

INAIL
Quaderni di ricerca

**Una proposta di ontologia per
gli *Open Data* Inail sugli infortuni:
dalla numerosità dei dati
alla gestione della conoscenza**

Patrizia Agnello, Silvia Maria Ansaldi

I *Quaderni di ricerca* dell'Inail sono lo strumento a disposizione dei ricercatori e dei ruoli professionali dell'Istituto per rendere pubblici i risultati più rilevanti delle loro attività; accolgono anche contributi di ricercatori esterni, che partecipano ai progetti dell'Inail.

Gli autori hanno la piena responsabilità delle opinioni espresse nei *Quaderni*, che non vanno intese come posizioni ufficiali dell'Inail.

I *Quaderni* sono disponibili online all'indirizzo www.inail.it.

Comitato di redazione: Paolo Bragatto, Massimo De Felice,
Benedetta Persechino, Patrizio Rossi

Segreteria di redazione: Toni Saracino

INAIL - Piazzale Giulio Pastore, 6 | 00144 Roma

ISBN 978-88-7484-481-4

Stampato presso la Tipografