in questo caso una risorsa web) come istanza dell'ontologia, vengono descritti i rapporti, le caratteristiche e le proprietà della classe in cui è collocato (Masolo, Oltramaria, Gangemia, Guarino, Vieu, 2003).

Da un punto di vista formale, l'ontologia può essere rappresentata da un grafo che contiene tutti i concetti relativi al dominio di interesse e le relazioni tra questi concetti (Basti, 2014). Le relazioni tra concetti sono espresse attraverso una struttura gerarchica: le classi che rappresentano i concetti di livello superiore si scompongono in sottoclassi più dettagliate che ereditano le caratteristiche dei concetti di livello superiore.

In aggiunta a questa strutturazione esiste anche una relazione tra classi espresso con il termine "proprietà". Le proprietà sono usate per descrivere le caratteristiche (o attributi) di concetti, e possono anche essere utilizzate per collegare diverse classi, superando il modello strettamente gerarchico appena presentato: le relazioni tra le classi non sono dunque solo di tipo di classe sottoclasse, ma anche espresse in termini di proprietà. Aver definito i concetti e le relazioni tra loro, è il primo passo per codificare la

conoscenza del dominio di interesse in modo tale da poter essere interpretato automaticamente.

I vantaggi dell'ontologia

L'utilizzo dell'ontologia porterà questi vantaggi (Basti, 2014):

- Permette di stabilire una serie di concetti chiave e le loro definizioni (relativi a un dominio specifico) che possono essere condivisi, fornendo la terminologia corretta per chi vuole creare i propri documenti (condivisione della conoscenza).
- Permette il riutilizzo della conoscenza codificata in altre ontologie o per il loro completamento (nessuna ripetizione di informazioni).
- Interpretato dal computer, consente il trattamento automatico della conoscenza e apre nuove prospettive alla valorizzazione della informazione su web.

È questo il motivo per cui il semantic web e la concettualizzazione in termini ontologici è una linea di ricerca importante per la costruzione di sistemi informativi statistici perché si configura come una modalità alternativa per pubblicare set di dati standardizzati, documentati, collegati ed elaborabili.

4.3. Il web nella prospettiva di un database integrato: il modello a grafo

Affinchè il web assuma una dimensione come quella appena prefigurata, come prima cosa occorre riflettere a quale modello di database ci stiamo riferendo. La tabella seguente sintetizza i principali approcci di modellazione dei dati: prima di esplorare i benefici del modello RDF può essere utile riconsiderare le principali caratteristiche di questi filoni di

modellazione: la tabella seguente⁸ mostra l'unicità di alcune caratteristiche del modello semantico.

comparing the features of the mainstream ways of modeling data versus the semantic web model						
model	example format	Data	metadata	identifier	query syntax	semantics (meaning)
object serialization	.net clr object serialization	object property values	object property names	e.g. filename	linq	n/a
relational	ms SQL, oracle, mySQL	table cell values	table column definitions	primary key (data column) value	SQL	n/a
hierarchical	xml	tag/attribute values	xsd/dtd	unique attribute key value	xpath	n/a
Graph	rdf/xml, turtle	Rdf	rdfs/owl	Uri	Sparql	yes, using rdfs and owl

Nei sistemi informativi statistici generalmente si fa riferimenti a modelli di archiviazione gerarchici o relazionali. Nel semantic web, come evidenziato anche dalla tabella, il paradigma è quello dei graph database.

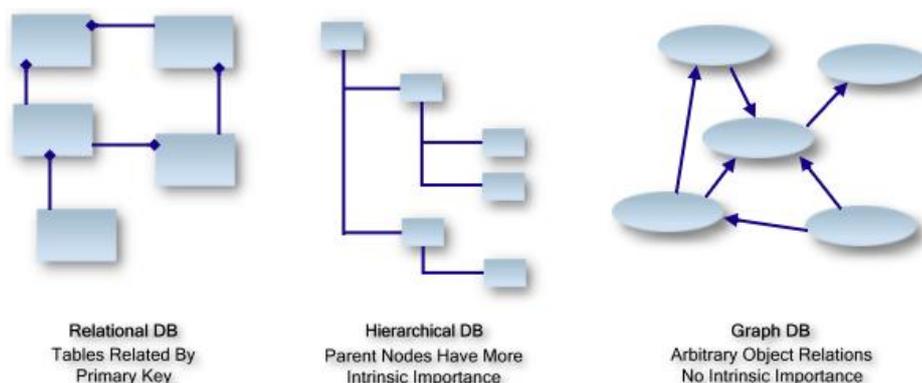


Figura 4 modelli di archiviazione

⁸ Guarda <http://www.linkeddatatools.com/introducing-rdf>

Una base di dati a grafo, o database a grafo, (Webber, Eifrem, Ian Robinson, 2015) usa nodi e archi per rappresentare e archiviare l'informazione. La rappresentazione dei dati mediante grafi offre un'alternativa al modello relazionale che fa uso di tabelle, ai database orientati al documento (che usano documenti), o altri, come i sistemi ad archivi strutturati (structured storage) basati su colonne o su cesti non interpretati di dati.

Attualmente, i modelli di riferimento per l'implementazione dei database a grafo sono due: il property graph model e il resource description framework graph (RDF). Nell'ambito di questa tesi si farà riferimento al secondo che è, appunto, quello utilizzato nel web semantico. I database a grafo che utilizzano il modello RDF sono anche noti come Triple Store, Quad Store, o RDF Store.

Per entrambe le tipologie di database a grafo (Vicknair et al., 2010) esistono dei linguaggi di interrogazione specifici, ma solo per RDF esiste uno standard riconosciuto in SPARQL.

Per meglio riflettere sulle potenzialità del web semantico nella costituzione di sistemi informativi statistici, si propone una breve presentazione dei suoi elementi fondamentali. Si tratta di aspetti molto noti in ambito informatico che non sono stati peraltro ancora approfonditamente riflettuti nel contesto statistico ufficiale.

4.4. Dal graph database al sistema informativo semantico: la tripla RDF

Secondo la sua definizione più accreditata, RDF è un modello di dati standard orientato al web data exchange. RDF (Raimond & Raimond, 2014) ha caratteristiche che facilitano le operazioni di merge, anche tra schemi differenti e supporta, in modo specifico, l'evoluzione temporale di schemi, senza la necessità di cambiare le profilature di utenza.

RDF appartiene alla famiglia dei linguaggi XML e ne condivide lo scopo fondamentale: generare informazioni che non siano soltanto destinati alla lettura, ma che possano essere riutilizzati per applicazioni automatiche;

RDF rappresenta, peraltro una profonda evoluzione rispetto a XML, perché non si limita a consentire l'interrogazione di documenti: nelle sue più avanzate applicazioni ne rappresenta il significato (Berners-Lee, 1998).

Qualunque cosa descritta da RDF è detta **risorsa**. Tipicamente una risorsa è reperibile sul web, ma RDF può descrivere anche risorse che non si trovano direttamente sul web. Ogni risorsa è identificata da un URI, Universal Resource Identifier. Le potenzialità di un approccio per URI nel campo della standard di concetti statistici è evidentemente molto promettenti.

Il modello di dati RDF è formato da risorse, proprietà e valori. Le proprietà sono delle relazioni che legano tra loro risorse e valori, e sono anch'esse identificate da URI. Un valore, invece, è un tipo di dato primitivo, che può essere una stringa contenente l'URI di una risorsa.

L'unità base per rappresentare un'informazione in RDF è lo statement. Uno statement (Klyne, Carroll, 2004) è una tripla del tipo Soggetto – Predicato – Oggetto, dove il soggetto è una risorsa, il predicato è una proprietà e l'oggetto è un valore o una risorsa.

4.4.1. Come funziona il RDF

Una risorsa di Internet è definita come qualsiasi risorsa con un Uniform Resource Identifier (URI). Ciò include la Uniform Resource Locator (URL) che identifica interi siti web così come pagine web specifiche (Hamon, 2013). La descrizione dichiarazioni di RDF, racchiusa all'interno di una sezione di Extensible Markup Language (XML), potrebbe essere inserita all'interno di una pagina Web (cioè, un linguaggio Hypertext Markup - HTML - file) o potrebbe essere in separate file.

RDF è una raccomandazione formale del W3C (Cyganiak, Wood, Lanthaler, 2014). Attualmente, secondo una raccomandazione del W3C che è ancora in fase di "Proposta", si ipotizza un sistema in cui le descrizioni relative ad un particolare scopo (ad esempio, tutte le descrizioni relative alla sicurezza e alla privacy) costituirebbero una classe

Tali classi potrebbero adattarsi a uno schema o una gerarchia di classi e farebbero ereditare alle loro sottotoclassi le descrizioni relative al loro specifico scopo

Nell'ambito dei sistemi informativi statistici, una simile modalità consentirebbe di specificare in modo parsimonioso vincoli particolari come quello riferibile alle modalità di rispetto del segreto statistico

Il data model RDF permette di definire un modello semplice per descrivere le relazioni tra le risorse, in termini di proprietà identificate da un nome e relativi valori. Tuttavia, RDF data model non fornisce nessun meccanismo per dichiarare queste proprietà, né per definire le relazioni tra queste proprietà ed altre risorse. Tale compito è definito da RDF Schema.

4.4.2. RDF Schema

RDF rappresenta relazioni tra risorse: poiché si vuole rappresentare anche il significato di queste risorse, si è utilizzato RDF per produrre un contesto, RDF Schema, che è orientato alla gestione dei metadati.

In RDF Schema (RDFS) (Brickley, Guha, 2014) ogni predicato è in relazione con altri predicati e permette di dichiarare l'esistenza di proprietà di un concetto, che permettano di esprimere con metodo sistematico affermazioni simili su risorse simili. RDF Schema permette di definire nuovi tipi di classe. Inoltre specificando il concetto di classe e sottoclasse, consente di definire gerarchie di classi. In RDF si possono rappresentare le risorse come istanze di classi e definire sottoclassi e tipi.

Si capisce dunque come il contesto RDF sia perfettamente adeguato alla descrizione di sistemi di metadati che sono essenziali anche per la predisposizione di sistemi informativi statistici.

4.5. *Web Ontology Language (OWL)*

Per ottenere la rappresentazione della conoscenza descritta nei paragrafi precedenti occorre anche poter gestire le classi concettuali e le loro relazioni: per questo obiettivo, lo strumento da utilizzare è OWL.

OWL (McGuinness, Harmelen, 2004) è un linguaggio proposto nel 2004 per descrivere ontologie; nel 2007, è stato istituito un nuovo gruppo di lavoro del OWL nel W3C per lo scopo di rivedere la proposta del 2004 con una nuova versione chiamata OWL2 (Motik, Parsia, 2012). OWL2 ha costituito, a ottobre 2009, una nuova raccomandazione del W3C.

Di seguito, con il termine OWL si farà riferimento a OWL2.

Questo linguaggio è disponibile in tre diverse forme di espressione e di complessità crescente:

- I. OWL Lite (Kriglstein, Wallner, 2010) è il più semplice fra le tre lingue ed è quello dal più basso potere espressivo. Può essere utilizzato in modo efficace soprattutto quando abbiamo bisogno di definire una classificazione gerarchica e vincoli semplici. E' facile da usare quando abbiamo bisogno di fare un rapido passaggio da un thesaurus ad un altro sistema di organizzazione della conoscenza.
- II. OWL DL (Description Logic) (Matentzoglou, Bail, Parsia, 2013) è un linguaggio più espressivo di Lite ed è decidibile ed ha procedure di deduzione con complessità nota.
- III. OWL Full (McGuinness, Harmelen, 2004) permette la massima espressività ma non è decidibile .

4.6. *Utilizzare i sistemi semantici: SPARQL Protocol and RDF Query Language (SPARQL) per la costruzione degli statistical data sets*

Dopo aver riflettuto sugli strumenti per la descrizione e la strutturazione della conoscenza, ci si sofferma adesso sui linguaggi per interrogare tali sistemi e per estrarre i set su cui poi applicare i sistemi di misura e di valutazione degli indicatori statistici.

SPARQL è un linguaggio di query progettato specificamente per interrogare gli RDF store (Grobe, 2009). Le queries SPARQL sono inviate da un client a un server conosciuto come *SPARQL-endpoint* utilizzando il protocollo HTTP. L'interazione tra il client e *SPARQL-endpoint* è definito da un protocollo adatto alla interpretazione automatica e non è destinato ad essere interpretato da esseri umani. Per questo motivo l'uso di

SPARQL richiede un'interfaccia che consente all'utente di inserire le query e visualizzare i risultati in modo comprensibile.

4.6.1. Protégé editor

Protégé (Horridge, 2011) è una piattaforma open source gratuita che fornisce una suite di strumenti per la costruzione di modelli di un dominio e le applicazioni basate sulla conoscenza con ontologie. Esso supporta la creazione, la visualizzazione e la manipolazione di ontologie in diversi formati di rappresentazione. Protégé può essere personalizzato per fornire il supporto del dominio per la creazione dei modelli di conoscenza e l'inserimento dei dati. Inoltre, Protégé può essere esteso per mezzo di un'architettura plug-in e una Application Programming basato su Java Interface (API) che serve a creare strumenti e applicazioni basati sulla conoscenza. La piattaforma Protégé supporta due modi principali di ontologie di modellazione: l'editor Protégé-frame e l'editor Protégé OWL. L'editor Protégé OWL permette agli utenti di costruire ontologie per il semantico web, in particolare del W3C Web Ontology Language (OWL).

Nelle sue più recenti evoluzioni Protégé sta integrandosi in modo sempre più stretto con i packages statistici, in particolare R. Tali applicazioni sono particolarmente orientate a funzioni di temporal data mining (Tusch, Huang, O'Connor, Das, 2009) (Tusch, O'Connor, Redmond, Shankar, Das, 2007) o alla costruzione di reti bayesiane basate su ontologie già esistenti⁹.

4.7. *Integrare le ontologie*

Il tema della costituzione di un sistema di conoscenza passa, come si è visto, attraverso la fattibilità dei processi di integrazione. Si porta adesso questa riflessione al suo più alto livello di generalità: quello della integrazione tra ontologie.

La costruzione di sistemi di conoscenza è infatti grandemente avvantaggiata dalla possibilità di integrare ontologie diverse: nella pratica

⁹ Guarda [http://protegewiki.stanford.edu/wiki/Bayesian_Network_Tab_\(BNTab\)](http://protegewiki.stanford.edu/wiki/Bayesian_Network_Tab_(BNTab))

statistica questa è una esigenza normale. Si pensi, nuovamente a titolo di esempio, agli studi di epidemiologia occupazionale (nel cui campo si inquadra anche il sistema informativo statistico presentato nel caso di studio) nei quali si integrano i domini del lavoro, quello della sanità e della prevenzione.

Nel processo di integrazione le ontologie di partenza sono aggregate insieme in modo da formare una nuova ontologia. Il concetto di integrazione tra ontologie (Pinto, Martins, 2001) è legato all'idea di allineamento. L'ontologia risultante dal processo di integrazione è costituita da regioni di conoscenza estratte dalle ontologie originarie. Tipicamente la conoscenza di queste regioni rimane sostanzialmente invariata.

Anche se il termine "integrazione" è utilizzato in molti settori, la sua definizione non è ancora consensuale.

Nell'ambito di questo lavoro si intenderà (Namyoun Choi, Song, Han, 2006) come l'atto di unire o integrare due ontologie concettualmente non coincidenti o i dati delle istanze di due ontologie. Questo processo di integrazione può essere eseguito in vari modi, manualmente, in modo semiautomatico o completamente automatico. Le tecniche di integrazione possono essere guidate anche da approcci statistici, tenendo conto della somiglianza dei concetti e delle istanze o attraverso la corrispondenza esatta tra i nomi degli oggetti o la corrispondenza dei loro Uniform Resource Identifier.

In questo paragrafo discutiamo l'integrazione tra le ontologie da due punti di vista; la prima: l'integrazione tra le strutture e i dati delle ontologie e la seconda: l'integrazione tra i *triple store* delle ontologie.

Si tratta di un tema molto importante, dal punto di vista dei sistemi informativi statistici, perché fa intravedere nuove potenzialità alle operazioni di linkage tra fonti.
--

4.7.1.L'integrazione tra le strutture e i dati delle ontologie

Il processo di integrazione inizia con l'analisi delle due ontologie da integrare: devono essere studiate nel dettaglio per individuare e studiare i punti concettuali di unione. Il programma "Protégé" permette di implementare questa integrazione in modo automatico (Ameen, Rani, Khan, 2014) attraverso l'opzione "Merge ontologies" nella lista "Refactor".

4.7.2.L'integrazione tra i triple store delle ontologie

Un *triple store* è un framework che, attraverso l'adozione di un linguaggio di query, consente l'archiviazione e il recupero di triple (dati RDF¹) (Allemang, Hendler, 2011): è dotato di strumenti che consentono la memorizzazione e l'accesso ai grafi RDF.

Tramite i *triple store* le triple possono essere importate e esportate utilizzando RDF o altri formati.

Un *triple store* comprende come una funzionalità fondamentale, la possibilità di unire insieme due set di dati. Tale unione è un insieme di dati che include le triple dai data set di origine. Tutte le risorse con lo stesso URI (indipendentemente dalla sorgente di dati di origine) sono considerate equivalenti nel set di dati generato dall'unione.

L'efficienza del processo di unione può dunque configurarsi come un requisito di qualità aggiuntivo per valutare una fonte.

4.8. Ruolo delle ontologie per la costituzione di archivi amministrativi interoperabili e giacimenti di dati riusabili per la costituzione di registri statistici

Nei paragrafi precedenti si sono discussi i vari metodi per condividere e unire domini di conoscenza: quando si tratta di sistemi informativi statistici la semplice condivisione della conoscenza può non bastare perchè i requisiti di qualità sono particolarmente stringenti e severi.

Gia si erano discussi (§3.6) gli elementi costitutivi dei principali *frameworks* di indicatori di qualità per fonti amministrative. Si tratta di

sets di indicatori che descrivono e documentano la qualità complessiva delle fonti di dati analizzati, e generalmente non entrano nel merito dei processi di costruzione di una fonte di qualità accettabile per il riuso statistico.

Tale obiettivo può essere raggiunto attraverso l'adozione di ontologie, che rivestono un ruolo importante sia nel processo di generazione delle singole fonti amministrative che in quello di sintesi dei registri statistici.

4.9. Ontologie e condivisione della conoscenza

Gli aspetti del semantic web, tratteggiati nei paragrafi precedenti, mostrano quali importanti avanzamenti nel processo di condivisione della conoscenza sia possibile ottenere: l'obiettivo iniziale della costruzione di sistemi informativi statistici adeguati alla rappresentazione e alla misura di contesti complessi trovano, in questo approccio un importante alleato.

Si era riflettuto (§3.2) sulle potenzialità del riuso dei dati amministrativi in ambito statistico e di come la modalità totalmente indipendente dei processi di generazione di tali archivi fosse uno dei limiti più severi rispetto all'obiettivo della loro integrazione in un sistema unificato e armonizzato. Si vede dunque come le tecnologie semantiche, operando a supporto della condivisione della conoscenza, siano particolarmente importanti in un contesto in cui si fa sempre più affidamento sulle fonti amministrative ma non si può contare fino in fondo sulla capacità di controllarne il processo di costruzione.

Si considerino, ad esempio, alcuni archivi originati in modo indipendente, come tipicamente avviene nel caso del riuso di dati amministrativi per analisi sociali, sanitarie, economiche, di popolazione. Il link tra ambienti informativi diversi è un tipico caso di sistema informativo statistico (come quello pensato per le analisi di epidemiologia del lavoro, oggetto del caso di studio presentato in questo elaborato).

Tipicamente quando gli ambienti informativi sono nati in modo totalmente autonomo:

- Ogni integrazione di dati tra sistemi diversi non può, evidentemente essere fatta sulla base del semplice join tra i database. Innanzitutto sono stati disegnati in modo indipendente, dovrebbero essere mappati e, sul piano più propriamente tecnico i loro sistemi potrebbero non essere compatibili;
- Per collaborare, dunque, dovrebbe essere deciso un data format condiviso, ad esempio creando un endpoint xml su ognuno dei loro siti web in modo che ognuno dei due possa richiedere informazione all'altro;
- È importante notare che questo approccio allo scambio di informazione tra sistemi di dati indipendenti e potenzialmente non perfettamente compatibili richiede investimento di tempo, denaro e di una interpretazione umana contestuale dei due diversi *datasets*;
- È inoltre limitato al *data domain* dei sistemi e ogni ulteriore aggiunta al loro sistema di conoscenza da fonti terze richiederebbe uno sforzo analogo: in altri termini, occorre che degli umani capiscano il significato dei dati e che si mettano d'accordo su formati comuni per farli collaborare adeguatamente i due data sets.

Con l'introduzione di RDF e della semantica questo processo si semplifica: le operazioni prima accennate diventano automatiche e non manuali (Buccella, Cechich, Brisaboa, 2004). I sistemi adottano, infatti, un comune vocabolario standard per descrivere i propri dati: ad esempio "mansione del lavoratore" deve avere lo stesso significato in entrambi i sistemi: questo può essere fatto adottando la stessa ontologia di base per esprimere il significato che sta dietro ai dati pubblicati. Tornando, per esemplificare, al tema della epidemiologia del lavoro trattato nel caso di studio:

- I sistemi possono interrogarsi l'un l'altro utilizzando gli stessi termini;
- Il sistema che descrive le caratteristiche delle opere e dei cantieri può avere i dettagli sanitari sugli infortuni occorsi durante i propri lavori;

- Il sistema che descrive le biografie sanitarie può acquisire più informazioni sulle condizioni di lavoro dei cantieri in cui il lavoratore ha operato;
- Grazie alle relazioni contestuali definite in una ontologia web formalizzata, ulteriori informazioni possono essere individuate attraverso una linked standard terminology senza che l'utente finale debba necessariamente essere a conoscenza della disponibilità di questa informazione;
- Tutto questo avviene senza il bisogno di trasformare o mappare;

Allo stato attuale esistono già numerose ontologie standard disponibili sul web ma manca, come vedremo nel corso di questo lavoro, una riflessione specifica, dal punto statistico, delle ontologie che consentano un merge di archivi amministrativi e gestionali orientati alla costituzione di registri statistici.

La condivisione della conoscenza avviene, dunque, attraverso alcuni strumenti fondamentali (Pieroni, Franchini, Mariani, Fortunato, Molinaro, 2013) (Scorza, 2009) per i quali si può auspicare una presenza della agenzie statistiche nazionali nella formulazione e certificazione di:

- vocabolari, intesi come una collezione di termini corredati di definizioni consistenti in specifici domini;
- le ontologie, che permettono di definire relazioni contestuali a sostegno di specifici vocabolari.

I diversi archivi amministrativi possono adottare una stessa ontologia di base, o un vocabolario comune, per esprimere il significato dei propri dati e pubblicandoli su un *queryable endpoint* in modo che possano comunicare tra loro attraverso la rete. In altre parole, i diversi sistemi si potranno interrogare (Costamagna, Spanò, 2012) a vicenda utilizzando gli stessi termini.

La condivisione della conoscenza secondo i principi precedentemente presentati si applicano ai siti, alle basi di conoscenza ed alle organizzazioni.

Alcuni esempio di vocabolari standard e di ontologie formali (media terms, terminologia medica o scientifica) sono:

- Dublin Core Metadata Initiative (DCMI)¹⁰ - Creates ontologies for a range of subjects, particularly focusing on common, every day terms and terms important in media.
- Friend Of A Friend (FOAF)¹¹ - focuses on developing a standard vocabulary/ontology for social networking purposes (cfr.§5.3.1).
- OpenCyc¹² - An ontology of everyday, common sense terms.

4.10. *Ontologie e database*

Se le ontologie possono supportare nella integrazione di fonti amministrative per l'uso statistico, occorrerà approfondire il nesso tra le ontologie e i database. È un aspetto molto importante per le seguenti ragioni:

- La maggior parte delle fonti amministrative che si intendono integrare in un sistema statistico sono originariamente contenute in databases;
- Si vuole esplorare il ruolo di ontologie ufficiali per generare sistemi informativi capaci di generare dati buoni non solo per il servizio ma anche per il riuso.

La differenza tra ontologie e database risiede essenzialmente nello scopo per cui sono stati creati (Sir, Bradac, Fiedler, 2015): le ontologie sono focalizzate sul tema della comprensione automatica del significato, mentre i database si concentrano sulla memorizzazione e sul *retrival* dei dati.

Le ontologie, tendenzialmente, si costruiscono in modo incrementale e, ogni volta che si può, si cerca di sfruttare ontologie esistenti o quantomeno la loro struttura.

¹⁰ Guarda <http://dublincore.org/>

¹¹ Guarda <http://www.foaf-project.org/>

¹² Guarda <http://meta-guide.com/opencyc-natural-language-2014>

Nella creazione di un sistema di database, applichiamo la normalizzazione delle tabelle; tale normalizzazione è utilizzata per eliminare dati ridondanti ed è la migliore garanzia della qualità dei dati.

È prassi sempre più accreditata ricavare i flat file utilizzati dai packages statistici per le elaborazioni, da queries applicate su sottostanti database relazionali: in questo modo si è sicuri di lavorare su dati consistenti.

La metodologia che crea ontologie non include però le forme normali e in questo senso la loro diffusione per la generazione di sistemi informativi statistici potrebbe aprire importanti questioni legate alla qualità dei dati.

4.10.1. Comunicazione tra un'ontologia e un database

I sistemi informativi statistici potrebbero avvantaggiarsi da un buon livello di comunicazione e di allineamento tra ontologie e database gestionali. Nel proseguo di questo capitolo si esamineranno le diverse strategie di comunicazione e allineamento tra databases e ontologie, che, in linea di massima è possibile ottenere quando le informazioni memorizzate nell'ontologia corrispondono ai dati memorizzati nel database. Vysniauskas e Nemuratie descrivono tre possibili opzioni per veicolare dati tra questi due domini (Vysniauskas, Nemuraite, 2006):

- Usando la stessa tecnica modellazione concettuale per rappresentare l'ontologia e il database.
- Generando uno schema di database per l'ontologia.
- Ottenendo un database dall' ontologia.

In particolare esamineremo questi diversi casi: (i) come ottenere una ontologia da un database; (ii) come derivare databases dalle ontologie.

4.10.2. Ottenere una ontologia da un database

Quando si converte un database in una ontologia occorre che tutte le componenti del modello del database (ER diagramma) (Watt, Eng, 2012) siano riferibile all'ontologia.

Per raggiungere questo obiettivo si passa attraverso quattro fasi (Chujai, Kerdprasop, Kerdprasop, 2014): che spiegheremo nel modo seguente.

4.10.2.1. Il metodo

- i. Trasformare le entità

Ogni entità nel diagramma ER può essere una classe nell'ontologia.

- ii. Trasformare la relazione

Ogni relazione può essere trasformata in un object property nell'ontologia.

- iii. Trasformare gli attributi

Ogni attributo può essere trasformato in una data property nell'ontologia.

- iv. Trasformare la cardinalità

Restrizioni cardinalità vengono utilizzati per specificare ulteriormente le proprietà dell'ontologia (Pasha, Sattar, 2012), minCardinality e maxCardinality della proprietà sarà 1 se l'attributo è la chiave primaria o estere. Il minCardinality della proprietà è 1 se ogni attributo è dichiarato come NOT NULL. Inoltre, il maxCardinality della proprietà è 1, se qualsiasi attributo è dichiarato come UNICO.

Per i dettagli dell'applicazione del metodo si veda appendice 5.

4.10.2.2. Gli strumenti di conversione

Ci sono numerosi strumenti per raggiungere questo obiettivo, ma in questo paragrafo ci si concentrerà su due in particolare.

Il primo strumento "DataMaster" (Ravi, Sivaranjini, 2012) crea un'ontologia basata sul database relazionale che la converte. In questo caso il database e l'ontologia non avranno alcuna connessione tra loro e si applicherà la query SPARQL direttamente all'ontologia. Se nella prospettiva di utilizzo si dovrà inserire nuovi dati solo nel database, dovremo creare una nuova ontologia per per recepire i cambiamenti.

Il secondo strumento "D2R Server" (Chen, Zhao, Zhang, 2013) crea un collegamento tra le applicazioni semantic web e il database relazionale, senza creare alcuna ontologia; quando si applicherà una query SPARQL sarà applicata nel database direttamente perché lo strumento permette di convertirla in una query SQL prima di applicarla. Ogni volta che l'applicazione vuole applicare una nuova query SPARQL, si deve creare una nuova connessione con il database relazionale. Le applicazioni semantic web potranno avere a disposizione un sistema di conoscenza sempre aggiornato.

4.10.3. Ottenere un database da un'ontologia

Con questa metodologia si genera lo script SQL che descrive i concetti e le relazioni presentate nell'ontologia, consentendo così di generare un database attraverso una serie di regole di mappatura (Vysniauskas, Nemuraite, 2006).

4.10.3.1. Il metodo

I passi per avere un database da un'ontologia sono i seguenti:

- Il primo passo in cui lo strumento di generazione riceve l'ontologia come oggetto di ingresso.
- Si creano oggetti tipizzati come non lessicale, lessicale, o relazione (Saccol, Andrade, Piveta, 2011).
- Al terzo passo si applicano le regole di trasformazione e memorizzano le informazioni negli oggetti tipizzati.
- Al quarto passo si crea il file XML di output e lo script SQL corrispondente, che può essere modificato ed eseguito direttamente nel database.

Per maggiori dettagli si veda appendice 6.

4.10.3.2. Gli strumenti di conversione

Il tema di come gestire la conversione di un'ontologia in un database relazionale ha ricevuto una particolare attenzione da parte dei ricercatori, anche se ancora non sembra esserci una soluzione totalmente soddisfacente (Humaira, Tabbasum, Ayesha, 2015).

In questo lavoro si è utilizzato (applicato nel caso di studio) un plugin di Protégé : OWL2ToRDB¹³. Si tratta di un plugin Protégé per trasformare una ontologia espressa in OWL in un database relazionale.

Questo strumento affronta il problema di mantenere la semantica della ontologia una volta trasformata in un database relazionale: OWL2ToRDB segue un approccio ibrido, secondo il quale una parte dei concetti della ontologia viene direttamente mappata nello schema relazionale sulla base della comune semantica. I costrutti della ontologia che non hanno elementi equivalenti nello schema del database vengono archiviati in tabelle di metadati: le classi della ontologia, le proprietà e le istanze vengono direttamente mappate in tabelle del database. Gli assiomi e le restrizioni in *metatables*.

Lo strumento converte i contenuti dell'ontologia in statements SQL in modo che gli utenti possano vedere lo script della trasformazione e possano anche decidere di intervenire direttamente con aggiustamenti o cambiamenti.

OWL2ToRDB è stato applicato nel caso di studio. Si rimanda al caso di studio (§4.10.3.2) per la discussione sulle criticità dello strumento.

4.11. Ruolo delle ontologie per la costituzione di fonti amministrative e registri statistici

Gli elementi necessari per operare con le ontologie nelle politiche di riutilizzo delle fonti amministrative sono stati tutti discussi: (i) si è individuato nel contesto semantico importanti elementi concettuali e operativi; (ii) si sono discussi i ponti concettuali ed operativi per raccordare le ontologie

¹³ Guarda <http://protegewiki.stanford.edu/wiki/OWL2ToRDB>

con il mondo dei database, il contesto su cui sono costruiti gli archivi amministrativi.

Si può dunque affermare che una fonte amministrativa per la statistica può essere dunque ricondotta ad una ontologia, intesa come descrizione strutturata del suo contenuto informativo, basata sul suo modello concettuale (Di Fazio, 2015).

4.11.1. Ontologie e qualità della fonte statistica

Quando una fonte amministrativa può essere ricondotta ad una ontologia ufficiale o standard, allora esistono importanti garanzie sulla sua integrabilità in più ampi scenari informativi e, in ultima istanza sulla sua qualità: con un approccio di tipo hyperdimension (cfr.§3.7) niente si dice, infatti, sulle potenzialità di integrazione sistemica della fonte amministrativa valutata. Usando le ontologie, invece, questo aspetto può essere studiato fino al dettaglio delle singole istanze, anche grazie al supporto di utilities basate su metodi grafici e di inferenza logica (Brancato, 2015).

Sotto questo punto di vista, dunque, si può contenere e diversamente indirizzare l'attività delle agenzie nazionali che tipicamente si concentrano sulla produzione degli indicatori a supporto dei framework di valutazione della qualità; una agenzia nazionale, fungendo da garante sulle ontologie ufficiali, indirizzerebbe anche nella generazione di fonti amministrative adeguate al riuso, così come verrà successivamente discusso nel caso di studio.

4.11.2. Ruolo delle ontologie per la costituzione di sistemi gestionali interoperabili e dei registri

Come già illustrato precedentemente (cfr.§3.4), le fonti amministrative assumono un valore particolarmente significativo quando vengono strutturate (magari insieme a fonti di tipo statistico) nell'ambito dei registri statistici.

Nella prospettiva europea (Eurostat, 2013) i registri dei singoli stati membri devono uscire da una dimensione descrittiva puramente nazionale e devono riuscire a sintetizzare un punto di vista coerente ed europeo sui temi affrontati.

Durante i processi di costruzione di questi registri, si è dovuto affrontare e risolvere in sede di trattamento dei dati importanti problemi di eterogeneità delle fonti coinvolte che hanno risentito profondamente delle specificità nazionali: per questo motivo Eurostat ha deciso di affrontare il problema da un punto di vista interoperabile, lanciando il progetto europeo SBR (Statistical business register).

Anche nei riguardi del supporto alla interoperabilità per l'uso statistico il supporto delle ontologie può essere determinante.

Si ricordi la definizione di interoperabilità (cfr.§3.5).

Il web semantico affronta il tema dell'interoperabilità rafforzando la capacità dei sistemi partecipanti a lavorare in sinergia. Tale risultato si basa sulle capacità, tipiche delle ontologie, di rappresentare la conoscenza e di sostenere il ragionamento automatico e l'inferenza logica.

In questo senso, l'ontologia, definibile come una tupla (Blair et al., 2011) $\langle A, L, P \rangle$, dove A è un insieme di assiomi, L è un linguaggio in cui esprimere questi assiomi, e P la dimostrazione che supporta le conseguenze degli assiomi, assume evidentemente un ruolo importante per la sua capacità di comprendere quanto distanti siano due sistemi e in che misura possano lavorare insieme. A questo si aggiunga che la dimostrazione P permette di derivare anche relazioni che non sono state esplicitate e che sono implicite nella descrizioni dei sistemi che si intende far interoperare.

È interessante notare come questi tipi di applicazioni si basino su specifici *middleware* che potrebbero, in prospettiva rappresentare uno dei servizi innovativi delle agenzie statistiche nazionali.

4.11.3. Il ruolo delle ontologie per la costruzione di fonti integrabili e riusabili

Al centro di integrazione dei dati c'è, come si è visto, il concetto di ontologia, che è un'esplicita specificazione di una concettualizzazione condivisa (Guarino, 1998). Un uso comune delle ontologie è la standardizzazione dei dati e la concettualizzazione attraverso un linguaggio formale comprensibile dalla macchina. Per realizzare queste proprietà, sono stati utilizzati in uno dei tre modi seguenti (Wache et al., 2001):

- Approccio basato su una singola ontologia. Tutti gli schemi di origine sono direttamente correlati a un'ontologia globale condivisa da tutti i punti di vista che fornisce un'interfaccia uniforme per l'utente. Tuttavia, questo approccio richiede che tutte le fonti abbiano una visione molto simile su un dominio, con analogo livello di granularità.
- Approccio dell'ontologia multipla. Ogni fonte di dati è descritto da una propria ontologia (locale) separatamente. Invece di usare una ontologia comune, le ontologie locali sono mappate l'uno all'altro. A tal fine, un formalismo di rappresentazione supplementare è necessaria per definire le mappature tra ontologia.
- Approccio ibrido. Viene utilizzata una combinazione dei due precedenti approcci. Innanzitutto, un'ontologia locale è costruito per ogni schema di origine, che, tuttavia, non si associa ad altre ontologie locali, ma di un'ontologia globale condivisa. Nuove fonti possono essere facilmente aggiunte senza la necessità di modificare mappature esistenti.

Nel caso di studio discusso in questa tesi si è utilizzato l'approccio ibrido.

4.11.4. Ontologie e classificazioni statistiche

I sistemi di classificazione hanno un ruolo insostituibile nella costruzione dei sistemi informativi statistici: l'adozione di metodologie semantiche ne può potenziare l'applicazione e la gestione.

Nel sintetizzare infatti sistemi informativi statistici e registri a partire da fonti amministrative si devono sempre affrontare i seguenti problemi:

- Recepire e applicare il più possibile le classificazioni ufficiali;
- Armonizzare classificazioni che, pur provenendo da soggetti istituzionali sono divergenti, sovente contraddittorie e con granularità e concettualizzazioni diverse;
- Creare nuove classificazioni, magari a partire dalle concrete esperienze di gestione in tutti quei casi in cui tali categorizzazioni mancano.

Tutte queste situazioni sono state individuate ed affrontate nell'ambito del caso di studio affrontato in questo elaborato, e risolte, il più possibile con il supporto delle ontologie.

I sistemi di classificazione vengono affrontati, in contesto ontologico, attraverso gerarchie di categorie e sottocategorie popolate da individui rappresentati dai singoli codici.

In particolare, occorre notare che quando si inseriscono classificazioni ufficiali all'interno di una ontologia ci si imbatte in due situazioni diverse

- I. Sistemi di classificazione rappresentabili da gerarchie di categorie e sotto categorie. Quando si devono rappresentare questo tipo di codifiche si fanno delle rappresentazioni in cui ci sono le classi, ma non gli individui. A titolo di esempio, ISTAT non elenca tutte le professioni possibili che possono essere riferite al 5^a digit (NUP).
- II. Il secondo tipo è costituito da classificazioni che indirizzano ad un insieme definito di possibilità. Quando descritte con una ontologia queste classificazioni sono rappresentate da gerarchie di classi,

sotto classi e individui. Un esempio di questo tipo è la classificazione delle attività economiche.

La disponibilità di strumenti di inferenza logica e di intelligenza artificiale supporta nella individuazione di categorie, e nel loro popolamento.

Dal punto di vista pratico le classi che rappresentano i sistemi di classificazione possono essere create in diversi modi:

- imputazione da file di appoggio in formato testo,
- imputazione tramite fogli di calcolo
- Sfruttando tesauri in skos
- Attraverso file XML
- Tramite tabelle di database

Per maggiore informazione sulle trasformazioni si veda Appendice 1.

4.11.5. I sistemi gestionali supportati delle ontologie

Il tema dell'utilizzo delle ontologie si è fino a questo punto concentrato sulla rappresentazione di sistemi informativi statistici: vale però la pena anche riflettere sul ruolo che tali metodologie potrebbero rivestire nei sistemi puramente gestionali, ossia di sistemi con la capacità di interagire con il proprio dominio l'imputazione, attraverso la registrazione e la cancellazione di dati.

La grande variabilità organizzative e concettuale dei gestionali che insistono su uno stesso dominio rende, come già è stato ricordato, problematico il loro riutilizzo statistico. Per ovviare a questi problemi, in alcuni casi può essere utile impostare dei gestionali basati su ontologie.

Il sistema gestionale basato su una ontologia é un sistema che organizza i dati di un dominio e definisce tutti i componenti del contenuto del dominio, e ha la capacità di riusare i componenti che sono stati definiti all'interno di un altro sistema; i dati di questo sistema sono inoltre comprensibili da agenti automatici e conseguenza hanno la capacità di creare automaticamente una relazione con ulteriori sistemi attraverso processi di integrazione delle parti comuni.

- *I vantaggi del sistema gestionale basato su una ontologia*

In seguito vengono elencati i vantaggi dell'utilizzo di questo sistema:

- I. Riusabilità del sistema riduce il volume dei dati duplicati anche tra sistemi diversi.
- II. L'uso di concettualizzazioni standard nel sistema abbinato alla riusabilità del sistema evita di avere differenze tra i dati che descrivono una variabile in caso che la utilizziamo in più di un dominio;
- III. L'uso di stessi contenuti standard in più di un sistema integra i sistemi e supporta nella creazioni di una rete di sistemi informativi, particolarmente utile nel linkage;
- IV. La rete informativa dà la possibilità di interrogare i dati di più di un sistema nel stesso tempo usando solo una query;
- V. Il riuso di stessi contenuti standard in più di un sistema accelera il processo di aggiornamento dei sistemi perché basta aggiornare i contenuti originali;
- VI. Economico, perché, per migliorare le performances tra sistemi non c'è bisogno di potenziare i singoli nodi. Agendo su web, le prestazioni di integrazione non sono legate alle caratteristiche dei sistemi locali che memorizzano i dati, perché la velocità per avere i risultati di una simile query, non dipende localmente.

5. Caso di studio - Le grandi opere pubbliche

Nella riflessione fino a questo punto sviluppata si è visto come le fonti amministrative possono essere un importante strumento per la rappresentazione della complessità a condizione che il loro livello di integrazione sia profondo. Per questo si sono rilette, alla luce della teoria della complessità, le fonti amministrative e poi ci si è domandati se le nuove tecnologie semantiche potessero essere un supporto per fare assumere a tali fonti una connotazione da rete complessa.

Le considerazioni fino a qui sviluppate verranno adesso applicate ad un caso di studio che presenta tutte quelle caratteristiche di dispersione della conoscenza che ben si prestano ad essere supportate da una ontologia.

L'ambito di analisi è quello dei cantieri delle grandi opere: si tratta di contesti che sono il crocevia di importanti percorsi di rischio, per i quali è indispensabile predisporre adeguati sistemi informativi statistici che supportino nelle fasi di vigilanza, e di supporto a politiche attive di miglioramento.

La costruzione di questi sistemi informativi non è semplice: le fonti sono prodotte da soggetti diversi, caratterizzati da dialetti amministrativi e gestionali molto eterogenei. Non esiste un unico soggetto ufficiale in grado di presidiare il linguaggio: in alcuni casi ad uno stesso lemma corrispondono accezioni diverse a seconda che si parli, ad esempio, di norme di collaudo o di assicurazioni pubbliche. Il dettaglio a cui vengono rilasciati i dati ufficiali (tipicamente INAIL) non è mai al livello di granularità descrittiva necessaria per effettuare analisi sui percorsi di rischio.

5.1. *Database Monitor*

Il sistema Monitor¹⁴ è stato sviluppato dalle regioni Toscana ed Emilia Romagna in concomitanza con la costruzione della galleria dell'altra velocità ferroviaria tra Firenze e Bologna ed è attualmente adottato a

¹⁴ Guarda <http://www.infomonitor.it/>

supporto della vigilanza nei lavori della variante di Valico appenninico, nei lavori della tramvia di Firenze, nel nodo ferroviario di Firenze e nella stazione dell'alta velocità, nei lavori della terza corsia autostradale Barberino Calenzano e Barberino Firenze sud. Precedentemente era stato adottato anche per i lavori del nodo ferroviario di Bologna e del tratto ferroviario Bologna Piacenza Milano

Monitor è in prima istanza un gestionale pensato in funzione delle azioni delle autorità sanitarie per le loro azioni di vigilanza: in Monitor è stata costituita una memoria relazionale ottenuta armonizzando tutte le fonti che si generano nelle attività di cantiere (dalle buste paga ai verbali dei sopralluoghi alle registrazioni degli infortuni) in modo da avere una base informativa utile a indirizzare la vigilanza dove il rischio è più alto.

Il database relazionale sottostante a Monitor descrive infortuni, sopralluoghi, aziende e lavoratori, con un livello descrittivo incentrato sui processi lavorativi.

Questa impostazione ha permesso di produrre indicatori di rischio per processi, con un livello descrittivo fine sulle specifiche condizioni di lavoro in galleria.

Il sistema, nel corso degli anni, aveva subito numerosi rimaneggiamenti: il livello delle codifiche non era quasi mai allineato agli standard. In alcuni casi Monitor ha rappresentato un reale elemento di innovazione informativa, ad esempio quando sono stati identificati e codificati i processi lavorativi delle gallerie: nonostante fossero concettualizzazioni ovviamente familiari a livello gestionale, Monitor è l'unico contesto in grado di dare i profili infortunistici specifici per quei processi.

Con questo lavoro, dunque, (i) si è voluto ottenere una concettualizzazione standard per questi complessi contesti lavorativi a supporto delle analisi di epidemiologia occupazionale e di valutazione del costo economico degli infortuni; (ii) si è voluto creare un sistema in grado di generare gestionali utili al riuso ed alla sintesi di sistemi informativi statistici; (iii) si sono abbinati URI ai concetti analizzati, in modo da costituire una prima base per una ontologia delle grandi opere (iv) si è fatta una proposta di

processo di codifica, a partire dai dati raccolti da gestionali come Monitor per coprire lacune di codifica nella statistica ufficiale.

5.2. *La metodologia del trattamento delle tabelle*

Monitor si basa su un database relazionale, le cui tabelle sono state il punto di partenza per la costruzione della ontologia. Come primo passo dunque, coerentemente con quanto discusso nei paragrafi precedenti (cfr.§4.10.2) si è operato per trasformare il DB in un'ontologia.

La metodologia del trattamento delle tabelle di Monitor ha seguito le seguenti fasi:

- I. Studiare se la tabella sia una classe o sotto classe di un'ontologia esterna;
- II. analizzare e studiare le colonne della tabella;
- III. selezionare le colonne che possono essere rappresentate come una proprietà o classe di un'ontologia esterna;
- IV. creare le proprietà e le classi che sono state selezionate nella fase precedente e collegarle con l'ontologia esterna usando la proprietà di owl "owl:sameAs" per le proprietà e la proprietà di owl "owl:equivalentClass" per le classi;
- V. selezionare le colonne che possono essere rappresentate sulla base dei concetti o classificazioni standard;
- VI. tradurre le colonne che selezionate nella fase precedente in classi e proprietà;
- VII. creare il resto delle colonne come "Data properties".

Per meglio illustrare il lavoro che è stato effettuato, si presenta il lavoro di trattamento della tabella lavoratori: tutte le altre sono state affrontate nello stesso modo.

Per i dettagli costruttivi si veda la appendice 2 .

5.3. *Tabella Lavoratori*

Il lavoro di costruzione della ontologia delle Grandi opere è partita dalla rappresentazione del concetto di Lavoratori, descritto in Monitor della tabella lavoratore.

Questa tabella rappresenta le informazioni relative ai lavoratori che operano nel cantiere. La figura seguente mostra la tabella nel database gestionale originale.



Figura 5 la tabella lavoratore nel database Monitor

- *L'analisi della tabella Lavoratori*

Il percorso di rappresentazione della conoscenza sul lavoratore in una ontologia inizia dall'analisi delle colonne della tabella relazionale: si valuterà se alcuni di questi attributi possono essere riferiti a classificazioni internazionali, o se addirittura possono essere collegati ad altre ontologie.

5.3.1.ontologie esterne

Uno dei vantaggi più importanti dell'approccio per ontologie è riferibile al fatto che si possono utilizzare conoscenze già strutturate in ontologie esistenti. Nel caso del "lavoratore" sono disponibili domini di conoscenza già strutturati in ontologie.

Si tratta di ontologie che descrivono i seguenti concetti:

- Persona
- Indirizzo

I. Il concetto di persona- FOAF

Il lavoratore è una persona: per rappresentare questo concetto ci si è riferiti alla ontologia Friend Of A Friend (FOAF) , una struttura RDF progettata per rappresentare in modo standard metadati riferibili alle persone ed ai loro social networks (Brickley, Miller, 2010). È una ontologia pensata per descrivere le relazioni tra persone. In un contesto in cui tutte le persone si sono definite tramite FOAF, possono essere individuate in modo automatico, ad esempio, tutte le persone che hanno lavorato in un certo cantiere o che hanno dei contatti in comune. FOAF integra (i) reti sociali (ii) reti di rappresentazione (iii) reti di informazione. FOAF non è in concorrenza con i siti Web socialmente orientati; piuttosto fornisce un approccio in cui diversi siti possono unire diverse parti, e con cui gli utenti possono mantenere un certo controllo sulle loro informazioni in un formato non proprietario.

Utilizzando FOAF nella ontologia dei cantieri delle grandi opere, si assimila il contesto lavorativo ad un social network; le potenzialità
--

informative che si potrebbero cogliere, ad esempio negli studi di coorte sono evidenti.

Trattandosi di una applicazione RDF, FOAF può essere facilmente aggregata con altri vocabolari (come ad esempio Dublin Core) consentendo l'acquisizione di un ricco set di metadati (cfr.§4.9).

II. vCard ontology

L'ontologia vCard (Iannella, McKinney, 2014) ha una lunga storia. È stata proposta per la prima volta nel 1995 e poi standardizzata da IETF¹⁵ nel 1998. Da allora, è diventata un nuovo vocabolario, come ad esempio il vocabolario FOAF. La ontologia vCard è focalizzata sulla descrizione di persone e organizzazioni, comprese le informazioni sulla posizione e gruppi di tali entità.

Per rappresentare l'indirizzo italiano vCard non era sufficiente, per cui si è dovuto sviluppare una nuova ontologia che ha comunque messo in comune con vCard tutte le informazioni in possibili.

Con il riferimento alle ontologie esterne si è posto le basi per integrare i lavoratori nel panorama dei dati a cui sono riferibili su web. Se questa opzione è molto interessante per la costituzione di sistemi informativi statistici orientati alla costituzione di biografie, le implicazioni sulla privacy e il segreto statistico sono evidenti.

L'aggiornamento di questi concetti in fusione delle tecnologie semantiche è un altro degli appuntamenti a cui sono chiamati le agenzie nazionali e internazionali.

5.3.2.1 concetti standard

L'adozione di codifiche standard è lo strumento per eccellenza che i sistemi informativi statistici adottano per effettuare l'integrazione.

¹⁵ Internet Engineering Task Force

Questi sistemi di codifica non sono ancora ontologie e per il momento non consentono di cogliere tutte le opportunità discusse nei paragrafi precedenti.

In questo caso di studio si intende iniziare questo percorso.

Dopo aver dunque discusso le ontologie che potevano essere integrate nella descrizione del lavoratore, si valutano adesso quegli aspetti che non sono supportati da ontologie ma da concetti ufficiali standard.

Per questi schemi di codifica standard sono stati prodotti, nell'ambito della tesi le corrispondenti ontologie.

I concetti ufficiali per i quali si sono predisposte ontologie sono:

I. Classificazione delle professioni

Tale classificazione è stata adottata dall'ISTAT¹⁶ nel 2011 e riprende il formato della Nomenclatura delle Unità Professionali, elaborata nel 2006 in partnership con l'Isfol.

L'obiettivo della CP2011 è duplice: da un lato intende ricondurre tutte le professioni presente nel mercato del lavoro ad un numero limitato di raggruppamenti professionali grazie al concetto di competenza (Istat, 2013); dall'altro lato ha lo scopo di fornire informazioni preziose sulle caratteristiche di ogni profilo professionale.

Ogni unità professionale è un insieme di professioni omogenee rispetto a conoscenze, competenze, abilità ed attività lavorative. Ciascuna unità professionale è identificata da un codice a 5 cifre ed è accompagnata da un elenco di professioni dette voci professionali che fungono come esempio senza pretendere di essere esaustive.

¹⁶ <http://cp2011.istat.it/>

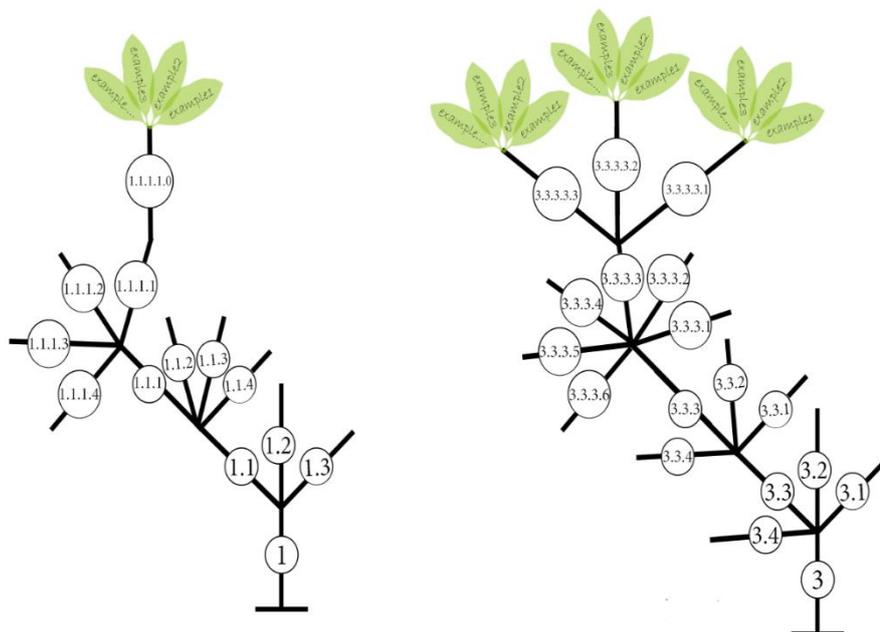


Figura 6 Rappresentazione delle gerarchica della classificazione delle professioni

II. La classificazione internazionale delle malattie

La classificazione internazionale delle malattie (ICD) è un sistema di classificazione che organizza le malattie e le lesioni in gruppi sulla base di criteri definiti (Im, Rm, Aht, 2011).

L'ICD-10¹⁷ è la decima revisione della ICD, che è la classificazione internazionale delle malattie e dei problemi connessi, proposto dalla OMS, classificati sopra di 2000 malattie. La traduzione ufficiale in italiano realizzata dall'ISTAT e l'Ufficio di Statistica del Ministero della Salute.

La classificazione è divisa in ventidue gruppi. Ogni gruppo di essa fa riferimento a uno specifico tipo di malattie e ha diversi sottogruppi; in questi sottogruppi ci sono i nomi delle malattie. Talvolta anche i sottogruppi hanno sottogruppi, e in questi sottogruppi sono elencati i nomi delle malattie.

¹⁷ <http://www.istat.it/it/archivio/6708>

III. Classificazione internazionale standard dell'istruzione

UNESCO ha sviluppato la classificazione internazionale dell'istruzione (UNESCO, 2012) (ISCED) per facilitare i confronti di statistiche sull'istruzione e indicatori tra i paesi sulla base di uniformi e definizioni concordate a livello internazionale. Nel 2011, una revisione ISCED è stata formalmente adottata dall'UNESCO Stati membri. Il prodotto di ampie consultazioni internazionali e regionali tra gli esperti di educazione e di statistica, ISCED 2011 tiene conto delle significative variazioni nei sistemi educativi in tutto il mondo rispetto all'ultima versione ISCED nel 1997.

IV. Il concetto di qualifica

Questo concetto è stato modellato a base dell'Art. 2095 Categorie dei prestatori di lavoro.

I prestatori di lavoro subordinato si distinguono in dirigenti, quadri, impiegati e operai (att. 95) (Comma così sostituito dalla Legge 13 maggio 1985, n.390).

Le leggi speciali (e le norme corporative), in relazione a ciascun ramo di produzione e alla particolare struttura dell'impresa, determinano i requisiti di appartenenza alle indicate categorie.

V. Il concetto di ruolo sicurezza

Il Rappresentante dei Lavoratori per la Sicurezza (RLS), nel diritto del lavoro italiano, è la figura, eletta o designata, che ha il compito in un'azienda di rappresentare i lavoratori per quanto concerne la salute e sicurezza nei luoghi di lavoro. E in particolare di individuare e valutare i rischi e le relative misure di sicurezza¹⁸.

¹⁸ Guarda

<http://www.sorgatoarchitettura.it/DOCUMENTI/CLIENTI/Le%20figure%20della%20sicurezza%20da%20A.S.L.E%20RLST%20di%20Milano-Lodi.pdf>

VI. L'indirizzo Italiano

L'Italia è stata divisa in venti regioni che hanno il codici dal 01 al 20. ogni regione è stata divisa in provincie (codice dal 001 al 107). Ogni provincia è divisa in comuni (codice di sei cifra dove le primi tre cifre indicano la provincia, le altre tre indicano il codice del comune)¹⁹.

5.4. *La modulazione ontologica dei Lavoratori*

I concetti della tabella relazionale possono ora essere mappati in una ontologia costituita da classi. I tali classi saranno riferite, quando possibile a ontologie esistenti, altrimenti verranno create Ex-novo a partire dalle classificazioni ufficiali ISTAT.

Le informazioni nella tabella del database relazionale presentata in apertura di capitolo sono adesso espresse in una ontologia rappresentata da classi relazionate per le quali sono state costituite le corrispondenti URI.

Qualsiasi applicativo che decida di adottare questa connotazione e descriversi in questi termini potrà adesso godere di tutte quello potenzialità di sintesi e di integrazione che erano state presentate e discusse nei capitoli iniziali.

Le classi che sono state create sono le seguenti.

- *Class lavoratori*

Class Persona	Sub class of	Thing
Class Persona	Equivalent to	Foaf:Person
Class lavoratori	Sub class of	Class persona

Dominio	Object properties	Range	Restrictions
Lavoratori	sonoImpiegatoDa	Ditta	Only
Persona	haMalattia	Malattia	Some
Lavoratori	haOccupatoRuoloSicurezza	RuoloSicurezza	Only

¹⁹ Guarda <http://www.istat.it/it/archivio/6789>

RuoloSicurezza	ruoloSicurezzaSonoStatiOccupatiDa	Lavoratori	Some
Lavoratori	haProfessione	Professioni	Only
Professioni	professioniSonoStatiOccupatiDa	Lavoratori	Some
Lavoratori	haLivelloDiIstruzione	Istruzione	Only
Lavoratori	haQualifica	Qualifica	Only
Qualifica	qualificaSonoStatiOccupatiDa	Lavoratori	Some
Persona	haSesso	Sesso	Only
Persona	cittàDIResidenza equivalent To vcard:hasLocality	Comuni	Only
Persona	HaLuogoDiNascita	Comuni	Only
Persona	capDIResidenza equivalent To vcard:hasPostalCode	Cap	Only
Persona	cittàDIabitazione	Comuni	Max 1
Persona	capDIabitazione	Cap	Max 1
Persona	haStatoCivile	statoCivile	Only
Lavoratori	siTrovaIn	Cantieri	Some
Lavoratori	HaCapacita	Capacita	Some
Lavoratori	HaConoscenze	Conoscenze	some

Dominio	Data properties	Range	Restrictions
Lavoratori	haIdLavoratori	String	Only
Persona	haCognome equivalent To Foaf:familyname	String	Only
Persona	haNome equivalent To Foaf:firstname	String	Only
Persona	hadataDiNascita equivalentTo Foaf:birthday	Date time	Only
Thing	Note	String	Only
Lavoratori	haAlloggio	Boolean	Only
Persona	viaDIResidenza equivalent To vcard:street-address	string	Only
Persona	viaDIabitazione	String	Max 1

Dopo aver rappresentato il lavoratore si presentano le classi con cui il lavoratore ha una relazione (via object property). Si noti come, le informazioni che prima venivano racchiuse negli attributi della tabella del database, adesso sono rappresentate da autonome classi standard

relazionate con la classe lavoratore, con un evidente miglioramenti delle capacita espressive e di integrazione.

- *Class RuoloSicurezza*

Class RuoloSicurezza	Sub class of	Thing
----------------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Lavoratori	haOccupatoRuoloSicurezza	RuoloSicurezza	Only
RuoloSicurezza	ruoloSicurezzaSonoStatiOccupatiDa	Lavoratori	Some

Dominio	Data properties	Range
RuoloSicurezza	rdfs:label	String
RuoloSicurezza	rdfs:commento	String
RuoloSicurezza	haCompiti	String

Individui		
AddettiAlleEmergenze	rdfs:label	Addetti Alle Emergenze
	rdfs:commento	Gli addetti alle emergenze sono i lavoratori incaricati dal datore di lavoro per l'attuazione delle misure di protezione e prevenzione e della gestione delle emergenze.
	haCompiti	Gli addetti alle emergenze intervengono direttamente nei casi di pericolo grave ed immediato sul cantiere. Per la particolare importanza del loro compito, i lavoratori nominati non possono rifiutare la designazione se non per giustificato motivo, devono essere formati e disporre di attrezzature adeguate. La loro formazione deve avvenire durante l'orario di lavoro e sono esenti da responsabilità soggette a sanzione. È compito del datore di lavoro, in collaborazione con il servizio di prevenzione, prevedere procedure per la gestione delle emergenze, vale a dire attuare quelle procedure operative indicate per interventi di evacuazione dei lavoratori in caso di pericolo grave e immediato, di salvataggio, di pronto soccorso, di prevenzione incendi e lotta antincendio.
	rdfs:label	RSPP

Responsabile Del Servizio Di Prevenzione e Protezione	rdfs:commento	È la persona, con attitudini e capacità adeguate, incaricata dal datore di lavoro, per l'individuazione e valutazione dei rischi e delle relative misure di sicurezza.
	haCompiti	Il responsabile del servizio di prevenzione e protezione deve essere nominato dal datore di lavoro previa consultazione dell'RLS / RLST. Anche gli addetti al servizio di prevenzione e protezione (ASPP) possono essere interni o esterni all'unità produttiva. Le loro capacità, nonché i requisiti professionali devono essere adeguati alla entità dei rischi presenti sul luogo di lavoro e relativi alla attività lavorativa: devono comunque essere in possesso di un titolo di studio, almeno di un diploma di istruzione secondaria superiore, con attestato di frequenza a specifici corsi di formazione. Sono inoltre tenuti a frequentare corsi di aggiornamento. Nelle società con meno di 30 dipendenti può essere il datore di lavoro stesso, previo specifico corso di formazione. Le attitudini e capacità adeguate derivano da compiti svolti in precedenza in materia di Prevenzione e Protezione. Il nominativo del Responsabile deve essere segnalato all'ASL e alla Direzione Provinciale del lavoro, allegando curriculum professionale e requisiti, considerando che può anche essere una persona esterna e non è sanzionato per lo svolgimento dei suoi compiti.
Rappresentante Dei Lavoratori Per La Sicurezza	rdfs:label	Rappresentante dei lavoratori per la sicurezza
	rdfs:commento	È il rappresentante dei lavoratori per quanto concerne gli aspetti della salute e della sicurezza. Viene eletto dai lavoratori con un'apposita assemblea aziendale (RLS). Questa figura può essere individuata anche in ambito territoriale (RLST).
	haCompiti	Deve ricevere una adeguata formazione e seguire uno specifico corso presso l'Organismo Paritetico, che rilascerà un attestato di avvenuta formazione. Nel considerare il suo profilo, il suo ruolo e i suoi compiti, si tenga conto di questi elementi: a) è esente da responsabilità sanzionabili b) non può subire alcun pregiudizio nell'espletamento delle sue funzioni c) ha libertà di accesso ai luoghi di lavoro d) riceve il documento di valutazione dei rischi e accede al Registro Infortuni

		<p>e) promuove iniziative idonee a tutelare la salute dei lavoratori</p> <p>f) formula osservazioni in caso di visite delle autorità competenti</p> <p>g) segnala al responsabile dell'azienda i rischi individuati</p> <p>h) partecipa alla riunione periodica di prevenzione e può richiederne la convocazione</p> <p>i) dispone del tempo e dei mezzi necessari per l'esercizio delle sue funzioni come previsto dal contratto collettivo di lavoro.</p>
LavoratoreAutonomo	rdfs:label	Lavoratore Autonomo
	rdfs:commento	È quella persona fisica che partecipa alla realizzazione dell'opera con risorse proprie, senza rapporti di subordinazione o dipendenti propri.
	haCompiti	Il lavoratore autonomo deve possedere i requisiti di idoneità tecnico-professionale verificabili attraverso l'iscrizione alla C.C.I.A.A. In materia di sicurezza deve premurarsi di utilizzare attrezzature di lavoro conformi alla normativa di sicurezza, così come idonei dispositivi di protezione individuale, quelli relativi a rischi specifici della propria attività. Inoltre si deve adeguare a quanto indicato dal coordinatore in fase esecutiva. Anche il lavoratore autonomo può incorrere in sanzioni di tipo penale per l'inosservanza alle disposizioni che lo riguardano.

- *Class StatoCivile*

Class StatoCivile	Sub class of	Thing
-------------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Persona	haStatoCivile	statoCivile	only

Individuo
Celibe
Nubile
Sposato
Sposata
Divorziato
Divorziata
Vedovo
Vedova

- *Class Professioni*

Professioni sono stati rappresentati come una gerarchia di classi e sotto classi basate sulla classificazione dei professionisti. In questo paragrafo usiamo la testa della gerarchia (classe Professioni) per rappresentare questa gerarchia.

Class Professioni	Sub class of	Thing
-------------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Lavoratori	haProfessione	Professioni	Only
Professioni	professioniSonoStatiOccupatiDa	lavoratori	some

Dominio	Data properties	Range
Professioni	rdfs:label	string
Professioni	rdfs:commento	string

- *Classe Malattia*

Malattie sono stati rappresentati come una gerarchia di classi e sotto classi basate sulle classificazioni delle malattie. In questo paragrafo usiamo la testa della gerarchia (classe Malattia) per rappresentare questa gerarchia.

Class Malattia	Sub class of	Thing
----------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Persona	ha Malattia	Malattia	some

Dominio	Data properties	Range
Malattia	rdfs:label	string
Malattia	rdfs:commento	string

- *Class Sesso*

Class Sesso	Sub class of	Thing
-------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Persona	haSesso	Sesso	only

Dominio	Data properties	Range
Persona	rdfs:label	string

Individuali		
Maschile	rdfs:label	M
Femminile	rdfs:label	F

- *Class Istruzione*

L'istruzione è stata rappresentata come una gerarchia di classi e sotto classi a basi della classificazione di ISCAED. In questo paragrafo usiamo la testa della gerarchia (classe Istruzione) per rappresentare questa gerarchia.

Class Istruzione	Sub class of	Thing
------------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Persona	haLivelloDiIstruzione	Istruzione	only

Dominio	Data properties	Range
Istruzione	rdfs:label	string
Istruzione	rdfs:commento	string

- *Class Qualifica*

Class Qualifica	Sub class of	Thing
-----------------	--------------	-------

Dominio	Object properties	Range	Restrictions
Lavoratori	haQualifica	Qualifica	Only
Qualifica	qualificaSonoStatiOccupatiDa	Lavoratori	some

Dominio	Data properties	Range
Qualifica	rdfs:label	string

Qualifica	rdfs:commento	string
-----------	---------------	--------

Individuali		
Dirigenti	rdfs:commento	qualifica attribuita a chi ricopre un ruolo dall'elevato grado di professionalità. Il dirigente ha la possibilità di esercitare in modo molto autonomo i propri poteri decisionali, per realizzare gli scopi dell'impresa. Assieme a quadri, impiegati e operai, i dirigenti sono una delle quattro categorie di lavoratori subordinati previste dal codice civile (art. 2095 del c.c.).
	rdfs:label	Dirigenti
Quadri	rdfs:commento	assieme a dirigenti, impiegati ed operai, sono una delle quattro categorie di lavoratori subordinati previste dal codice civile: sono lavoratori che svolgono attività di rilevante importanza ai fini dello sviluppo e dell'attuazione degli obiettivi dell'impresa, ma non hanno ruoli dirigenziali.
	rdfs:label	Quadri
impiegati	rdfs:commento	termine che indica generalmente il lavoro dipendente presso enti privati o pubblici (ove si parla di "pubblico impiego"). Assieme a dirigenti, quadri e operai, sono una delle quattro categorie di lavoratori subordinati previste dal codice civile.
	rdfs:label	Impiegati
Operai	rdfs:commento	assieme a dirigenti, quadri e impiegati, sono una delle quattro categorie di lavoratori subordinati previste dal codice civile: sono coloro cui è demandato il lavoro prettamente di manodopera.
	rdfs:label	Operai

- *Class IndirizzolItaliano*

Class Italia	Sub class of	Thing
Class Regione	Sub class of	Class Italia
Provincia	Sub class of	Class Italia
Comuni	Sub class of	Class Italia
Cap	Sub class of	Class Italia

Dominio	Object properties	Range	Restrictions
Cap	haComuni	Comuni	Only
Comuni	haProvincia	Provincia	Only
Provincia Comuni	& haRegione	Regioni	Only

Persona	cittàDIResidenza equivalent vcard:hasLocality	to	Comuni	Only
Persona	HaLuogoDiNascita		Comuni	Only
Persona	capDIResidenza equivalent vcard:hasPostalCode	to	Cap	Only
Persona	cittàDIabitazione		Comuni	Only
Persona	capDIabitazione		Cap	Only

Dominio	Object properties	Range	Restrictions
IndirizzoItaliano	Rdfs:label	String	
Regioni	haPosizioneGeografico	String	Only

5.5. *I concetti standard della ontologia di grandi opera*

Dopo aver esaminato nel dettaglio il concetto di lavoratore ed aver trasformato molti dei suoi attributi in riferimenti ad ontologie già esistenti o a classi rappresentanti concetti ufficiali, si presentano adesso le altre classi presenti nel dominio del cantiere delle grandi opere. Esse Sono:

- La classificazione delle attività economiche
- Agente Materiale
- Tipologia di lavoro
- Tariffa Inail
- Tipo Verbali
- Motivazione Sopralluogo
- Sede Lesione e Natura della Lesione
- Modalità infortunio
- Fasi costruttive
- Area di Attività (ADA)

I. La classificazione delle attività economiche

La classificazione delle attività economiche²⁰ –ATECO2007- costituisce la versione nazionale della nomenclatura europea Nace Rev. 2, pubblicata sull'Official Journal il 20 dicembre 2006 (Regolamento (CE) n.1893/2006 del PE e del Consiglio del 20/12/2006). L'obiettivo era tenere conto delle specificità della struttura produttiva italiana ed individuare attività particolarmente rilevanti nel nostro Paese.

La classificazione Ateco 2007 presenta le varie attività economiche raggruppate, dal generale al particolare, in sezioni, divisioni, gruppi, classi, categorie e sottocategorie.

Ciascuna attività economica è quindi codificata con un numero di sei cifre; le prime due cifre sono separate da un punto dalle seconde due, le seconde due sono a loro volta separate da un punto dalle ultime due.

II. Agente Materiale

Nella descrizione delle modalità di infortunio un ruolo importante e rivestito dall'agente materiale coinvolto.

Ci sono tre variabili che forniscono informazioni sugli agenti materiali coinvolti nell'infortunio (Eurostat, 2012):

- l'agente materiale associato all'attività fisica specifica'
- l'agente materiale associato all'deviazione'
- l'agente materiale associato al contatto - modalità di lesione'.

I tre agenti materiali non devono necessariamente essere diversi. Ma è ovviamente possibile che ogni punto corrisponda ad un diverso 'Agente materiale'. In alcuni casi non c'è 'Agente materiale' da registrare o codificare, e questo il caso, ad esempio di una commessa in piedi che volta per servire un cliente. Il movimento provoca una lesione ma nessuno agente materiale è coinvolta.

²⁰ Guarda <http://www.istat.it/it/strumenti/definizioni-e-classificazioni/ateco-2007>

III. Tipologia di lavoro

La tipologia di lavoro descrive il tipo di lavoro o di attività (attività generale) in corso di esecuzione da parte della vittima al momento dell'incidente (Eurostat, 2012).

Questa variabile non è l'occupazione della vittima: la descrizione del tipo di lavoro e di compito, in termini generali, che la vittima stava eseguendo durante un periodo di tempo che termina all'istante dell'incidente.

IV. Tariffa Inail

La Tariffa inail²¹ indica i differenti premi per le assicurazioni contro gli infortuni sul lavoro e le malattie professionali applicabili in concreto in base all'attività aziendale:

- industria
- artigianato,
- terziario,
- altre attività

Per determinare la tariffa dei premi assicurativi, le lavorazioni sono classificate in dieci grandi gruppi, di norma articolati in gruppi, sottogruppi e voci. La voce di tariffa è, quindi, l'unità elementare di classificazione alla quale corrisponde una o più lavorazioni assicurate. A ogni voce di tariffa è abbinato un tasso medio nazionale detto anche tasso medio di tariffa.

V. Tipo Verbali

Il Verbale è un documento redatto dopo ogni sopralluogo. I verbali possono essere di tre tipi:

- Prescrittivo: è un atto con cui l'organo di vigilanza stabilisce e prescrive qualcosa al datore di lavoro, il quale è tenuto ad osservare.
- Disposizione: è un atto con cui l'organo di vigilanza impartisce, sulla base di un apprezzamento discrezionale, un'ingiunzione al

²¹ Guarda http://www.inail.it/internet_web/appmanager/internet/home

datore di lavoro di adottare le misure prevenzionali ad integrazione e specificazione dell'obbligo generico contenuto in una norma primaria in bianco o di portata molto ampia ed indeterminata.

- Richiamo alla norma: è un atto con cui l'organo di vigilanza accerta una violazione facendo riferimento ad una norma.

VI. Motivazione Sopralluogo

Ogni indagine sul luogo è effettuata con differenti motivazioni quali²²:

- Sopralluogo Iniziativa
- Sopralluogo Inchiesta Infortunio
- Inchiesta Malattia Professionale
- Verifiche Impianti Macchine
- Indagini Ambientali
- Su Segnalazioni di Problemi
- Verifica di Adempimenti

VII. Sede Lesione e Natura della Lesione

L'INAIL²³ ha definito una classificazione sia per la sede della Lesione sia per la Natura della lesione indicando per "sede lesione" il luogo o meglio la parte del corpo che ha subito una lesione; mentre per "natura della lesione" si indica il tipo di alterazione a carico di un tessuto o di un organo che comporti un cambiamento.

VIII. Modalità infortunio

È una classificazione Inail utile per descrivere come è avvenuto l'infortunio. La sua tassonomia è la seguente:

- Attivo
- Passivo
- Incidenti
- Ambiente
- Cadute

²² Guarda <http://www.infomonitor.it>

²³ Guarda http://www.inail.it/internet_web/appmanager/internet/home

IX. Fasi costruttive

Sono dei processi edili che consentono la realizzazione dell'opera, tra loro collegate e coordinate che richiedono la gestione di mezzi (macchine, strumenti, materiali) di uomini (mano d'opera) tempi e dena²⁴.

X. Area di Attività (ADA)

Indica un insieme di compiti omogeneo per tipo o prodotto realizzato, orientato alla produzione di un risultato ed identificabili all'interno di uno specifico processo lavorativo²⁵. Tale insieme rappresenta un descrittore dei Repertori regionali delle Competenze.

5.6. *La forma finale dell'ontologia di Grande Opera*

Il diagramma precedente rappresenta tutti i classi e tutti gli object properties che collegano i classi tra loro. Per maggiore informazione sulla struttura delle classi si veda Appendice 2.

²⁴ Guarda <http://www.infomonitor.it>

²⁵ Guarda <http://web.rete.toscana.it/RRFP/pdf/Glossario.pdf>

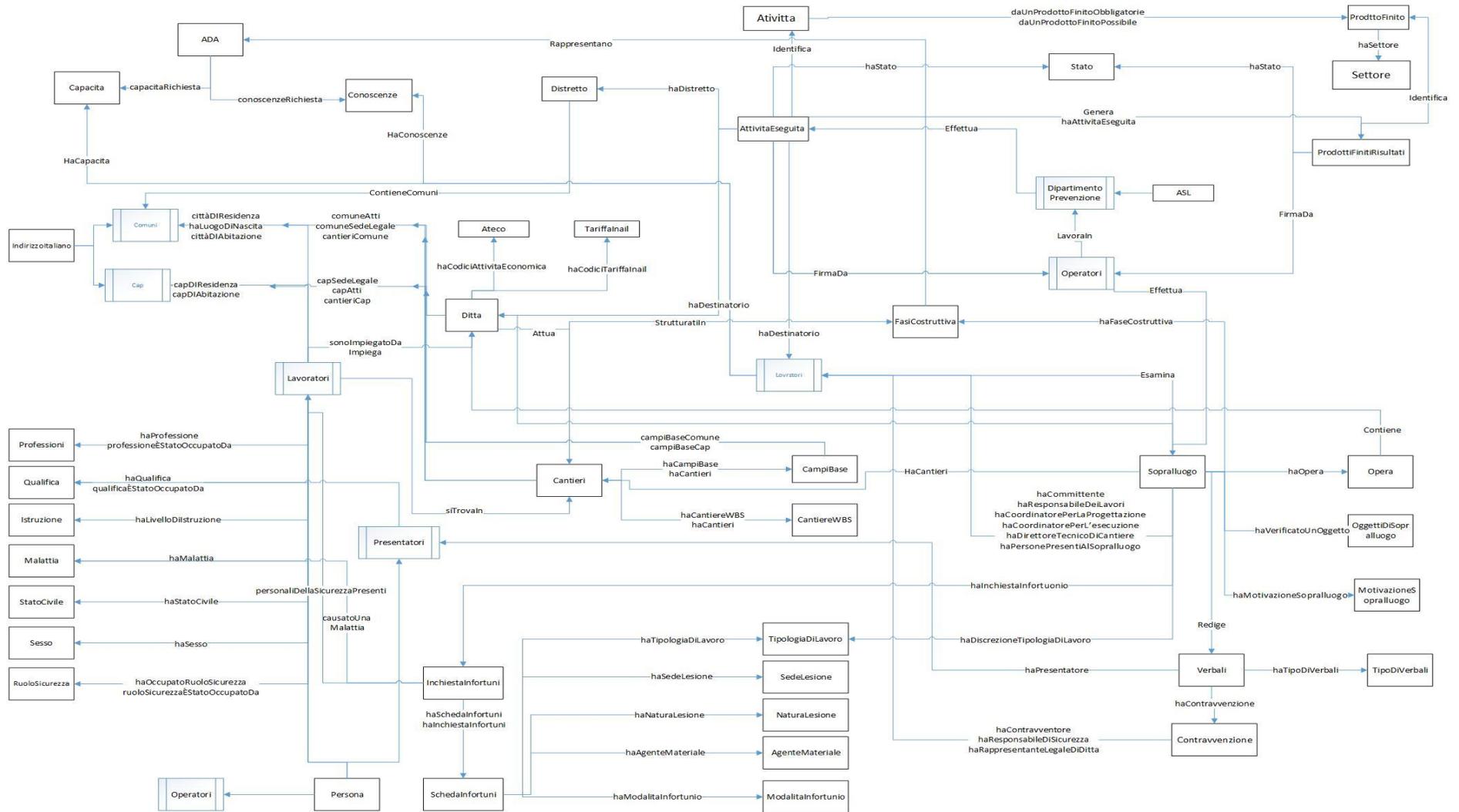


Figura 7 Ontologia della Grande Opera: classi e object properties

5.7. *La visualizzazione dell'ontologia*

Per la manipolazione grafica delle ontologie sono disponibili interessanti utilities che consentono un più veloce approfondimento delle potenzialità informative del sistema.

Di seguito verranno presentate alcune applicazioni.

Per visualizzare l'ontologia e applicare le query usiamo il programma Gruff²⁶: si tratta di un navigatore interattivo triple-store, gestore di query. Funziona su AllegroGraph da Franz Inc. e in misura minore su qualsiasi endpoint SPARQL. Le informazioni possono essere sfogliate come grafi di nodi e linee di collegamento, e come le tabelle descrittive delle proprietà di specifici nodi. Le query possono essere scritte come SPARQL o in Prolog, o gestite graficamente come diagrammi di nodi e linee di collegamento. Quando si utilizza gli AllegroGraph stores, i dati possono essere creati e modificati compilando tavole di valori di proprietà, oppure collegando i nodi rappresentati dai grafi con le linee di collegamento per generare così le nuove triple. I vari punti di vista e gli strumenti sono strettamente integrati in un impianto pensato per una rapida navigazione, interrogazione e modifica.

Dopo il caricamento dell'ontologia possiamo ricavare lo schema RDF dove, nella parte centrale della schermata si vede una porzione della ontologia mentre, nella parte sinistra in basso si vede, in un rettangolo, lo schema nella sua interezza. In alto a sinistra sono rappresentate le indicazioni dell'ontologia come se vede nella foto successiva.

²⁶ Guarda http://franz.com/agraph/gruff/gruff_documentation.html

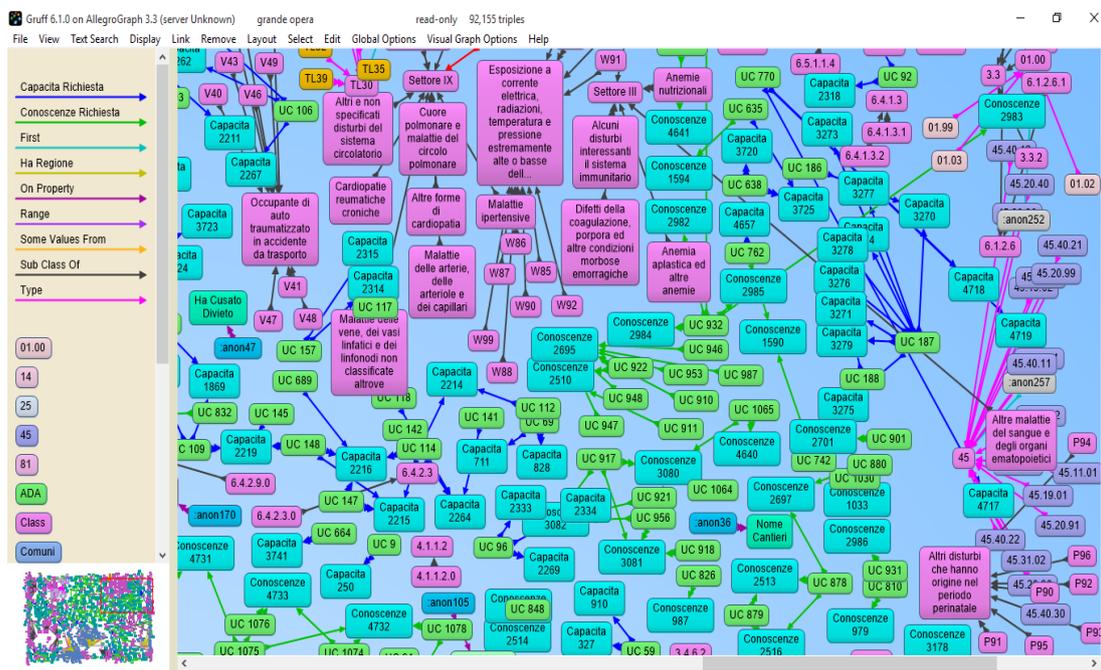


Figura 8 Una vista parziale delle triple dell'ontologia Grande Opera

Al di là della specificità della rappresentazione, si noti come sia aumentata l'espressività semantica di questa descrizione rispetto alla tabella di Fig.5; praticamente ogni elemento di questa rete ha un suo identificativo unico, tipicamente memorizzato in rete, che, qualora riferito ad altre fonti e altri applicativi, consente l'innescò di operazioni automatiche di integrazione.

In questo senso dunque una ontologia realizza l'auspicio espresso nei primi capitoli: il sistema delle informazioni diventa una rappresentazione fedele della complessità dei cantieri delle grandi opere pubbliche. Se gli uffici nazionali di statistica, unitamente alle authority che governano gli appalti pubblici, regolamentassero i gestionali di aziende appaltanti, delle istituzioni sanitarie e di vigilanza, si potrebbero ricavare informazioni fino ad ora sepolte in archivi e faldoni incapaci di fare sistema.

5.7.1. Interrogare graficamente una ontologia

Per applicare una query su questa base di conoscenza si sceglie da view list l'opzione query view, si imposta la query e si applica. Ad esempio ipotizziamo che si voglia sapere quali sono le attivita che danno il prodotto finito " Inchiesta per tossinfezione alimentare (PF_5) " e a quale settore viene questo attribuito. Nella foto successiva vediamo la query e i risultati.

The screenshot shows the AllegroGraph 3.3 interface. At the top, the title bar reads "Gruff 6.1.0 on AllegroGraph 3.3 (server Unknown) grande opera read-only 92,155 triples". The menu bar includes "File", "View", "Text Search", "Display", "Edit", "Global Options", "Query Options", "Table Options", and "Help". The interface has several tabs: "SPARQL" (selected), "Use Planner", "Reindent", "Name Query", "Revisit", "Graph View", "Table View", and "Graphical Query View". The "Run Query" button is highlighted. Below the tabs, there is a text area for the query:

```
PREFIX : <http://www.semanticweb.org/adhan/ontologies/2015/10/GrandeOpera#>
SELECT ?daProdotto ?Attivita ?label ?settore WHERE { ?Attivita ?daProdotto :PF_5.
?Attivita rdfs:label ?label.
:PF_5 :haSettore ?settore
}
```

To the right of the query area, there is a text box with instructions: "Enter a SPARQL select or describe query in the text widget to the left, then press the Run Query button. All known namespace abbreviations will be in effect. Or first select the Prolog radio button and enter a Prolog query (perhaps with additional list forms as well). Basic Emacs keystrokes can be used to edit the text. In the results table, left-click a column header cell to sort the table by that column. Left-click a node cell to visit that". Below the query area, there are buttons for "Create Visual Graph", "Add to Visual Graph", "Write Text Report", and "Save as CSV". The results section shows "12 Results" and a table with the following data:

?daProdotto	?Attivita	?label	?settore
Da Un Prodotto Finito Obbligatorie	Attivita 7	aggiornamento archivio cartaceo informatico	IAN
Da Un Prodotto Finito Obbligatorie	Attivita 46	inchiesta epidemiologica	IAN
Da Un Prodotto Finito Obbligatorie	Attivita 91	relazione finale	IAN
Da Un Prodotto Finito Obbligatorie	Attivita 14	approfondimenti tecnico scientifici / legislativi / documentali	IAN
Da Un Prodotto Finito Obbligatorie	Attivita 38	elaborazione/analisi dati	IAN

Below the table, there are two panels: "Explicit Nodes from Query" showing "PF_5" and "Explicit Predicates from Query" showing "Ha Settore" and "Label". At the bottom left, the URL "http://www.w3.org/2000/01/rdf-schema#label" is visible.

Figura 9 Esempio di risultati di una query sparql

Il programma ci da la possibilità di avere il risultati in diverse forme; una di queste è la forma grafica come vediamo nella foto successiva.

Converrà sottolineare come questo risultato differisca dalla interrogazione di un normale database. Le risorse che concorrono alla costruzione del risultato potrebbero trovare ovunque. La stessa avvertenza vale, ovviamente anche per tutti gli esempi successivi.

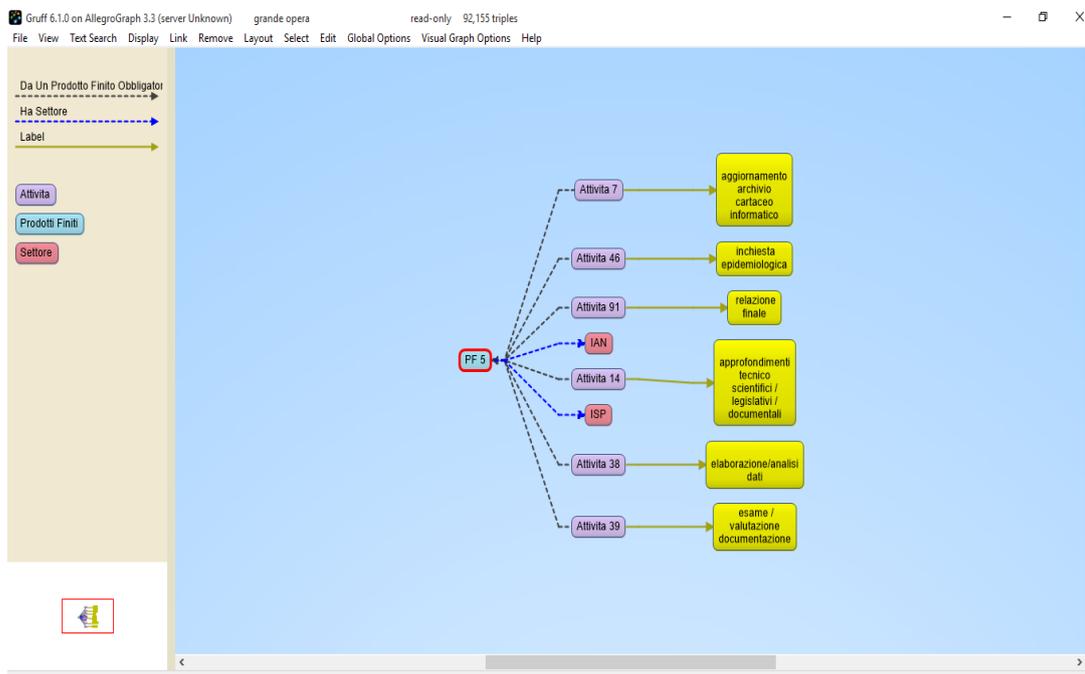


Figura 10 esempio di risultati di una query sparql in modo grafico

5.8. Il collegamento tra due individui non sono collegati direttamente

È essenziale riuscire a creare collegamenti tra individui della ontologia attraverso una serie di object properties.

Qui di seguito un esempio eseguito con il programma Gruff: si consideri il caso dell'individuo PISLL (Settore Prevenzione, igiene, sicurezza nei luoghi di lavoro) e l'individuo Attivita_39 (esame / valutazione documentazione) come vediamo nella figura successiva.

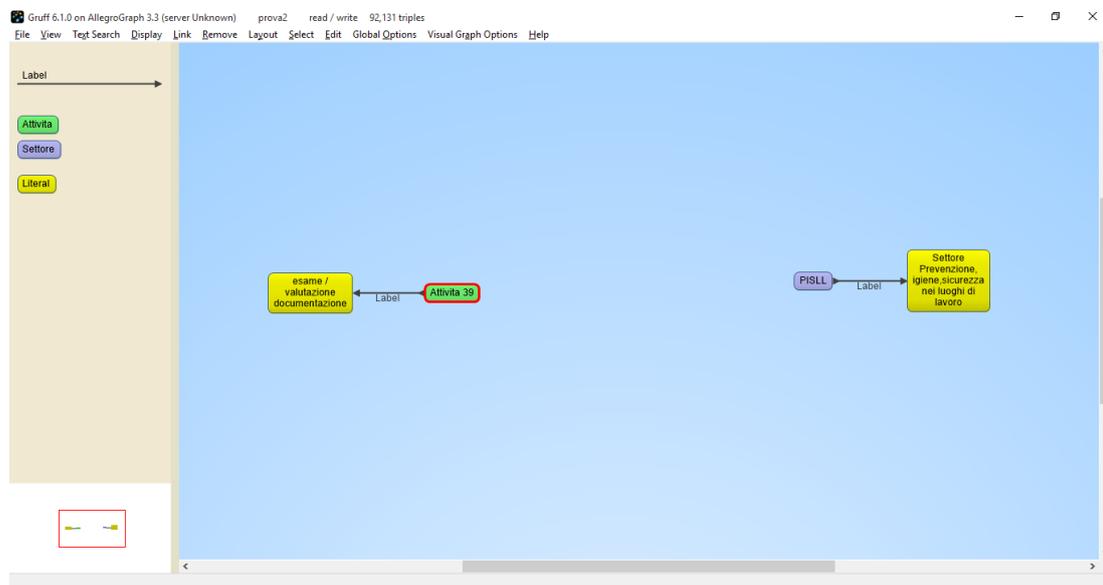


Figura 11 Individui della ontologia Grande Opera

Il programma apre una lista di properties per scegliere quelle che vogliamo usare, come vediamo nella figura successiva. Nel caso in cui non sappiamo le properties che servono per creare il percorso possiamo scegliere tutte le properties.

Sostanzialmente il sistema consente di esplorare quali processi (informativi in ultima istanza, ma gestionali in origine) consentono di legare tra loro questi concetti.

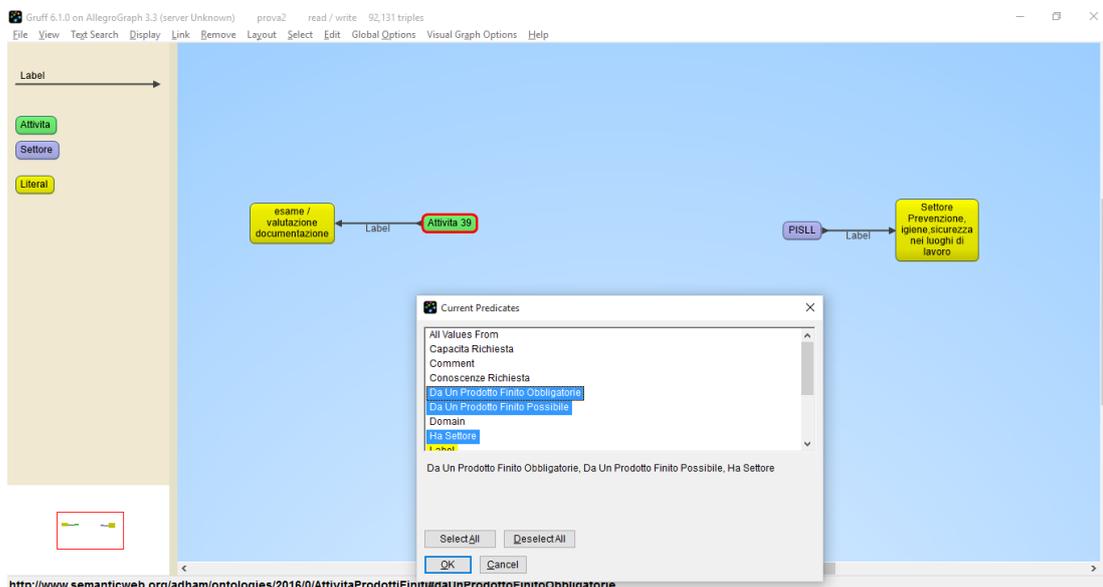


Figura 12 Collegamenti tra individui della ontologia attraverso una serie di object properties

Il programma mostra una schema contiene tutti i percorsi che partano dall'Attività_39 e arrivano al PISLL come vediamo nella figura successiva.

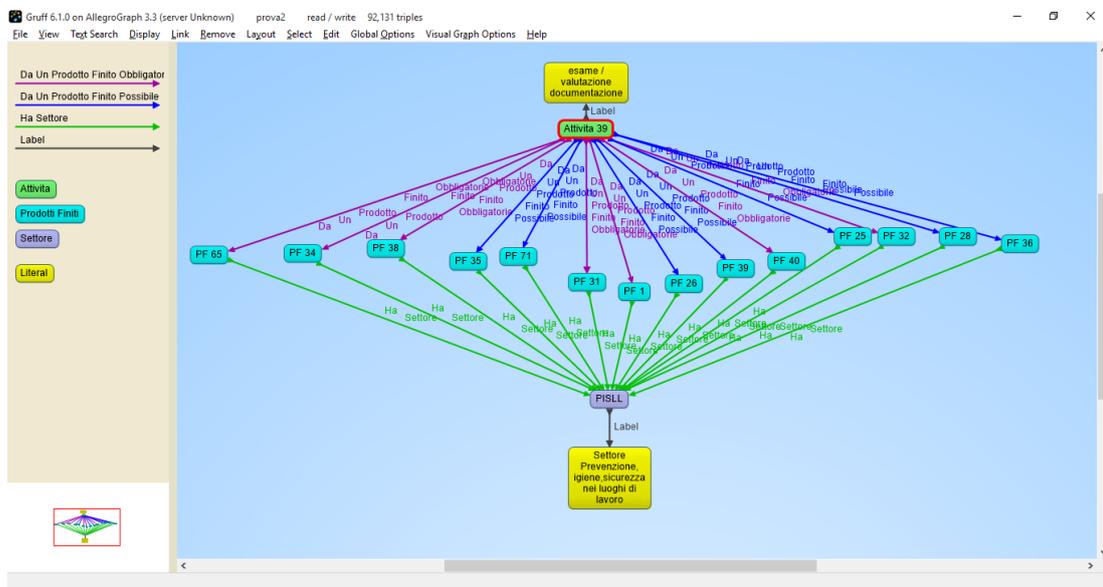


Figura 13 Risultati di un collegamenti tra individui della ontologia attraverso una serie di object properties

Si tratta a tutti gli effetti di funzionalità che consentono al ricercatore di esplorare strade alternative di linkage, nel caso si vogliano individuare le fonti per aggregare insieme dei caratteri che non compaiono in nessuna fonte comune.

Sono chiarissimi i rischi che si corre da un punto di vista dei sistemi informativi statistici: il fatto che i percorsi di connessione siano tecnicamente fattibili (questo mostra l'utilizzo del sistema) non garantisce che siano corretti da un punto di vista statistico.

Per questo, è urgente una forte presenza della ricerca statistica ad affiancare questi applicativi informatici.

5.9. Estrarre e combinare le informazioni da due ontologie diverse

Come abbiamo spiegato nei capitoli precedenti, per estrarre le informazione da due ontologie diverse abbiamo bisogno di trovare almeno un elemento in comune tra loro o avere un concetto in un'ontologia definito come un sinonimo (owl:sameAs) or (owl:equivalentClass) di un altro concetto nell'altra ontologia.

Nell'ontologia della Grande Opera abbiamo definito il concetto Malattia in base alla classificazione delle malattie. Nella rete c'è una ontologia che si chiama HumanDiseaseOntology; tra queste due ontologie ci sono tante cose in comune. Per integrarle, basta definire ogni class dalla tassonomia della class malattia come (owl:sameAs) or (owl:equivalentClass) con il corrispondente nel HumanDiseaseOntology.

Per vedere come si estraggono le informazione si usa il programma Gruff per caricare entrambi i dataset.

Poi applichiamo questa ad esempio una query che un legame tra le informazioni della class (Batteri, virus ed altri agenti infettivi) e la sua corrispondente la class disease by infectious agent.

I dettagli di questa query sono riportati in appendice 7

nella figura successiva se vede le informazioni estratte.

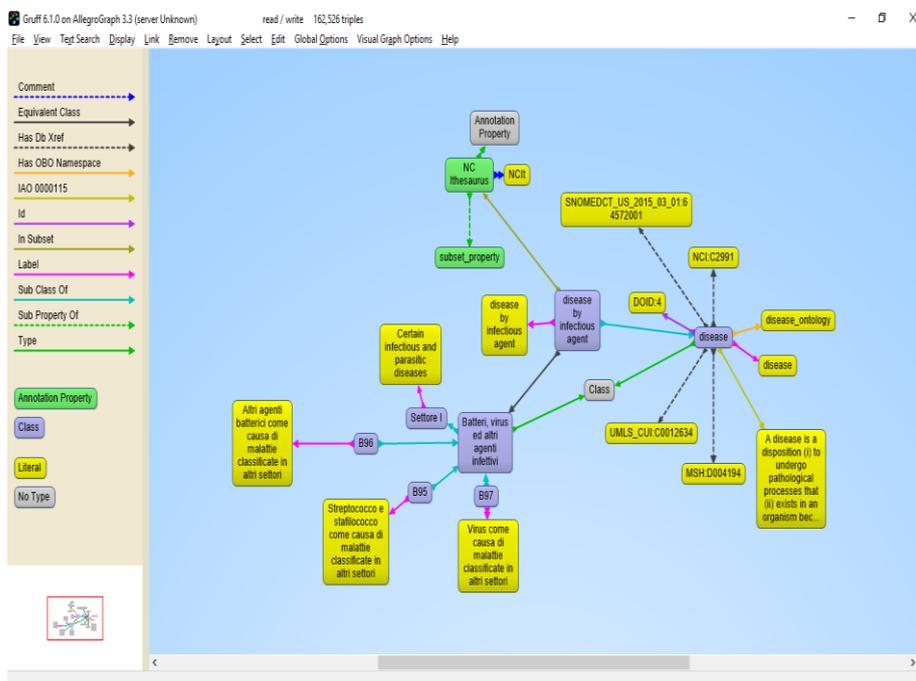


Figura 14 I risultati del legame tra ontologie diverse

La Figura mostra graficamente il risultato:

Ad ogni lavoratore che abbia contratto una infezione il sistema aggancia anche tutte le informazioni presenti nella HumanDiseaseOntology: è bastato affermare l'equivalenza del concetto di malattia.

Questo esempio rafforza l'importanza di un presidio statistico ufficiale sui concetti sulla cui base, in modo tecnicamente così semplice, si effettuano i linkage.

5.10. La creazione di un nuovo database Monitor sulla base dell'ontologia della Grande Opera

Tutte le considerazioni precedentemente svolte sono fortemente rafforzate se il lavoro di riferimento all'ontologia non viene fatto ex post, a partire da dati già generati, ma ex ante. In altri termini si vuole sperimentare in questo specifico caso di studio, quanto discusso a livello teorico nel paragrafo (cfr.§4.10.3)

A partire dalla ontologia appena presentata si ricava dunque una nuova versione del database di Monitor, in modo tale che i dati archiviati dal sistema siano già direttamente riferibili alle URI dell'ontologia e integrabili in un sistema informativo statistico delle grandi opere, magari di respiro nazionale.

Per la creazione delle nuove tabelle del database avevamo la scelta tra due metodi (cfr.§4.10.1):

I. Il metodo automatico

In questo metodo possiamo usare il plugin Owl2Rdb di Protege, con l'avvertenza che la metodologia adottata da questo plugin trasforma ogni classe in una tabella. Nel nostro caso, se avessimo usato questo plugin avremmo ricavato centinaia di tabelle che non servono: sostanzialmente questo plugin può essere usato utilmente solo quando l'ontologia non ha sotto classi.

Un approccio a difficoltà incrementale, ad esempio, può prevedere di condizionare la conversione delle classi a tabelle solo quando le classi sono popolate da individui.

II. Il metodo manuale

Per creare una nuova versione di Monitor allineata con l'ontologia si è adottata questa metodologia:

1. si comincia la trasformazione dalla class principale nell'ontologia (nel nostro caso è la class Lavoratori)
2. si crea una tabella con il nome della classe
3. si trasforma ogni data property in una colonna
4. si trasforma ogni object property: ci sono due possibilità a seconda del tipo di classe che si intende collegare alla classe principale
 - a. se la seconda classe ha un numero finito di individui e ha solo *ha* come data property, allora la seconda classe diventerà una colonna nella tabella della classe principale. Ad esempio: per esempio l' object property haSesso e la classe Sesso diventa l'attributo "sesso";

- b. Nelle altre situazioni la seconda classe diventerà una tabella collegata con la tabella della classe principale tramite un chiave o una terza tabella costituita da due colonne che rappresentano le chiavi delle due tabelle.
5. Nel caso in cui nella ontologia ci siano delle tassonomie (ad esempio persona--> lavoratore si trasforma in una tabella relazionale solo quella popolata da individui, ossia il lavoratore. Queste tassonomie hanno infatti un senso in un contesto come quello delle ontologie che devono supportare il ragionamento automatico ma non ne hanno nei database.

Applicando questa procedura si è ricavato una nuova versione del database Monitor, normalizzato e aderente alla ontologia. Per maggiore informazione sulla struttura delle tabelle si veda Appendice 3.

.

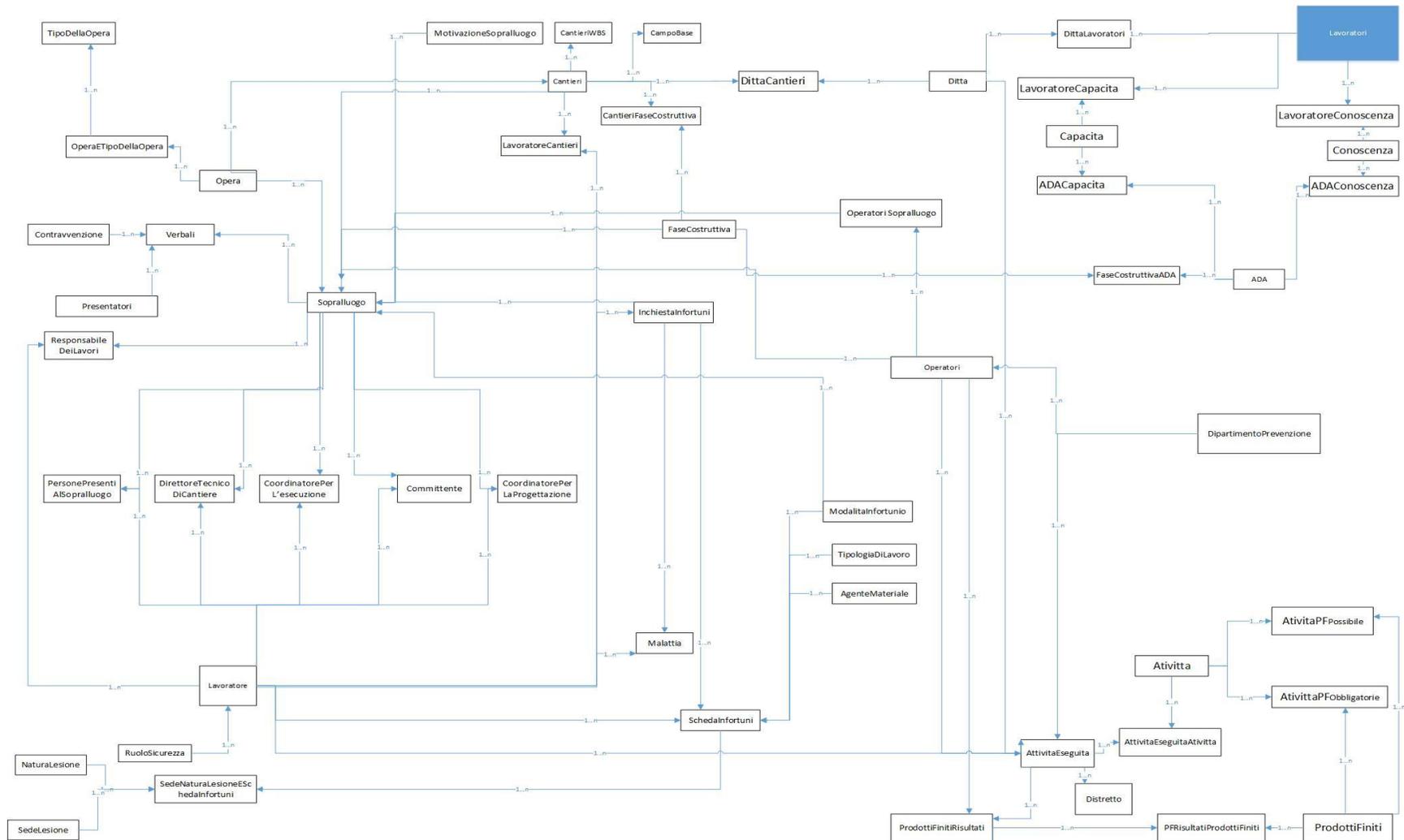


Figura 15 La struttura del nuovo database Monitor

5.11. Il funzionamento congiunto di una ontologia e di un database

L'impiego diffuso dei database nella nostra vita quotidiana rende molto difficile sostituire l'utilizzo dei database con le ontologie. Questo lavoro ha suggerito una metodologia per usare entrambi (databases e ontologie) in un sistema gestionale unificato.

Il senso di questa operazione dal punto di vista dei sistemi informativi statistici, va visto nella necessità di impostare delle strategie di raccolta di semantiche utilizzate in specifici contesti nella prospettiva futura di impostare dei sistemi di codifica condivisa.

Considerando il caso di studio esaminato, l'esempio dei processi produttivi è illuminante.

Non esiste una codifica unica dei processi produttivi in edilizia: a seconda dei contesti vengono utilizzate delle denotazioni diverse. Si può dunque immaginare di impostare su alcuni gestionali selezionati come caso di studio dei meccanismi di aggiornamento di una ontologia in funzione di quanto viene imputato nei database.

Questa si configurerebbe, dunque, come una base da cui far partire dei processi di condivisione che dovrebbero portare alla ontologia ufficiale della agenzia nazionale statistica.

In seguito spieghiamo i passi dell'aggiornamento suggerito:

I. La creazione di sparql.txt file

Con la creazione del programma si crea un unico text file con il nome " sparql" in cui si scrive tutti i "namespace" che servano per l'ontologia per esempio:

PREFIX : < qui deve essere URI dell'ontologia di grande opera >

PREFIX owl: <<http://www.w3.org/2002/07/owl#>>

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

PREFIX xml: <http://www.w3.org/XML/1998/namespace>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

- II. L'aggiornamento della programmazione delle icone (inserisce, aggiorna e cancella)

I gestionali dovranno essere dotati di icone per l'inserimento delle codifiche (come individui) nella ontologia (cfr.§Appendice 4)

I passi di lavoro saranno i seguenti:

(i) inserimento dei dati nel database;(ii) scrittura di una sparql insert query nel sparql.txt file.(Questo viene applicato usando un algoritmo che in questo lavoro è stato scritto nel linguaggio Pseudocode prendendo in considerazione che le variabili prendano i loro valori dei campi nella finestra e anche prendendo in considerazione che ci sono alcuni campi non obbligatori. Per il codice pseudocode si veda appendice 4 (iii) gli applica la query sparql: in qualunque momento, possiamo ricavare le sparql query dal sparql.txt file dove sono state scritte dal programma gestionale e applicarle all'ontologia tramite uno sparql endpoint.

6. Conclusioni

Le nuove tecnologie semantiche rappresentano una sfida, un'opportunità ed un rischio per la statistica ufficiale: consentono infatti di gestire le informazioni quantitative con una potenza elaborativa inedita, ma rischiano di generare sistemi informativi molto lontani da quegli standard di qualità che sono indispensabili per l'analisi statistica.

In questo lavoro si è ripensato al riuso statistico delle fonti amministrative alla luce delle potenzialità che potrebbero venire da una consapevole integrazione con la tecnologia semantica. Si è pensato cioè ad una strategia che consenta di contenere il grande spreco di memoria pubblica dovuto alla difficoltà di integrazione delle fonti: sistemi informativi adeguati alla gestione di problemi e servizi risultano inservibili se ripensati per il riuso dei loro dati.

Naturalmente esistono e sono sempre più diffuse le metodologie big data, ma difficilmente con questi approcci si può arrivare al dettaglio informativo e alla qualità e precisione di misura necessari quando si devono affrontare problemi di grande specificità e delicatezza, come tipicamente sono quelli sanitari o quelli legati alle condizioni sociali e di lavoro.

Per questo ci si è concentrati su come usare queste tecnologie per supportare tutto il processo di generazione degli archivi, fin dal momento della loro concettualizzazione.

Al di là dei risultati di fattibilità tecnica, discussi nel caso di studio, è interessante aver evidenziato degli inediti ambiti di presenza delle agenzie statistiche: il presidio del linguaggio e delle concettualizzazioni che, se adottate il modo ampio, consentirebbero una ben diversa qualità delle fonti amministrative. Si tratta di una integrazione che non va solo nel senso di una più vasta diffusione delle codifiche ufficiali, ma che supporta anche il percorso inverso: quello

del riconoscimento delle modalità di denotazione degli esperti dei diversi domini, per conoscerli, apprenderli, integrarli e diffonderli ufficialmente.

La possibilità di connotare ogni concetto di un suo identificativo ufficiale memorizzato su internet, la scelta di far adottare queste modalità agli attori sociali ed economici che agiscono sui territori e nei processi, la costituzione di grandi testi interpretabili automaticamente, sposta gli usuali orizzonti di chi si occupa di sistemi informativi statistici: processi come quelli di linkage o di test delle condizioni di rispetto della privacy assumono una scala e una problematicità di ordine estremamente maggiore e nuovo per la comunità statistica.

Bibliografia

Allemang, D. & Hendler, J., 2011. *Semantic Web for the Working Ontologist: Effective modeling in RDFS and OWL* Second Edi. D. J. H. Allemang, USA: elsevier. Available at: <http://www.sciencedirect.com/science/article/pii/B9780123859655100160>.

Ameen, A., Rani, B.P. & Khan, K.U.R., 2014. Semi-Automatic Merging of Ontologies using Protégé. *International Journal of Computer Applications*, 85(12), pp.35–42.

Angles, R., 2012. A Comparison of Current Graph Database Models. *IEEE Intelligent Systems and Their Applications*, 28th Inter, p.171.

Backstrom, L. & Ugander, J., 2012. Four Degrees of Separation. , pp.33–42.

Baldacci, E., 2014. L ' uso statistico degli archivi amministrativi : una risorsa per il Sistan e per il Paese. ISTAT.

Barabasi, A.-L., Albert, R. & Jeong, H., 2000. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A*, 281, pp.69–77. Available at: http://ac.els-cdn.com/S0378437100000182/1-s2.0-S0378437100000182-main.pdf?_tid=916ba2a6-d949-11e5-957d-00000aacb35f&acdnat=1456134750_e5b1607da889e215012c6608a1ed6e9f.

Berners-lee, T. et al., 2001. The Semantic Web. *Scientific American*, 284(5), pp.35–43.

Berners-lee, T., 1998. Why RDF model is different from the XML model. *Design Issues note*, (September). Available at: <https://www.w3.org/DesignIssues/RDF-XML.html>.

Bititci, U.S. et al., 2004. Creating and Managing Value in Collaborative Networks. *International Journal of Physical Distribution & Logistics Management*, 34(3-4), pp.251–268.

- Blair, G.S. et al., 2011. The role of ontologies in emergent middleware: Supporting interoperability in complex distributed systems. HAL, 7049 LNCS, pp.410–430.
- Brancato, G., 2015. che utilizzano dati amministrativi Linee guida. ISTAT, Versione 1.
- Brickley, D. & Guha, R.V., 2014. RDF Schema 1 . 1. W3C Recommendation. Available at: <https://www.w3.org/TR/rdf-schema/>.
- Brickley, D. & Miller, L., 2010. FOAF Vocabulary Specification 0 . 98. Namespace Document, (August). Available at: <http://xmlns.com/foaf/spec/>.
- Buccella, A., Cechich, A. & Brisaboa, N.R., 2004. Taking Advantages of Ontology and Contexts to Determine Similarity of. Universidad Nacional Del Comahue, 048.
- Calzaroni, M., 2008. Le fonti amministrative nei processi e nei prodotti della statistica ufficiale. Atti della Nona Conferenza Nazionale di Statistica, pp.15–16. Available at: <http://www.istat.it/it/files/2011/02/Calzaroni.pdf>.
- Calzaroni, M. & Contini, B., 2004. La cooperazione Inter-istituzionale: il valore aggiunto dell'integrazione di informazioni. VII Conferenza nazionale di statistica, Roma 9-10 novembre.
- Cerroni, F., Bella, G. Di & Galiè, L., 2014. Evaluating administrative data quality as input of the statistical production process. RIVISTA DI STATISTICA UFFICIALE, n.1-2, p.117.
- Chandrasekaran, B., Josephson, J.R. & Benjamins, V.R., 1999. What are ontologies, and why do we need them? IEEE Intelligent Systems and Their Applications, 14, pp.20–26.
- Chen, Y., Zhao, X. & Zhang, S., 2013. Publishing RDF from Relational Database Based on D2R Improvement. WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS, 10(8), pp.241–248.

Chujai, P., Kerdprasop, N. & Kerdprasop, K., 2014. On Transforming the ER Model to Ontology Using Protégé OWL Tool. *International Journal of Computer Theory and Engineering*, 6(6), pp.484–489. Available at: <http://www.ijcte.org/index.php?m=content&c=index&a=show&catid=59&id=1092>.

Colander, D. & Kupers, R., 2014. *Complexity and the art of public policy*, New Jersey: Princeton university.

Costamagna, E. & Spanò, A., 2012. Integrazione di database spaziali multiscala in ambito urbano Original. *Conferenza Nazionale ASITA*, Vicenza, pp.483–490.

Cyganiak, R., Wood, D. & Lanthaler, M., 2014. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation. Available at: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/Overview.html>.

Daas, P.J.H. et al., 2008. *Proposal for a quality framework for the evaluation of administrative and survey data*. Statistics Netherlands.

Eurostat, 2013. *European profiling of the actual economic activity The European System of interoperable Business Registers (ESBRs)*. European Union, pp.92–93.

Eurostat, 2012. *European Statistics on Accidents at Work (ESAW) Summary methodology 2012 editi.*, Luxembourg: Publications Office of the European Union.

Fazio, D., 2013. *BLUE-Enterprise and Trade Statistics final results*. BLUE-ETS has, (July).

Di Fazio, S., 2015. *Istituto Centrale per gli Archivi Modello concettuale, ontologia e schemi RDF SAN. Linked Open Data del SAN. Dataset aggregati da SAN*. Istituto Centrale per gli Archivi.

Fisher, D., 2006. *An emergent perspective on interoperation in systems of systems*. Software Engineering Institute, (March), pp.1–67. Available at: <http://repository.cmu.edu/sei/411/>.

Franceschetti, M. & Meester, R., 2006. navigation in small-world networks: a scale-free continuum model. *journal of applied probability*, 43(4), pp.1173–1180.

Gasti, G., 2014. l' ontologia formale del “realismo naturale”, cosmologia evolutiva e partecipazione dell' essere. *divus thomas* 117-2, 2, pp.229–334.

Grobe, M., 2009. RDF, Jena, SparQL and the “Semantic Web.” *Proceedings of the ACM SIGUCCS fall conference on User services conference*. Available at: <http://dl.acm.org/citation.cfm?id=1629501.1629525>.

Guarino, N., 1998. Formal Ontology and Information Systems. *Proceedings of the first international conference*, 46(June), pp.3–15. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.1776&rep=rep1&type=pdf>.

Hamon, V., 2013. Malicious URI resolving in PDF documents. *Journal of Computer Virology and Hacking Techniques*, 9(2), pp.65–76. Available at: <http://link.springer.com/10.1007/s11416-013-0179-2>.

HorrIDGE, M., 2011. A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools Edition 1.3. The University Of Manchester. Available at: http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf.

Humaira, A., Tabbasum, N. & Ayesha, S., 2015. A Survey on Automatic Mapping of Ontology to Relational Database Schema. *Research Journal of Recent Sciences*, 4(4), pp.66–70.

Iannella, R. & Mckinney, J., 2014. vCard Ontology - for describing People and Organizations. W3C Interest Group Note 22. Available at: <http://www.w3.org/TR/vcard-rdf/>.

IM, M., RM, L. & AHT, R.-S., 2011. History of the Statistical Classification of Diseases and Causes of Death. National Center for Health Statistics.

ISO/IEC, 2013. Information Technology — Metadata registries (MDR) — Part 6 : Registration. ISO/IEC, N1884(JTC 1/SC 32/WG 2).

Istat, 2013. La classificazione delle professioni, Roma: Istituto nazionale di statistica.

Karr, A.F., Sanil, A.P. & Banks, D.L., 2006. Data quality: A statistical perspective. *Statistical Methodology*, 3(2), pp.137–173.

Klyne, G. & Carroll, J.J., 2004. Resource Description Framework (RDF) : Concepts and Abstract Syntax. W3C Recommendation. Available at: <https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>.

Kriglstein, S. & Wallner, G., 2010. Knoocks - A Visualization Approach for OWL Lite Ontologies. *Intelligent and Software Intensive Systems Knoocks*, pp.950–955. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5447389>.

Laurini, R. & Murgante, B., 2008. Interoperabilità semantica e geometrica nelle basi di dati geografiche nella pianificazione urbana. Franco Angeli, Milano, pp.229–244.

Lawrence, E.E., 2015. Ant colony optimization.

Martelli, C., 2011. A linguistic approach to the construction of complex statistical information systems. Association for Survey Computing.

Martelli, C. & Bellini, E., 2012. The role of a semantic approach of complex statistical information system planning and management Tourism observatories in social dialogue and participative governance processes. *IEEE Intelligent Systems and Their Applications*.

Martini, M., 2004. Qualità degli archivi amministrativi: aspetti teorici e ipotesi di miglioramento degli archivi degli agenti economici in Italia. Franco Angeli, Milano.

- Masolo, C. et al., 2003. La Prospettiva dell ' Ontologia Applicata La Prospettiva dell ' Ontologia Applicata. *Rivista di Estetica*, pp.170–183.
- Matentzoglou, N., Bail, S. & Parsia, B., 2013. A corpus of OWL DL ontologies. *CEUR Workshop Proceedings*.
- McGuinness, D.L. & van Harmelen, F., 2004. OWL Web Ontology Language Overview. W3C Recommendation. Available at: <https://www.w3.org/TR/owl-features/>.
- Motik, B. & Parsia, B., 2012. OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. W3C Recommendation. Available at: <https://www.w3.org/TR/owl2-syntax/>.
- Namyoun Choi, Song, I.-Y. & Han, H., 2006. A Survey on Ontology Mapping. *ACM New York*, 35(3), pp.34–41.
- Pasha, M. & Sattar, A., 2012. Building domain ontologies from relational database using mapping rules. *International Journal of Intelligent Engineering and Systems*, 5(1), pp.20–27.
- Pironi, S. et al., 2013. Technical Report / Jul . 2013 Ontologie e modellazione di dati sanitari. Technical Report.
- Pinto, H.S. & Martins, J.P., 2001. A methodology for ontology integration. *Proceedings of the international conference on Knowledge*, pp.131, 138. Available at: <http://portal.acm.org/citation.cfm?doid=500737.500759>.
- Quintano, C.C., 2008. Le fonti amministrative , una risorsa primaria della statistica ufficiale. *Nona conferenza nazionale di statistica*.
- Raimond, Y. & Raimond, Y., 2014. RDF 1 . 1 Primer. W3C Working Group, pp.1–14.
- Ravi, L. & Sivaranjini, N., 2012. Towards Ontology Development Based on Relational Database. *International Journal of Web Technology*, 01(02), pp.106–111.

- Rethemeyer, R., 2005. Conceptualizing and Measuring Collaborative Networks. *Public Administration Review*, 65, pp.117–121.
- Rinaldi, A., 2002. Fonti informative e indicatori statistici per l'analisi socio - economica territoriale. , Istituto G. Tagliacarne, Roma, Paper n. 3, p.17.
- Saccol, D. de B., Andrade, T. de C. & Piveta, E.K., 2011. Mapping OWL ontologies to relational schemas. *IEEE International Conference on Information Reuse Integration*, (10), pp.71–76. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5298643>.
- Scorza, F., 2009. il progetto di un'ontologia della pianificazione come strumento di supporto alla programmazione dello sviluppo regionale europeo. conferenza italiana di scienze regionali il.
- Singh, J., 2005. Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science*, 51 n.5, pp.756–770.
- Sir, M., Bradac, Z. & Fiedler, P., 2015. Ontology versus Database. *scienceDirect*, 48(4), pp.220–225. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S2405896315008113>.
- The Center for Survey Statistics and Methodology, 2015. Improving the Methodology for Using Administrative Data in an Agricultural Statistics System. *Global Strategy Working Papers*, (1), p.12.
- Tronti, L., 2007. I dati amministrativi per le statistiche sui mercati del lavoro locali : il progetto Guida.
- Tusch, G. et al., 2009. Exploring Microarray Time Series with Protégé. *th International Protege Conference*.
- Tusch, G. et al., 2007. The Protégé-Owl SWRLTab and Temporal Data Mining in Surgery. *th International Protege Conference*.
- Udrea, O., Pugliese, A. & Subrahmanian, V.S., 2007. GRIN : A Graph Based RDF Index. *AAAI*.

- UNESCO, 2012. International Standard Classification of Education Isced 2011, Canada: UNESCO Institute for Statistics.
- United Nations, 2000. terminology on statistical metadata. conference of european statisticians statistical standards and studies, 53. Available at: <http://www.unece.org/fileadmin/DAM/stats/publications/53metadaterminology.pdf>.
- Vale, S., 2006. How Real are the Barriers to the Use of Administrative Sources for Statistical Purposes? , pp.2–5.
- Vale, S., 2013. Using Administrative Data in Statistical Registers.
- Vega-Redundo, F., 2007. Complex social networks. Cambridge University Press.
- Vicknair, C. et al., 2010. A Comparison of a Graph Database and a Relational Database. ACM.
- Vyšniauskas, E., Nemurait, L. & Paradauskas, B., 2012. Preserving Semantics of Owl 2 Ontologies in Relational Databases Using Hybrid Approach. information technology and control, 41(2), pp.103–115.
- Vysniauskas, E. & Nemuraite, L., 2006. Transforming ontology representation from OWL to relational database. Information technology and control, 35(3), pp.333–343.
- Wache, H. et al., 2001. Ontology-Based Integration of Information - A Survey of Existing Approaches. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.12.8073>.
- Watt, A. & Eng, N., 2012. Database Design 2nd Editio.,
- Webber, J., Eifrem, E. & Ian Robinson, 2015. Graph Databases 2nd rditio., United States of America: O'Reilly Media.
- Wesley, Y., Lavallée, P. & Julie, T., 2012. The use of Administrative Data at Statistics Canada Wesley. Statistics Canada, XXXIII(2), pp.81–87. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15003161>.

Whitman, L. et al., 2006. An enterprise model of interoperability e Panetto
To cite this version : , Saint Etie.

Wing, J.M., 2010. Understanding Network Complexity. Proceedings of
the Second IEEE International Workshop on Network Science for
Communications Networks, San Diego.

Yahia, E., Aubry, A. & Panetto, H., 2012. Formal measures for semantic
interoperability assessment in cooperative enterprise information systems.
Computers in Industry, 63, pp.443–457.

Zavanella, 2004. Qualità degli archivi amministrativi e qualità
dell'informazione statistica. Franco Angeli, Milano.

Siti

http://www.iso.org/iso/catalogue_detail?csnumber=42180

<http://dublincore.org>

<http://www.foaf-project.org/>

[http://protegewiki.stanford.edu/wiki/Bayesian_Network_Tab_\(BNTab\)](http://protegewiki.stanford.edu/wiki/Bayesian_Network_Tab_(BNTab))

<http://protegewiki.stanford.edu/wiki/OWL2ToRDB>

<http://cp2011.istat.it>

<http://www.istat.it/it/archivio/6708>

<http://www.istat.it/it/archivio/6789>

http://www.inail.it/internet_web/appmanager/internet/home

<http://web.rete.toscana.it/RRFP/pdf/Glossario.pdf>

http://franz.com/agraph/gruff/gruff_documentation.html

<http://www.sorgatoarchitettura.it/DOCUMENTI/CLIENTI/Le%20figure%20della%20sicurezza%20da%20A.S.L.E%20RLST%20di%20Milano-Lodi.pdf>

<http://www.infoMonitor.it>

<http://ec.europa.eu/eurostat/data/sdmx-data-metadata-exchange>

Appendice 1

Metodi per l'imputazione di sistemi di classificazione in una ontologia

I Modi di creare un classificazione all'interno di un'ontologia

La creazione di una classificazione all'interno di un'ontologia è basata sul tipo di file che è usato per creare la classificazione. Sulla base di questo possiamo dividere i modi in cinque tipi:

I. La classificazione scritta a mano su carta o in file word o Pdf

Questo tipo di classificazione siamo obbligati a crearla manualmente in due modi:

- Si crea la classificazione usando il programma Protégé;
- si scrive la classificazione in linguaggio turtle dove Turtle è la sintassi per RDF che permette ad un grafico RDF di essere completamente scritto in forma di testo compatto e naturale. Si ricorda che i file Turtle hanno estensione ".ttl". Possono essere importati all'interno della nostra ontologia usando il programma Protégé.

II. La classificazione scritta su file Excel

Scriviamo la classificazione di questo tipo in linguaggio turtle in modo semi-automatico seguendo alcuni passaggi ed attivandoli sulla classificazione delle attività economiche descritte nella figura seguente.

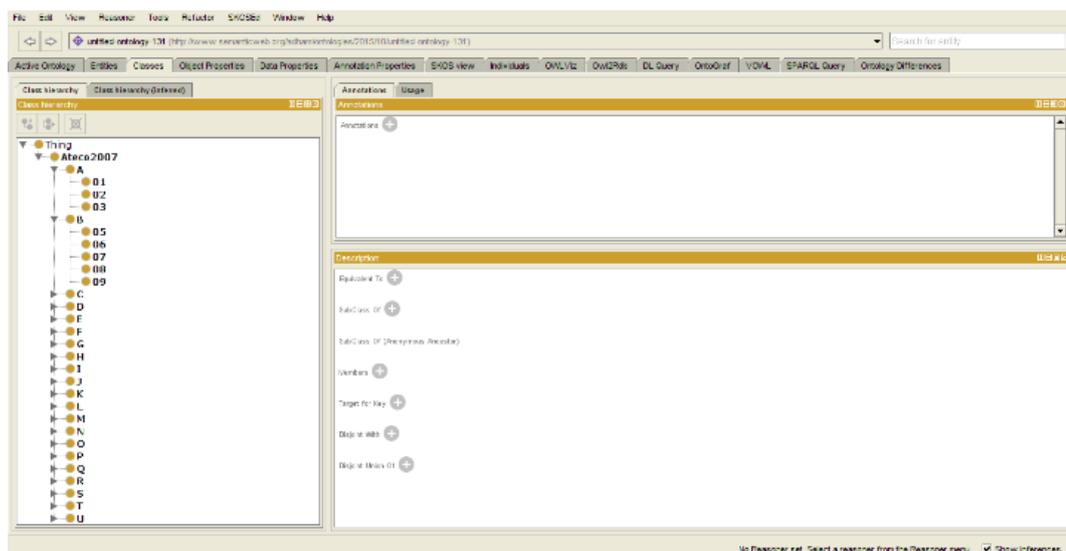
Tabella dei titoli a sei cifre della classificazione delle attività economiche Ateco 2007	
CODICE	DESCRIZIONE
A	AGRICOLTURA, SILVICOLTURA E PESCA
01	COLTIVAZIONI AGRICOLE E PRODUZIONE DI PRODOTTI ANIMALI, CACCIA E SERVIZI CONNESSI
01.11.10	Coltivazione di cereali (escluso il riso)
01.11.20	Coltivazione di semi oleosi
01.11.30	Coltivazione di legumi da granella
01.11.40	Coltivazioni miste di cereali, legumi da granella e semi oleosi
01.12.00	Coltivazione di riso
01.13.10	Coltivazione di ortaggi (inclusi i meloni) in foglia, a fusto, a frutto, in radici, bulbi e tuberi in piena aria (escluse barbabietola da zucchero e patate)
01.13.20	Coltivazione di ortaggi (inclusi i meloni) in foglia, a fusto, a frutto, in radici, bulbi e tuberi in colture protette (escluse barbabietola da zucchero e patate)
01.13.30	Coltivazione di barbabietola da zucchero
01.13.40	Coltivazione di patate
01.14.00	Coltivazione di canna da zucchero
01.15.00	Coltivazione di tabacco
01.16.00	Coltivazione di piante per la preparazione di fibre tessili
01.19.10	Coltivazione di fiori in piena aria
01.19.20	Coltivazione di fiori in colture protette
01.19.90	Coltivazione di piante da foraggio e di altre colture non permanenti
01.21.00	Coltivazione di uva
01.22.00	Coltivazione di frutta di origine tropicale e subtropicale

- Analisi della classificazione a base della creazione ontologica

la classificazione contiene una classe padre che si chiama Ateco2007 e le sue sotto classi che composte dalle linee che hanno un codice letterale; a loro volta queste sotto classi hanno le loro sotto classi composte da tutte le linee che hanno un codice numerico da due cifre. Tutte le linee che hanno un codice numerico da sei cifre sono individuali di tipo della sotto classe che ha le stesse prime due cifre del codice.

- Creare la gerarchia delle classi

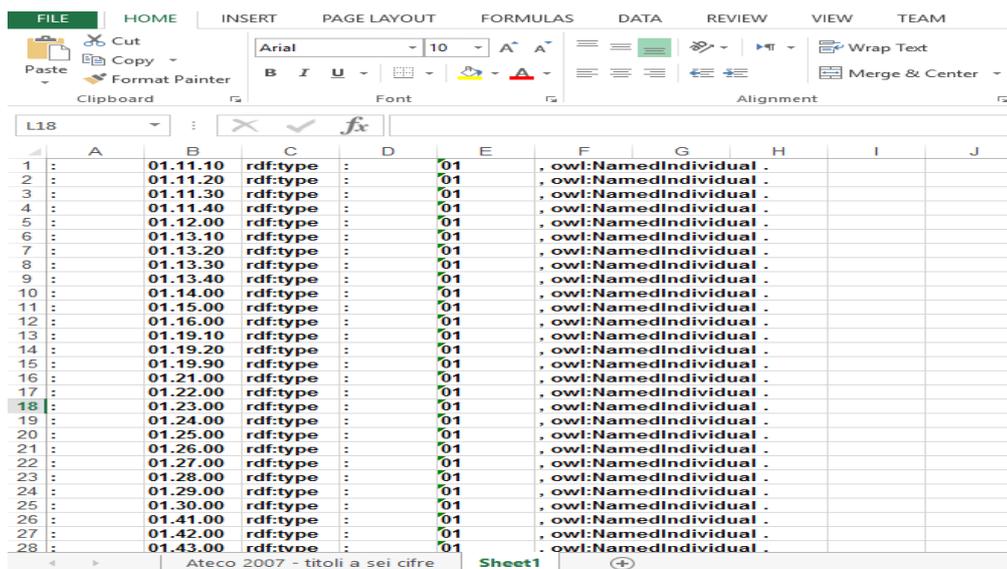
usando il programma Protégé si crea la gerarchia delle classi della classificazione come vediamo nella figura seguente.



si salva il file Protégé in formato turtle, si copia il contenuto del file e lo incolliamo nel file tipo ".ttl" vuoto.

- L'inserimento degli individuali

In un'altra pagina Excel incolliamo la colonna dei codici e creamo altre colonne²⁷ come si mostra nella figura seguente.



copiamo tutto il contenuto della pagina e lo incolliamo in modo "text only" dentro un file Word. in questo caso ogni linea ha questa forma:

```
: 95.29.02 rdf:type: 95 , owl:NamedIndividual .
```

abbiamo bisogno di fare una modifica alle linee per essere adeguata al linguaggio turtle. si fa questa modifica usando la proprietà "sostituire" che è disponibile nel programma Word. Dopo questa modifica abbiamo le linee in questa forma.

```
:95.29.02 rdf:type:95, owl:NamedIndividual .
```

A questo punto possiamo copiare il contenuto del file Word e incollarlo nel file ".ttl" precedente.

- L'aggiornamento del label per le classe e gli individuali

Questo passaggio assomiglia al passaggio precedente è per avere la forma

```
:A rdfs:label "AGRICOLTURA, SILVICOLTURA E PESCA " .
```

²⁷ abbiamo messo ":" nella colonna "A" perche il programma protege usa questo simbolo per indicare alla namespace delle sue ontologie.

:01 rdfs:label "COLTIVAZIONI AGRICOLE E PRODUZIONE DI PRODOTTI ANIMALI, CACCIA E SERVIZI CONNESSI " .

:01.11.10 rdfs:label "Coltivazione di cereali (escluso il riso) " .

dove il contenuto della colonna DESCRIZIONE è il label.

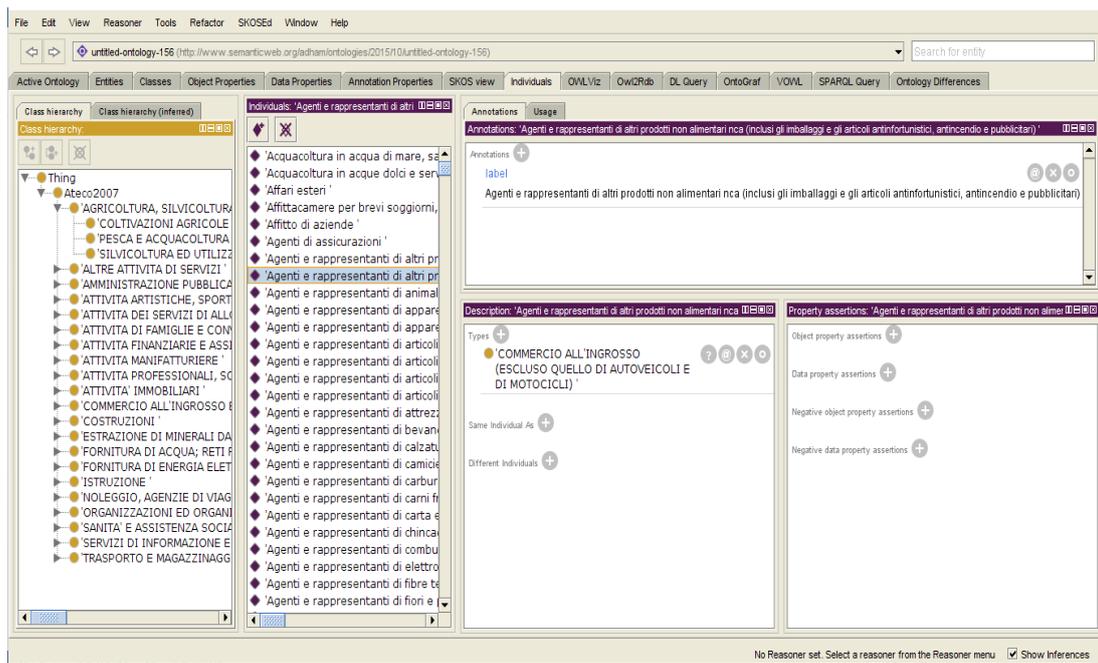
A questo punto possiamo copiare il contenuto del file Word e incollarlo nel file ".ttl" precedente.

- Convalidare i contenuti del file ".ttl"

Per sapere se il contenuto del file è adeguato completamente alla grammatica del linguaggio Turtle; si può usare "Fuseki server" che caricando il file ".ttl" ci indica la line in cui c'è il primo errore e ci informa il tipo di errore. Dopo la correzione dell'errore ricarichiamo il file; ripetiamo questo passo fino a far diventare tutto il file corretto.

- Importare il file ".ttl" nel Protégé

quando importiamo il file abbiamo l'ontologia mostrata nella figura seguente.



III. La classificazione è stata fatta come un tesoro (skos file)

Ci sono strumenti che creano questo tipo di classificazione in modo automatico, uno di questi strumenti è SKOS2OWL dove SKOS2OWL è uno strumento online che converte classificazioni gerarchiche disponibili nel W3C SKOS formato (Simple Knowledge Organization Systems) in ontologie RDF-S o OWL. In molti casi, le ontologie risultanti

possono essere usate direttamente. In caso contrario, possono essere raffinate utilizzando gli strumenti standard di costruzione di ontologie come ad esempio Protégé.

IV. La classificazione scritta su file XML:

Ci sono strumenti che creano questo tipo di classificazione in modo automatico. uno di questi strumenti è un protege plugin che si chiama XML2OWL che permette la mappatura dei nuovi dati XML per un'ontologia OWL esistente attraverso Protege. Le mappature sono svolte secondo regole fornite dall'utente, e si traducono in nuovi individui e tassonomia in ontologia.

V. La classificazione fatta come una tabella in un database

Ci sono strumenti che creano questo tipo di classificazione in modo automatico. uno di questi strumenti è DataMaster che è un Protege plugin per la creazione dello schema delle struttura e per l'importazione dei dati provenienti da database relazionali all'interno di un'ontologia.

Appendice 2

La modulazione ontologica della grande opera

modulazione è stata fatta tramite la creazione di una serie di classe che sono create a base dei concetti standard e il database Monitor che viene spiegata come il seguente:

I. Classe Ditta

Class Ditta	Sub class of	Thing
PartitaIVA	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Ditta	haCodiciAttivitaEconomica	Ateco	Only
Ditta	haCodiciTariffaInail	TariffaInail	Only
Ditta	Impiega	Lavoratori	Min 1
Ditta	comuneSedeLegale	Comuni	Only
Ditta	capSedeLegale	Cap	Only
Ditta	comuneAtti	Comuni	Only
Ditta	capAtti	Cap	Only

Dominio	Data properties	Range	Restrictions
Ditta	haDittaNome	String	Only
Ditta	haRagioneSociale	String	Only
Ditta	viaSedeLegale	String	Only
Ditta	numeroTotaleDeiDipendenti	int	Only
Ditta	viaAtti	String	Only
Ditta	haTelefono	String	Some
Ditta	haFax	String	Some
Ditta	haEmail	String	some
Ditta	haDataChiusura	Date	Only
Ditta	note	String	
Ditta	haEsercente	String	some
Ditta	haEnunce	boolean	Only

II. Classe TariffaInail

Class TariffaInail	Sub class of	Thing
codiceTariffaInail	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Ditta	haCodiciTariffaInail	TariffaInail	Some

Dominio	Data properties	Range	Restrictions
TariffaInail	haSettore	String	Only
TariffaInail	haClasse	String	Only
TariffaInail	haCategoria	string	Only

III. Classe Ateco

Class Ateco	Sub class of	Thing
codiceAteco	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Ditta	haCodiciAttivitaEconomica	Ateco	some

Dominio	Data properties	Range
Ateco	rdfs:label	string

IV. Class Cantieri

Class Cantieri	Sub class of	Thing
codiceCantiere	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Cantieri	haCantiereWBS	CantiereWBS	Some
Cantieri	haCampiBase	CampiBase	Some
Cantieri	cantieriComune	Comuni	Only
Cantieri	cantieriCap	Cap	Only
CantiereWBS	haCantieri	Cantieri	Only
CampiBase	haCantieri	Cantieri	Only
Lavoratori	siTrovaIn	Cantieri	some
Cantieri	StrutturatiIn	Fasi costruttive	Min 1
Opera	Contiene	cantieri	Min 1

Dominio	Data properties	Range	Restrictions
Cantieri	nomeCantieri	String	Only
Cantieri	cantieriIndirizzo	String	Only
Cantieri	cantierizone	String	Only
Cantieri	haCoordinataGaussBoaga_X	String	Only
Cantieri	haCoordinataGaussBoaga_Y	String	Only
Cantieri	LatitudineWGS84	String	Only
Cantieri	Longitudine WGS84	String	Only

V. Classe CantiereWBS

Class CantiereWBS	Sub class of	Thing
codiceCantiereWBS	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
CantiereWBS	haCantieri	Cantieri	Only

Dominio	Data properties	Range	Restrictions
CantiereWBS	nomeCantieriWBS	String	Only
CantiereWBS	haFronte	String	Only
CantiereWBS	haGalleria	String	Only

VI. Class CampiBase

Class CampiBase	Sub class of	Thing
codiceCampiBase	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
CampiBase	haCantieri	Cantieri	Only
CampiBase	campiBaseComune	Comuni	Only
CampiBase	campiBaseCap	Cap	Only

Dominio	Data properties	Range	Restrictions
CampiBase	codiceCampoBaseRif	String	Only
CampiBase	campiBaseIndirizzo	String	Only

VII. Class InchiestaInfortuni

Class InchiestaInfortuni	Sub class of	Thing
CodInchiesta	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
InchiestaInfortuni	causatoUnaMalattia	Malattia	Some
InchiestaInfortuni	haSchedaInfortuni	SchedaInfortuni	Some
InchiestaInfortuni	personaliDellaSicurezzaPresenti	Lavoratori	some

Dominio	Data properties	Range	Restrictions
InchiestaInfortuni	Cod_atto	String	Only
InchiestaInfortuni	haInchiestaDiInfortunio	boolean	Only

InchiestaInfortuni	dataInfortunio	data	Only
InchiestaInfortuni	haCompetenzaCompiti	Boolean	Only
InchiestaInfortuni	altriInfortunati	Boolean	Only
InchiestaInfortuni	sospesoDalLavoro	Boolean	Only
InchiestaInfortuni	DurataDiSospensione	String	Only
InchiestaInfortuni	dataDelritornoAllavoro	Date	Only
InchiestaInfortuni	scaturitaUn'indagine	Boolean	Only
InchiestaInfortuni	dataInizioIndagine	Date	Only
InchiestaInfortuni	dataFineIndagine	Date	Only
InchiestaInfortuni	sonoPresentiDeiCollegli	Boolean	Only
InchiestaInfortuni	sonoPresentiDellePersonalidellaSicurezza	Boolean	Only
InchiestaInfortuni	sonoPresentiDeiServiziDiEmergenzaSanitaria	Boolean	Only
InchiestaInfortuni	durataTraInfortunioESoccorsoRicevuto	String	Only
InchiestaInfortuni	portatoIN	String	Some
InchiestaInfortuni	primaDiagnosi	String	Only
InchiestaInfortuni	previsioneDellaDurataDellaMalattia	String	Max 1
InchiestaInfortuni	definitivaDellaDurataDellaMalattia	String	Only
InchiestaInfortuni	causatoLaMorte	Boolean	Only
InchiestaInfortuni	lesioniPermanenti	Boolean	Only
InchiestaInfortuni	durataDiAssenzaDalLavoro	String	Only
InchiestaInfortuni	dinamicaDiInfortunio	String	Only
InchiestaInfortuni	causeDiInfortunio	String	Only
InchiestaInfortuni	stataChiusaInchiesta	Boolean	Only
InchiestaInfortuni	violazioniDiCodicePenale	Boolean	Only
InchiestaInfortuni	descrizioneDelleViolazioni	String	Only

VIII. Class SchedaInfortuni

Class SchedaInfortuni	Sub class of	Thing
CodiceInfortunio	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
SchedaInfortuni	haInchiestaInfortuni	InchiestaInfortuni	Only
SchedaInfortuni	haModalitaInfortunio	ModalitaInfortunio	Only

SchedaInfortuni	haAgenteMateriale	AgenteMateriale	Some
SchedaInfortuni	haSedeLesione	SedeLesione	Some
SchedaInfortuni	haNaturaLesione	NaturaLesione	Some
SchedaInfortuni	haTipologiaDiLavoro	TipologiaDiLavoro	Only

Dominio	Data properties	Range	Restrictions
SchedaInfortuni	Inabilita	Boolean	Only
SchedaInfortuni	accadutoInItinere	Boolean	Only
SchedaInfortuni	informazioneContieneLaFonteInfermeria	Boolean	Only
SchedaInfortuni	informazioneContieneLaFonteDitta	Boolean	Only
SchedaInfortuni	informazioneContieneLaFonteUSL	Boolean	Only
SchedaInfortuni	schedaCompleta	Boolean	Only
SchedaInfortuni	Fonte	String	Some

IX. Class ModalitaInfortunio

Class ModalitaInfortunio	Sub class of	Thing
codiceModalitaInfortunio	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
SchedaInfortuni	haModalitaInfortunio	ModalitaInfortunio	Only

Dominio	Data properties	Range
ModalitaInfortunio	rdfs:label	String

Individual of Class ModalitaInfortunio		
Ambiente	rdfs:label	Ambiente
	rdfs:commento	infortunio causato dall' ambientale (es: se c'è stata una frana)
Attivo	rdfs:label	Attivo
	rdfs:commento	se il soggetto infortunato è stato lui stesso a causarlo
Cadute	rdfs:label	Cadute
	rdfs:commento	se l'infortunio è stato a causa di una caduta
Incidenti	rdfs:label	Incidenti
	rdfs:commento	se l'infortunio è stato a causa di un incidente.
Passivo	rdfs:label	Passivo
	rdfs:commento	se ha subito l'infortunio e non è stato il soggetto infortunato a causarlo

X. Class AgenteMateriale

Class AgenteMateriale	Sub class of	Thing
codiceAgenteMateriale	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
SchedaInfortuni	haAgenteMateriale	AgenteMateriale	Only

Dominio	Data properties	Range
AgenteMateriale	rdfs:label	String

XI. Class SedeLesione

Class SedeLesione	Sub class of	Thing
codiceAgenteMateriale	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
SchedaInfortuni	haSedeLesione	SedeLesione	Only

Dominio	Data properties	Range
SedeLesione	rdfs:label	String

XII. Class NaturaLesione

Class NaturaLesione	Sub class of	Thing
codiceNaturaLesione	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
SchedaInfortuni	haNaturaLesione	NaturaLesione	Only

Dominio	Data properties	Range
NaturaLesione	rdfs:label	String

XIII. Class TipologiaDiLavoro

Class AgenteMateriale	Sub class of	Thing
codiceTipologiaDiLavoro	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
SchedaInfortuni	haTipologiaDiLavoro	TipologiaDiLavoro	Only

Dominio	Data properties	Range
TipologiaDiLavoro	rdfs:label	String

XIV. Class Sopralluogo

Class Sopralluogo	Sub class of	Thing
codiceSopralluogo	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Sopralluogo	HaCantieri	Cantieri	Only
Sopralluogo	haOpera	Opera	Only
Sopralluogo	haTipologiaDiLavoro	TipologiaDiLavoro	Only
Sopralluogo	haFaseCostruttiva	FaseCostruttiva	Some
Sopralluogo	haMotiviDiSopralluogo	MotivazioneSopralluogo	Only
Sopralluogo	haCommittente	Lavoratori	Only
Sopralluogo	haResponsabileDeiLavori	Lavoratori	Only
Sopralluogo	haCoordinatorePerLaProgettazione	Lavoratori	Only
Sopralluogo	haCoordinatorePerL'esecuzione	Lavoratori	Only
Sopralluogo	haDirettoreTecnicoDiCantiere	Lavoratori	Only
Sopralluogo	haPersonePresentiAlSopralluogo	Lavoratori	Some
Sopralluogo	haInchiestaInfortuonio	InchiestaInfortuonio	Some
Sopralluogo	haVerificatoUnOggetto	OggettiDiSopralluogo	Only
Sopralluogo	haEsecutoreDiSopralluogo	Operatori	Some
Sopralluogo	Esamina	Lavoratori	Some
Sopralluogo	Esamina	Ditta	Some
Sopralluogo	Redige	Verbali	only

Dominio	data properties	Range	Restrictions
Sopralluogo	dataDelSopralluogo	Data	Only
Sopralluogo	OraInizio	String	Only
Sopralluogo	OraFine	string	Only
Sopralluogo	USL	Boolean	Only
Sopralluogo	haAltroOperatore	Boolean	Only
Sopralluogo	haSopralluogoIniziativa	Boolean	Only
Sopralluogo	haInchiestaMalattiaProfessionale	string	Some
Sopralluogo	NumeroDiFogliCompilati	int	Only
Sopralluogo	DataCompilazioneDelVerbale	Data	Only
Sopralluogo	DataRegistrazioneDelVerbale	Data	Only

XV. Class MotivazioneSopralluogo

Class Sopralluogo	Sub class of	Thing
codiceMotivazioneSopralluogo	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Sopralluogo	haMotiviDiSopralluogo	MotivazioneSopralluogo	only

Individual		
SopralluogoIniziativa	rdfs:label	Sopralluogo iniziativa
SopralluogoInchiestaInfortunio	rdfs:label	Sopralluogo inchiesta infortunio
InchiestaMalattiaProfessionale	rdfs:label	Inchiesta malattia professionale
VerificheImpiantiMacchine	rdfs:label	Verifiche impianti macchine
IndaginiAmbientali	rdfs:label	Indagini ambientali
SuSegnalazioniDiProblemi	rdfs:label	Su segnalazioni di problemi
VerificaDiAdempimenti	rdfs:label	Verifica di adempimenti

XVI. Class OggettiDiSopralluogo

Class OggettiDiSopralluogo	Sub class of	Thing
codiceOggettiDiSopralluogo	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Sopralluogo	haVerificatoUnOggetto	OggettiDiSopralluogo	Only

Individual		
ImpiantiMacchine	rdfs:label	Impianti macchine
	rdfs:commento	è indagine sul luogo per verificare il funzionamento e la presenza degli impianti tecnologici presenti sul cantiere.
ImpiantiElettrici	rdfs:label	Impianti elettrici
	rdfs:commento	è indagine sul luogo per verificare il funzionamento e la presenza dell'insieme di apparecchiature elettriche, meccaniche e fisiche atte alla trasmissione e all'utilizzo di energia elettrica. La figura professionale che si occupa della realizzazione e manutenzione di tali impianti è quella dell'elettricista. Esistono due grandi categorie di impianti elettrici: impianti civili e impianti industriali. I primi si utilizzano nelle abitazioni private e nei luoghi di pubblico accesso come scuole e ospedali;

		i secondi si utilizzano nei luoghi di lavoro e di produzione e sono spesso utilizzati per movimentare e automatizzare le "macchine" nelle industrie.
ApparecchiDiPressione	rdfs:label	Apparecchi di pressione
	rdfs:commento	<p>è indagine sul luogo per verificare il funzionamento e la presenza di apparecchiature in pressione soggette alla PED (Direttiva apparecchi a pressione) le seguenti singole attrezzature e insiemi da queste composti:</p> <ul style="list-style-type: none"> • i recipienti: alloggiamenti progettati e costruiti per contenere fluidi pressurizzati quali compressori, autoclavi, condensatori, recipienti a gas o a vapore, reattori, scambiatori, sfere GPL, ecc. • tubazioni intese come tubo o insieme di tubi in pressione destinati al trasporto dei fluidi compresi gli eventuali componenti sottoposti a pressione quali giunti di smontaggio, giunti di dilatazione, flangie, raccordi, ecc.. Non sono comprese ad esempio le condotte idriche, per petrolio o gas (si veda punti seguenti); • accessori in pressione: le valvole idrauliche come le saracinesche, le valvole a farfalla, le valvole a fuso, gli sfiati, le valvole di non ritorno, ecc. • accessori di sicurezza: i dispositivi destinati alla protezione delle attrezzature in pressione contro il superamento dei limiti ammissibili; questi comprendono: <ul style="list-style-type: none"> • dispositivi di limitazione diretta della pressione: valvole di sicurezza, dispositivi a disco di rottura, aste pieghevoli, dispositivi di sicurezza pilotati per lo scarico pressione (CSPRS);

		<ul style="list-style-type: none"> • dispositivi di limitazione che attivano i sistemi di regolazione o che chiudono e disattivano l'attrezzatura: pressostati, termostati, interruttori di livello del fluido, dispositivi di misurazione, controllo e regolazione per la sicurezza (SRMCR); • insiemi: sono costituiti da varie attrezzature in pressione assemblati da un fabbricante per costituire un tutto integrato e funzionale.
Indagine Ambientale	rdfs:label	Indagine ambientale
	rdfs:commento	è una procedura amministrativa di supporto per l'autorità competente [come Ministero dell'Ambiente o Regione] finalizzata ad individuare, descrivere e valutare gli impatti ambientali di un'opera, il cui progetto è sottoposto ad approvazione o autorizzazione.
Rumore	rdfs:label	Rumore
	rdfs:commento	Serve a rilevare quanto è eccessiva l'esposizione a suoni e rumori di elevata intensità. Il rumore è un segnale di disturbo rispetto all'informazione trasmessa in un sistema.
Illuminazione	rdfs:label	Illuminazione
	rdfs:commento	Ispezione, indagine sul luogo per verificare se l'utilizzo di flussi luminosi, naturali (mediati da elementi architettonici) o emessi da sorgenti artificiali (apparecchiature generalmente elettriche) riescono ad ottenere determinati livelli di luce (illuminamenti) sull'oggetto (in senso lato) da illuminare.
Microclima	rdfs:label	Microclima
	rdfs:commento	indagine sul luogo per verificare se clima di una zona geografica locale in cui i parametri atmosferici medi differiscono in modo caratteristico e significativo da quelli delle zone circostanti a causa di peculiarità topografiche, orografiche, geomorfologiche e ambientali.
Impianti Di Ventilazione	rdfs:label	Impianti di ventilazione
	rdfs:commento	serve a verificare se i sistemi di ventilazione controllata o ventilazione meccanica controllata (VMC) consentono di gestire il ricambio dell'aria di un ambiente con l'esterno. Questo avviene senza

		l'apertura di finestre o porte, tramite condotte di ventilazione forzata, collegate con gli ambienti interni da aspiratori (per la rimozione dell'aria viziata o inquinata) e da diffusori (per l'immissione di aria nuova).
PrelievoDegliInquinanti	rdfs:label	Prelievo degli inquinanti
	rdfs:commento	Specifica se sono stati prelevate sostenze inquinanti cioè sono sostanze che, direttamente o indirettamente, costituiscono un pericolo per la salute dell'uomo o per l'ambiente, provocando alterazioni delle risorse biologiche e dell'ecosistema.
FoglioDiPrescrizione	rdfs:label	Foglio di Prescrizione
	rdfs:commento	è un foglio in cui vengono descritte ed elencate una serie di indicazioni dopo aver svolto il sopralluogo.

XVII. Class Verbali

Class Verbali	Sub class of	Thing
codiceVerbali	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Verbali	haTipoDiVerbali	TipoDiVerbali	Only
Verbali	haPresentatore	Presentatore	Min 1
Verbali	haContravvenzione	Contravvenzione	some

Dominio	Data properties	Range	Restrictions
Verbali	haNumeroDiVerbali	String	Only
Thing	Note	String	
Verbali	haParoleChiave	String	Some
Verbali	haDataDiVerbale	Date	Only
Verbali	haDataDiRicezioneAtto	Date	Only
Verbali	haDataDiAttoEmessoFatto	Date	Only
Verbali	haAttoProvvedimenti	String	Some
Verbali	consegnaVerbale	String	Only

XVIII. Class TipoDiVerbali

Class TipoDiVerbali	Sub class of	Thing
TipoDiVerbali	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Verbali	haTipoDiVerbali	TipoDiVerbali	only

Individual		
Disposizione	rdfs:label	Disposizione
	rdfs:commento	La disposizione è un atto con cui l'organo di vigilanza impartisce, sulla base di un apprezzamento discrezionale, un'ingiunzione al datore di lavoro di adottare le misure prevenzionali ad integrazione e specificazione dell'obbligo generico contenuto in una norma primaria in bianco o di portata molto ampia ed indeterminata.
Prescrittivo	rdfs:label	Prescrittivo
	rdfs:commento	prescrittivo è un atto con cui l'organo di vigilanza stabilisce e prescrive qualcosa al datore di lavoro, il quale è tenuto ad osservare.
RichiamoAllaNorma	rdfs:label	Richiamo alla norma
	rdfs:commento	è un atto con cui l'organo di vigilanza accerta una violazione facendo riferimento ad una norma.

XIX. Class Presentatori

Class Persona	Sub class of	Thing
Class Persona	Equivalent to	Foaf:Person
Class Presentatori	Sub class of	Class persona
codicePresentatori	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Presentatori	haQualifica	Qualifica	Only
Qualifica	qualificaSonoStatiOccupatiDa	Presentatori	Some
Verbali	haPresentatore	Presentatore	Min 1

Dominio	Data properties	Range	Restrictions
Persona	haCognome sameAs Foaf:familyname	String	Only
Persona	haNome sameAs Foaf:firstname	String	Only

XX. Class Contravvenzione

Class Contravvenzione	Sub class of	Thing
CodiceContravvenzione	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Contravvenzione	haContravventore	Lavoratori	Only
Verbali	haContravvenzione	Contravvenzione	some
Contravvenzione	haResponsabileDiSicurezza	Lavoratori	Only
Contravvenzione	haRappresentanteLegaleDiDitta	Lavoratori	Only

Dominio	Data properties	Range	Restrictions
Contravvenzione	haPresentatoIlContravventore	Boolean	Only
Contravvenzione	haDataDiContravvenzione	Date	Only
Contravvenzione	haCusatoSequestro	Boolean	Only
Contravvenzione	dataDiInizioSequestro	Date	Only
Contravvenzione	dataDiFineSequestro	Date	Only
Contravvenzione	haCusatoDivieto	Boolean	Only
Contravvenzione	dataDiInizioDivieto	Date	Only
Contravvenzione	dataDiFineDivieto	Date	Only

XXI. Class FasiCostruttiva

Class FasiCostruttiva	Sub class of	Thing
CodiceFasiCostruttiva	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Sopralluogo	haFasiCostruttiva	FasiCostruttiva	Some
Cantieri	Strutturati in	Fasi costruttive	Min 1
Fasi costruttive	Rappresentano	ADA	Min 1

Dominio	Data properties	Range
FasiCostruttiva	rdfs:label	string

Individual		
F01	rdfs:label	Preconsolidamento
F02	rdfs:label	Scavo del fronte
F03	rdfs:label	Smarino
F04	rdfs:label	trasporto materiale di scavo a scarica e sistemazione

F05	rdfs:label	Prerivestimento
F06	rdfs:label	scavo arco rovescio
F07	rdfs:label	getto arco rovescio
F08	rdfs:label	impermeabilizzazione
F09	rdfs:label	getto calotta
F10	rdfs:label	getto piedritti
F11	rdfs:label	Manutenzione
F12	rdfs:label	installazione e collaudo impianti
F13	rdfs:label	collaudi civili
F14	rdfs:label	consolidamento esterne
F15	rdfs:label	dima di attacco
F16	rdfs:label	allestimento cantiere
F17	rdfs:label	campo base
F18	rdfs:label	Sbancamento
F19	rdfs:label	fondazioni su pali in cls
F20	rdfs:label	bonifica ordigni bellici
F21	rdfs:label	rilievi e Monitoraggi aziendali
F22	rdfs:label	Costruzione strade
F23	rdfs:label	costruzione piazzali
F24	rdfs:label	Rilevato
F25	rdfs:label	Terrapieno
F26	rdfs:label	demolizioni opere esistenti
F27	rdfs:label	Pila
F28	rdfs:label	Spalla
F29	rdfs:label	plinto per pila
F30	rdfs:label	plinto per spalla
F31	rdfs:label	Pulvino
F32	rdfs:label	trave viadotto costruita in opera
F33	rdfs:label	trave varata
F34	rdfs:label	posata con gru
F35	rdfs:label	ponte di ferro ad una campata
F36	rdfs:label	allestimento armamento ferroviario
F37	rdfs:label	allestimento impianto di alimentazione elettrica per linea ferroviaria
F38	rdfs:label	galleria artificiale pre-scavata

F39	rdfs:label	galleria artificiale post-scavata
F40	rdfs:label	Tombino
F41	rdfs:label	sottopasso a cielo aperto
F42	rdfs:label	Nicchia
F43	rdfs:label	rivestimento con fresa
F44	rdfs:label	movimentazione materiale
F45	rdfs:label	Scotico
F46	rdfs:label	ripristino area di cantiere
F47	rdfs:label	lavori di falegnameria per armatura
F48	rdfs:label	ponte metallico
F49	rdfs:label	Trincea
F50	rdfs:label	edificio in superficie
F51	rdfs:label	sistemazione idraulica
F52	rdfs:label	muro di spinta
F53	rdfs:label	Muro di contenimento
F54	rdfs:label	Monolite
F55	rdfs:label	messa in opera del monolite
F56	rdfs:label	soletta ponte
F57	rdfs:label	prospezione archeologica
F58	rdfs:label	scatolare prescavato
F59	rdfs:label	servizi interrati
F60	rdfs:label	servizi esterni
F61	rdfs:label	spostamento servizi interferenti
F62	rdfs:label	edificio in superficie a servizio della linea
F63	rdfs:label	edifici in sotterraneo
F64	rdfs:label	Montaggio fresa
F65	rdfs:label	smontaggio fresa
F00	rdfs:label	altro

XXII. Class Opera

Class Opera	Sub class of	Thing
CodiceOpera	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Sopralluogo	haOpera	Opera	Only

Opera	Contiene	cantieri	Min 1
-------	----------	----------	-------

Dominio	Data properties	Range
Opera	rdfs:label	String
Opera	rdfs:commento	string

XXIII. Class ADA (Aria Di Attivita)

Class ADA	Sub class of	Thing
CodiceADA	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
ADA	capacitaRichiesta	Capacita	Some
ADA	conoscenzeRichiesta	Conoscenze	some

Dominio	Data properties	Range	Restrictions
ADA	rdfs:label	string	Only
ADA	Indicatori	Literal	Only
ADA	Osservazione	Literal	Only
ADA	Risultato	Literal	Only

XXIV. Capacita

Class Capacita	Sub class of	Thing
CodiceCapacita	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
ADA	capacitaRichiesta	Capacita	Some

Dominio	Data properties	Range
Capacita	rdfs:label	String

XXV. Conoscenze

Class Conoscenze	Sub class of	Thing
CodiceConoscenze	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
ADA	conoscenzeRichiesta	Conoscenze	Some

Dominio	Data properties	Range
ADA	rdfs:label	String

XXVI. Ativitta

Class Ativitta	Sub class of	Thing
CodiceAtivitta	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Ativitta	daUnProdottoFinitoObbligatorie	ProdttoFinito	Some
Ativitta	daUnProdottoFinitoPossibile	ProdttoFinito	Some

Dominio	Data properties	Range
Ativitta	rdfs:label	string

XXVII. ProdottoFinito

Class ProdttoFinito	Sub class of	Thing
CodiceAtivitta	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Ativitta	daUnProdottoFinitoObbligatorie	ProdttoFinito	Some
Ativitta	daUnProdottoFinitoPossibile	ProdttoFinito	Some
ProdttoFinito	haSettore	Settore	Only

Dominio	Data properties	Range
ProdttoFinito	rdfs:label	string

XXVIII. Settore

Class Ativitta	Sub class of	Thing
CodiceAtivitta	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
ProdttoFinito	haSettore	Settore	Only

Dominio	Data properties	Range
ProdttoFinito	rdfs:label	String

XXIX. Class ASL

Class ASL	Sub class of	Thing
DipartimentoPrevenzione	Sub class of	ASL
CodiceDipartimento	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
DipartimentoPrevenzione	Effettua	AttivitaEseguita	Some
DipartimentoPrevenzione	DipartimentoCap	cap	Only
DipartimentoPrevenzione	DipartimentoComune	comune	Only

Dominio	Data properties	Range	Restrictions
DipartimentoPrevenzione	NomeDipartimento	String	Only
DipartimentoPrevenzione	ViaDipartimento	String	Only

XXX. Class AttivitaEseguita

AttivitaEseguita	Sub class of	Thing
CodiceAttivitaEseguita	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
AttivitaEseguita	Identifica	Attivita	Only
AttivitaEseguita	haDestinatorio	Lavoratori	Some
AttivitaEseguita	haDestinatorio	Ditta	Some
AttivitaEseguita	Genera	ProdottoFinitoRisultato	Some
AttivitaEseguita	haStato	Stato	Only
AttivitaEseguita	FirmaDa	Operatori	Only
AttivitaEseguita	haDistretto	Distretto	Only

Dominio	Data properties	Range	Restrictions
AttivitaEseguita	DataDiAttivita	Data	Only
AttivitaEseguita	DataFirma	Data	Only
AttivitaEseguita	haZona	String	Only

XXXI. Class ProdottiFinitiRisultati

Class ProdottiFinitiRisultati	Sub class of	Thing
CodiceProdottiFinitiRisultati	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
ProdottiFinitiRisultati	Identifica	ProdottiFiniti	Min 1

ProdottiFinitiRisultati	haStato	Stato	Min 1
ProdottiFinitiRisultati	FirmaDa	Operatori	Only
ProdottiFinitiRisultati	haAttivitaEseguita	AttivitaEseguita	Only

Dominio	Data properties	Range	Restrictions
ProdottiFinitiRisultati	DataInizio	Data	Only
ProdottiFinitiRisultati	DataFine	Data	Only
ProdottiFinitiRisultati	DataFirma	Data	Only

XXXII. Class Operatori

Persona	Sub class of	Thing
Class Operatori	Sub class of	Persona
CodiceOperatori	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Operatori	LavoraIn	DipartimentoPrevenzione	Only
Operatori	Effettua	Sopralluogo	Some

Dominio	Data properties	Range	Restrictions
Operatori	Nome	String	Only
Operatori	Cognome	String	Only

XXXIII. Class Stato

Class Stato	Sub class of	Thing
CodiceStato	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
AttivitaEseguita	haStato	Stato	Only
ProdottiFinitiRisultati	haStato	Stato	Only

Individuals		
Incorso	rdfs:label	Incorso
	rdfs:commento	Assume valore INCORSO quando la data fine non è valorizzata
Chiuso	rdfs:label	Chiuso
	rdfs:commento	Assume valore CHIUSO quando la data fine è valorizzata

Nullo	rdfs:label	Nullo
	rdfs:commento	Assume valore NULLO quando la pratica viene annullata

XXXIV. Class Distretto

Class Distretto	Sub class of	Thing
CodiceDistretto	Rdf:type	owl:NamedIndividual

Dominio	Object properties	Range	Restrictions
Distretto	ContieneComuni	Comuni	Min 1

Dominio	Data properties	Range	Restrictions
Distretto	DistrettoNome	string	Only

Appendice 3

Il database Monitor

Tabella Cantieri		
Nome della colonna	Tipo della data	Nota
CodiceCantieri	String	Primary Key
nomeCantieri	String	
cantieriVia	String	
CantieriCap	int	
CantieriComune	String	
CantieriComuneCodice	String	Codice comune dell'ISTAT
CantieriProvincia	String	
CantieriRegione	String	
Cantierizone	String	
CodiceOpera	String	foreign key
CoordinataGaussBoaga_X	String	
CoordinataGaussBoaga_Y	String	
LatitudineWGS84	String	
LongitudineWGS84	String	

Tabella CantieriWBS		
Nome della colonna	Tipo della data	Nota
CodiceCantieriWBS	String	Primary Key
CodiceCantieri	String	foreign key
nomeCantieriWBS	String	
Fronte	String	
Galleria	String	

Tabella CampiBase		
Nome della colonna	Tipo della data	Nota
CodiceCampiBase	String	Primary Key
CodiceCantieri	String	Foreign key
campiBaseVia	String	
campiBaseCap	Int	
campiBaseComune	String	
campiBaseComunecodice	String	Codice comune dell'ISTAT
CampoBaseProvincia	String	
CampoBaseRegione	String	
codiceCampoBaseRif	String	

Tabella Ditta		
Nome della colonna	Tipo della data	Nota
PartitaIVA	String	Primary Key
CodiciAttivitaEconomica	String	Atico2007
CodiciTariffaInail	String	Inail
viaSedeLegale	String	
capSedeLegale	int	
comuneSedeLegale	String	
ComuneCodice	String	Codice comune dell'ISTAT
ProvinciaSedeLegale	String	
RegioneSedeLegale	String	
ViaAtti	String	
capAtti	int	
comuneAtti	String	
comuneAttiCodice	String	Codice comune dell'ISTAT
ProvinciaAtti	String	
RegioneAtti	String	

Tabella Malattia		
Nome della colonna	Tipo della data	Nota
ID	Autonumber	Primary key
Malattia	String	
CodiceMalattia	String	classificazione internazionale delle malattie
CFLavoratore	String	Foreign key
CodInchiesta	String	Foreign key

Tabella InchiestaInfortuni		
Nome della colonna	Tipo della data	Nota
CodInchiesta	String	Primary key
CFLavoratore	String	Foreign key
Cod_atto	String	
InchiestaDiInfortunio	boolean	
dataInfortunio	data	
CompetenzaCompiti	Boolean	
altriInfortunati	Boolean	
sospesoDalLavoro	Boolean	
DurataDiSospensione	String	
dataDelritornoAlLavoro	Date	
scaturitaUn'indagine	Boolean	
dataInizioIndagine	Date	
dataFineIndagine	Date	
PresentiDeiColleghi	Boolean	

PresentiDellePersonalitàDellaSicurezza	Boolean	
PresentiDeiServiziDiEmergenzaSanitaria	Boolean	
durataTraInfortunioESoccorsoRicevuto	String	
portatoIN	String	
primaDiagnosi	String	
previsioneDellaDurataDellaMalattia	String	
definitivaDellaDurataDellaMalattia	String	
causatoLaMorte	Boolean	
lesioniPermanenti	Boolean	
durataDiAssenzaDalLavoro	String	
dinamicaDiInfortunio	String	
causeDiInfortunio	String	
stataChiusaInchiesta	Boolean	
violazioniDiCodicePenale	Boolean	
descrizioneDelleViolazioni	String	

Tabella SchedaInfortuni		
Nome della colonna	Tipo della data	Nota
CodiceInfortunio	String	Primary key
CodInchiesta	String	Foreign key
CFLavoratore	String	Foreign key
codiceModalitàInfortunio	String	Foreign key
CodiceAgenteMateriale	String	Foreign key
CodiceTipologiaDiLavoro	String	Foreign key
Inabilita	Boolean	
accadutoInItinere	Boolean	
informazioneContieneLaFonteInfermeria	Boolean	
informazioneContieneLaFonteDitta	Boolean	
informazioneContieneLaFonteUSL	Boolean	
schedaCompleta	Boolean	
Fonte	String	

Tabella ModalitàInfortunio		
Nome della colonna	Tipo della data	Nota
codiceModalitàInfortunio	String	Primary key
ModalitàInfortunio	String	

Tabella AgenteMateriale		
Nome della colonna	Tipo della data	Nota
CodiceAgenteMateriale	String	Primary key, European Statistics on Accidents at Work (ESAW)
AgenteMateriale	String	

Tabella SedeLesione		
Nome della colonna	Tipo della data	Nota
codiceSedeLesione	String	Primary key, INAIL
SedeLesione	String	

Tabella NaturaLesione		
Nome della colonna	Tipo della data	Nota
codiceNaturaLesione	String	Primary key, INAIL
NaturaLesione	String	

Tabella SedeNaturaLesioneESchedaInfortuni		
Nome della colonna	Tipo della data	Nota
ID	Autonumber	Primary key
SchedaInfortuni	String	Foreign key
codiceSedeLesione	String	Foreign key
codiceNaturaLesione	String	Foreign key

Tabella TipologiaDiLavoro		
Nome della colonna	Tipo della data	Nota
CodiceTipologiaDiLavoro	String	Primary key, European Statistics on Accidents at Work (ESAW)
TipologiaDiLavoro	String	

Tabella Lavoratori		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
IdLavoratori	String	
Nome	String	
Cognome	String	
dataDiNascita	data	
CitaDiNascita	String	
codiceCitaDiNascita	String	Codice ISTAT
CodiceRuoloSicurezza	String	Foreign key

Professione	String	
CodiceLivelloProfessione	String	classificazione delle professioni 2011
Istruzione	String	
CodiceLivelloIstruzione	String	classificazione internazionale dell'istruzione (ISCED)
Qualifica	String	un data list di quattro scelte (Dirigenti, Quadri, Impiegati e Operai)
Sesso	String	un data list di due scelte (maschile e femminile)
ViaDIAbitazione	String	
CapDIAbitazione	Int	
cittàDIAbitazione	String	
codicecittàDIAbitazione	String	Codice ISTAT
ProvinciaDIAbitazione	String	
RegioneDIAbitazione	String	
ViaDIResidenza	String	
CapDIResidenza	Int	
cittàDIResidenza	String	
codicecittàDIResidenza	String	Codice ISTAT
ProvinciaDIResidenza	String	
RegioneDIResidenza	String	
StatoCivile	String	un data list di otto scelte (Celibe, Nubile, Sposato, Sposata, Divorziato, Divorziata, Vedovo e Vedova)
Alloggio	Boolean	
Note	String	

Tabella RuoloSicurezza		
Nome della colonna	Tipo della data	Nota
CodiceRuoloSicurezza	String	Primary key
RuoloSicurezza	String	

Tabella DittaLavoratori		
Nome della colonna	Tipo della data	Nota
PartitaIVA	String	Primary key
CFLavoratore	String	Primary key

Tabella Sopralluogo		
Nome della colonna	Tipo della data	Nota
codiceSopralluogo	String	Primary key
CodiceCantiere	String	Foreign key
CodiceOpera	String	Foreign key
CodiceTipologiaDiLavoro	String	Foreign key
CodiceFaseCostruttiva	String	Foreign key
CodiceMotivazioneSopralluogo	String	Foreign key
CodInchiesta	String	Foreign key
OggettiDiSopralluogo	String	Questa colonna contiene lista di dieci righe (ImpiantiMacchine, ImpiantiElettrici, ApparecchiDiPressione, IndagineAmbientale, Rumore, Illuminazione, Mircoclima, ImpiantiDiVentilazione, PrelievoDegliInquinanti e FoglioDiPrescrizione)
dataDelSopralluogo	Data	
OraInizio	String	
OraFine	string	
USL	Boolean	
AltroOperatore	Boolean	
haSopralluogoIniziativa	Boolean	
InchiestaMalattiaProfessionale	string	
NumeroDiFogliCompilati	Numero	
DataCompilazioneDelVerbale	Data	
DataRegistrazioneDelVerbale	Data	

Tabella FaseCostruttiva		
Nome della colonna	Tipo della data	Nota
CodiceFaseCostruttiva	String	Primary key
FaseCostruttiva	String	

Tabella Opera		
Nome della colonna	Tipo della data	Nota
CodiceOpera	String	Primary key
NomeDellaOpera	String	

Tabella TipoDellaOpera		
Nome della colonna	Tipo della data	Nota
CodiceTipoDellaOpera	String	Primary key

TipoDellaOpera	String	
----------------	--------	--

Tabella OperaETipoDellaOpera		
Nome della colonna	Tipo della data	Nota
CodiceOpera	String	Primary key
CodiceTipoDellaOpera	String	Primary key

Tabella MotivazioneSopralluogo		
Nome della colonna	Tipo della data	Nota
CodiceMotivazioneSopralluogo	String	Primary key
MotivazioneSopralluogo	String	

Tabella Committente		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
codiceSopralluogo	String	Primary key

Tabella ResponsabileDeiLavori		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
codiceSopralluogo	String	Primary key

Tabella CoordinatorePerLaProgettazione		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
codiceSopralluogo	String	Primary key

Tabella CoordinatorePerL'esecuzione		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
codiceSopralluogo	String	Primary key

Tabella DirettoreTecnicoDiCantiere		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
codiceSopralluogo	String	Primary key

Tabella PersonePresentiAlSopralluogo		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key

codiceSopralluogo	String	Primary key
-------------------	--------	-------------

Tabella Verbali		
Nome della colonna	Tipo della data	Nota
CodiceVerbali	String	Primary key
CodicePresentatore	String	Foreign key
CodiceContravvenzione	String	Foreign key
codiceSopralluogo	String	Foreign key
NumeroDiVerbali	String	
Note	String	
ParoleChiave	String	
DataDiVerbale	Date	
DataDiRicezioneAtto	Date	
DataDiAttoEmessoFatto	Date	
AttoProvvedimenti	String	
consegnaVerbale	String	
TipoDiVerbali	string	Questa colonna contiene lista di tre righe(Disposizione, Prescrittivo e RichiamoAllaNorma)

Tabella Presentatori		
Nome della colonna	Tipo della data	Nota
CodicePresentatori	String	Primary key
Nome	String	
Cognome	String	
Qualifica	String	un data list di quattro scelte (Dirigenti, Quadri, Impiegati e Operai)

Tabella Contravvenzione		
Nome della colonna	Tipo della data	Nota
CodiceContravvenzione	String	Primary key
CFLavoratoriContravventore	String	Foreign key
CFLavoratoriResponsabileDiSicurezza	String	Foreign key
CFLavoratoriRappresentanteLegaleDiDitta	String	Foreign key
PresentatoIlContravventore	Boolean	
DataDiContravvenzione	Date	
CusatoSequestro	Boolean	
dataDiInizioSequestro	Date	
dataDiFineSequestro	Date	
CusatoDivieto	Boolean	
dataDiInizioDivieto	Date	

dataDiFineDivieto	Date	
-------------------	------	--

Tabella ADA		
Nome della colonna	Tipo della data	Nota
CodiceUC	String	Primary key
ADA	String	
Osservazione	String	
Indictori	String	
Risultato	String	

Tabella Capacita		
Nome della colonna	Tipo della data	Nota
CodiceCapacita	String	Primary key
Capacita	String	

Tabella Conoscenza		
Nome della colonna	Tipo della data	Nota
CodiceConoscenza	String	Primary key
Conoscenza	String	

Tabella ADACapacita		
Nome della colonna	Tipo della data	Nota
CodiceUC	String	Primary key
CodiceCapacita	String	Primary key

Tabella ADAConoscenza		
Nome della colonna	Tipo della data	Nota
CodiceUC	String	Primary key
CodiceConoscenza	String	Primary key

Tabella FaseCostruttivaADA		
Nome della colonna	Tipo della data	Nota
CodiceFaseCostruttiva	String	Primary key
CodiceUC	String	Primary key

Tabella LavoratoreCapacita		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
CodiceCapacita	String	Primary key

Tabella LavoratoreConoscenza		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
CodiceConoscenza	String	Primary key

Tabella LavoratoreCantieri		
Nome della colonna	Tipo della data	Nota
CFLavoratore	String	Primary key
CodiceCantieri	String	Primary key

Tabella DittaCantieri		
Nome della colonna	Tipo della data	Nota
PartitaIVA	String	Primary key
CodiceCantieri	String	Primary key

Tabella CantieriFaseCostruttiva		
Nome della colonna	Tipo della data	Nota
CodiceFaseCostruttiva	String	Primary key
CodiceCantieri	String	Primary key

Tabella Ativitta		
Nome della colonna	Tipo della data	Nota
CodiceAtivitta	String	Primary key
nomeAtivitta	String	

Tabella ProdottiFiniti		
Nome della colonna	Tipo della data	Nota
CodiceProdottoFinito	String	Primary key
nomeProdottoFinito	String	
Settore	String	Questa colonna contiene lista di otto righe(IAN, ISP, PISLL, SPV, ML, MS, DIPART. (IAN/ISP/PISLL/SPV) e DIPART. (ISP/PISLL))

Tabella AtivittaPFObbligatorie		
Nome della colonna	Tipo della data	Nota
CodiceAtivitta	String	Primary key
CodiceProdottoFinito	String	Primary key

Tabella AttivitaPFPossibile		
Nome della colonna	Tipo della data	Nota
CodiceAttivitta	String	Primary key
CodiceProdottoFinito	String	Primary key

Tabella DipartimentoPrevenzione		
Nome della colonna	Tipo della data	Nota
codiceDipart	String	Primary key
nomeDipart	String	
viaDipart	String	
CapDipart	Int	
comuneDipart	String	
Codicecomuni	String	Codice ISTAT
ProvinciaDipart	String	
RegioneDipart	String	

Tabella AttivitaEseguita		
Nome della colonna	Tipo della data	Nota
codice AttivitaEseguita	string	Primary key
codiceDipartimentoPrevenzione	String	Foreign key
Destinatorio (CFLavoratore)	String	Foreign key
Destinatorio (PartitaIVAditta)	String	Foreign key
Firma (codice Operatori)	String	Foreign key
DataDiAttivita	Data	
DataFirma	Data	
haZona	String	
Stato	String	Questa colonna contiene lista di tre righe(Incorso, Chiuso e Nullo)

Tabella ProdottiFinitiRisultati		
Nome della colonna	Tipo della data	Nota
codiceProdottiFinitiRisultati	string	Primary key
codice AttivitaEseguita	string	Foreign key
Firma (codice Operatori)	String	Foreign key
DataFirma	Data	
DataInizio	Data	
DataFine	Data	
Stato	String	Questa colonna contiene lista di tre righe(Incorso, Chiuso e Nullo)

Tabella AttivitaEseguitaAtivitta		
Nome della colonna	Tipo della data	Nota
codiceAtivitta	string	Primary key
codice AttivitaEseguita	string	Primary key

Tabella PFRisultatiProdottiFiniti		
Nome della colonna	Tipo della data	Nota
codiceProdottoFinito	string	Primary key
codiceProdottiFinitiRisultati	string	Primary key

Tabella Distretto		
Nome della colonna	Tipo della data	Nota
codiceDistretto	string	Primary key
codiceAttivitaEseguita	string	Primary key
codiceComuni		

Tabella Operatori		
Nome della colonna	Tipo della data	Nota
codiceOperatore	string	Primary key
nome Operatori	string	
Cognome Operatori	string	
codiceDipartimentoPrevenzione	string	Foreign key

Tabella OperatoriSopralluogo		
Nome della colonna	Tipo della data	Nota
codice Sopralluogo	string	Primary key
codiceOperatore	string	Primary key

Appendice 4

Lavoratori			
CFLavoratore	<input type="text" value="palo2376"/>	ProvinciaDIResidenza	<input type="text" value="FI"/>
IdLavoratori	<input type="text" value="3256"/>	RegioneDIResidenza	<input type="text" value="toscana"/>
Nome	<input type="text" value="Paolo"/>	Qualifica	<input type="text" value="Operai"/>
Cognome	<input type="text" value="fiorentino"/>	Istruzione	<input type="text" value="dotoratto"/>
dataDINascita	<input type="text" value="01/01/1980"/>	CodiceLivelloIstruzione	<input type="text" value="Livello_6"/>
CitaDINascita	<input type="text" value="Firenze"/>	Professione	<input type="text" value="ingegnere_civile"/>
codiceCitaDINascita	<input type="text" value="048017"/>	CodiceLivelloProfessione	<input type="text" value="3.1.3.5.0"/>
Sesso	<input type="text" value="maschile"/>	CodiceRuoloSicurezza	<input type="text"/>
StatoCivile	<input type="text" value="sposato"/>	Alloggio	<input type="text" value="false"/>
ViaDIAbitazione	<input type="text" value="via de guicciardini 35"/>	note	<input type="text"/>
CapDIAbitazione	<input type="text" value="50125"/>		
cittàDIAbitazione	<input type="text" value="Firenze"/>		
codicecittàDIAbitazione	<input type="text" value="048017"/>		
ProvinciaDIAbitazione	<input type="text" value="FI"/>		
RegioneDIAbitazione	<input type="text" value="toscana"/>		
ViaDIResidenza	<input type="text" value="via del gegnoro 113"/>		
CapDIResidenza	<input type="text" value="50135"/>		
cittàDIResidenza	<input type="text" value="Firenze"/>		
codicecittàDIResidenza	<input type="text" value="048017"/>		

Main Program

BEGIN

Set codiceFiscale;

Set IdLavoratori;

Set haNome;

Set haCognome ;
Set hadataDiNascita;
Set codiceCittaDiNascita;
Set codicecittàDIResidenza;
Set viaDIResidenza;
Set capDIResidenza;
Set codicecittàDIabitazione;
Set viaDIabitazione;
Set capDIabitazione;
Set haAlloggio;
Set Note;
Set haStatoCivile;
Set haSesso;
Set Istruzione;
Set Cod_Istruzione;
Set haQualifica;
Set professione;
Set codProfessione;
Set haOccupatoRuoloSicurezza;
WriteToFile(codiceFiscale, IdLavoratori, haNome, haCognome, hadataDiNascita, codiceCittaDiNascita, codicecittàDIResidenza, viaDIResidenza, capDIResidenza, codicecittàDIabitazione, viaDIabitazione, capDIabitazione, haAlloggio, Note, haStatoCivile, haSesso, Istruzione, Cod_Istruzione, haQualifica, professione, codProfessione, haOccupatoRuoloSicurezza);

END

procedure WriteToFile(codiceFiscale, IdLavoratori, haNome, haCognome, hadataDiNascita, codiceCittaDiNascita, codicecittàDIResidenza, viaDIResidenza, capDIResidenza, cittàDIabitazione, viaDIabitazione, codicecapDIabitazione, haAlloggio, Note,

haStatoCivile, haSesso, Istruzione, Cod_Istruzione, haQualifica, professione, codProfessione, haOccupatoRuoloSicurezza)

BEGIN

Set FileWriter;

FileWriter.write('INSERT DATA

```
{ :'+ Istruzione +' rdf:type :'+ Cod_Istruzione + ',  
      owl:NamedIndividual .  
};
```

INSERT DATA

```
{  
  :'+ professione +' rdf:type :'+ codProfessione +' ,  
      owl:NamedIndividual .  
};
```

);

FileWriter.write('INSERT DATA

```
{ :'+ codiceFiscale +' rdf:type :Lavoratori ,  
      owl:NamedIndividual ;  
      :haNome " '+haNome +' "^^xsd:string ;  
      :haIdLavoratori " '+IdLavoratori +' "^^xsd:string ;  
      :hadataDiNascita "'+ hadataDiNascita +' "^^xsd:dateTime ;  
      :codiceCittaDiNascita : '+codiceCittaDiNascita +' ;  
      :haSesso :'+ haSesso +' ;  
      :cittàDIResidenza :'+ codicecittàDIResidenza +' ;  
      :viaDIResidenza "'+ viaDIResidenza +' "^^xsd:string ;  
      :capDIResidenza "'+capDIResidenza +' " ;  
      :haLivelloDiIstruzione :'+Istruzione +' ;  
      :haQualifica :'+haQualifica +' ;
```

```

        :haProfessione :'+professione +;');
IF (cittàDIabitazione is not empty)
BEGIN
FileWriter.write(':cittàDIabitazione:'+ codicecittàDIabitazione + ' ');
END
IF (viaDIabitazione is not empty)
BEGIN
FileWriter.write(':viaDIabitazione "'+ viaDIabitazione +'"^^xsd:string ');
END
IF (capDIabitazione is not empty)
BEGIN
FileWriter.write(':capDomicilio "'+ capDIabitazione +'"^^xsd:int ');
END
IF (haAlloggio is not empty)
BEGIN
FileWriter.write(':haAlloggio "'+ haAlloggio +'"^^xsd:boolean ');
END
FileWriter.write(':haCognome "'+haCognome+'"^^xsd:string .
}; ');
END

```

Il risultato di questo algoritmo è:

```

INSERT DATA
{ :dottorato rdf:type :Livello_6 ,
      owl:NamedIndividual .
};
INSERT DATA
{
      :ingegnere_civile rdf:type :3.1.3.5.0 ,

```

```

        owl:NamedIndividual .
};
INSERT DATA
{ :palo2376 rdf:type :Lavoratori ,
        owl:NamedIndividual ;
        :haNome " Paolo "^^xsd:string ;
        :haIdLavoratori " 3256 "^^xsd:string ;
        :hadataDiNascita " 01/01/1980 "^^xsd:dateTime ;
        :codiceCitaDiNascita :048017 ;
        :haSesso :Maschile ;
        :cittàDIResidenza :048017;
        :viaDIResidenza "via del gegnoro 133"^^xsd:string ;
        :capDIResidenza :50135 ;
        :haLivelloDiIstruzione :dottorato ;
        :haQualifica :Operai ;
        :haProfessione :ingegnere_civile;
        : cittàDIabitazione :048017;
        :viaDIabitazione "via di guicciardini 35"^^xsd:string ;
        :cap DIAbitazione :50125 ;
        :haAlloggio "false"^^xsd:boolean ;
        :haCognome "kahlawi"^^xsd:string .
};

```

- **Cancella**

Lo scopo dell'aggiornamento è come quello precedente, ma qui si cancella i dati nel database al posto di inserirli nel database, e per applicarlo usiamo questo algoritmo.

Main Program

BEGIN

Set codiceFiscale;

```

Set Istruzione;
Set professione;
Set haOccupatoRuoloSicurezza;
WriteToFile(codiceFiscale, Istruzione, professione);
END
procedure WriteToFile(codiceFiscale, Istruzione, professione)
BEGIN
Set FileWriter;
FileWriter.write('DELETE WHERE { :'+codiceFiscale+' ?product ?object};
                DELETE WHERE { :'+Istruzione+' ?product ?object};
                DELETE WHERE { :'+professione+' ?product ?object};');
END

```

Il risultato di questo algoritmo è:

```

DELETE WHERE { :palo2376 ?product ?object};
DELETE WHERE { :dottorato ?product ?object};
DELETE WHERE { :ingegnere_civile ?product ?object};

```

- **Aggiorna**

Lo scopo dell'aggiornamento è come lo scopo del primo ma qui si aggiornano i dati nel database al posto di inserirli nel database, e per applicarlo usiamo un algoritmo contiene prima l'algoritmo dell'icona "Cancella" e poi l'algoritmo dell'icona "Inserisce".

Appendice 5

Ottenere una ontologia da un database

Quando si converte un database in una ontologia occorre che tutte le componenti del modello del database (ER diagramma) (Watt, Eng, 2012) siano riferibile all'ontologia.

Per raggiungere questo obiettivo si passa attraverso quattro fasi (Chujai, Kerdprasop, Kerdprasop, 2014); che spiegheremo nel modo seguente.

Il metodo

i. Trasformare le entità

Ogni entità nel diagramma ER può essere una classe nell'ontologia come vedremo nella seguente tabella.

Componente ER	Componente ontologia
entità	Classi
entità con chiave esterna	Classi o subclassi di forte entità
Specifica ogni entità non si sovrappone	classi disgiunti

ii. Trasformare la relazione

Ogni relazione può essere trasformata in un object property nell'ontologia come vedremo attraverso la tabella seguente.

ER componente	Ontology componente
Relazione	Object propertie
Relazione binaria senza attributi	La via della trasformazione dovrebbe essere divisa in due proprietà dell'oggetto; la prima dovrebbe essere la stessa della relazione che viene mostrata nel modello ER, e la seconda dovrebbe essere una proprietà inversa della precedente
Relazione binaria con attributi	Le fasi sono le seguenti: 1. si prende il nome della relazione di modello ER per creare la terza classe.

	<p>2. si prendono gli attributi di relazione per creare data property che ha il dominio alla terza classe.</p> <p>3. Dopo che abbiamo fatto le prime fasi successive, facciamo la fase di relazione binaria senza attributi.</p>
Relazione da entità A a entità B	un proprietà che ha il dominio A e il range B e una inversa proprietà che ha il dominio B e il range A
1:1 relazione	Impostare la funzione proprietà e Set cardinalità massimo uno
1:N e relazione	Impostare la funzione proprietà e Set cardinalità massimo uno dalla proprietà inversa
N:1 relazione	Impostare la funzione proprietà e Set cardinalità massimo uno con lo stesso nome del modello ER
Specifica ogni relazione non si sovrappone	Proprietà disgiunti

iii. Trasformare gli attributi

Ogni attributo può essere trasformato in una data property nell'ontologia come vedremo attraverso la tabella seguente.

ER componente	Ontology componente
Attributi	Data properties
entità di attributi	Dominio
Tipo di dati (real, integer etc.)	Range

iv. Trasformare la cardinalità

Restrizioni cardinalità vengono utilizzati per specificare ulteriormente le proprietà dell'ontologia (Pasha, Sattar, 2012), minCardinality e maxCardinality della proprietà sarà 1 se l'attributo è la chiave primaria o estere. Il minCardinality della proprietà è 1 se ogni attributo è dichiarato come NOT NULL. Inoltre, il maxCardinality della proprietà è 1, se qualsiasi attributo è dichiarato come UNICO.

Appendice 6

Ottenere un database da un'ontologia

Per generare uno schema relazionale da file XML integrate, si definisce le seguenti regole di trasformazione (Saccol, Andrade, Piveta, 2011):

- Non-lessicale concetto: mappato ad una tabella. La tabella è chiamata come il concetto. La chiave primaria viene creata e chiamata come `cod_ <nome tabella>` automaticamente. Ad esempio, se la tabella si chiama `lavoratore`, la sua chiave primaria è `cod_lavoratore`.
- Lessicale concetto: mappato alla colonna della tabella corrispondente al concetto non lessicale a cui appartiene. Cioè, le relazioni tra un “non lessicale” e un concetto “lessicale” determina che la tabella generata per il concetto non lessicale ottenga una colonna con il nome e il tipo del concetto lessicale associato. Ad esempio, se un concetto `lavoratore` non lessicale che ha una relazione ad un nome concetto lessicale viene inserito, in questo caso una “`lavoratore tabella`” verrà aggiunta con una colonna denominata nome dello stesso tipo del loro concetto lessicale.
- relazioni: la cardinalità informa se la colonna è obbligatoria (not null). Ad esempio, se contiene (1, X) e X è 1 o N, quindi la colonna non è nulla. Nelle relazioni che presentano contemporaneamente dell'origine e della destinazione in concetti non lessicali, analizziamo le cardinalità quelli diretti e quelli inversi.
- Se la cardinalità è (1 a 1) o (1 a N): si crea una chiave esterna che fa riferimento alla tabella di origine con il nome della sua chiave primaria nella tabella di destinazione.
- Se la cardinalità è (N a 1): si crea una chiave esterna che fa riferimento alla tabella di destinazione con il suo nome chiave primaria della tabella di origine.
- Se la cardinalità è (N a N): viene creata una nuova tabella, che è chiamata come la concatenazione dei due concetti. Questa tabella contiene una chiave primaria costituita dalla colonna con lo stesso nome e tipo di chiave primaria della tabella di origine e la colonna con lo stesso nome e tipo di chiave primaria della tabella di destinazione.

Appendice 7

```
select    ?labelmalatia1    ?labelmalatia    ?sottoclass    ?Ontologia_della_Malattia
?HumanDiseaseOntology    ?disease    ?predicate_variable_4    ?predicate_variable_3
?predicate_variable_2 ?predicate_variable_1 ?predicate_variable_6 where

{ ?HumanDiseaseOntology ?predicate_variable_6 ?disease .

  ?Ontologia_della_Malattia rdfs:label ?labelmalatia .

  <http://purl.obolibrary.org/obo/DOID_0050117>                ?predicate_variable_2
?HumanDiseaseOntology ;

                                <http://www.w3.org/2002/07/owl#equivalentClass>
<http://www.semanticweb.org/adham/ontologies/2014/11/Malattie#Batteri,_virus_ed_altri_a
genti_infettivi> .

<http://www.semanticweb.org/adham/ontologies/2014/11/Malattie#Batteri,_virus_ed_altri_a
genti_infettivi> ?predicate_variable_3 ?Ontologia_della_Malattia .

  ?sottoclass                                ?predicate_variable_4
<http://www.semanticweb.org/adham/ontologies/2014/11/Malattie#Batteri,_virus_ed_altri_a
genti_infettivi> ;

                                rdfs:label ?labelmalatia1 . }
```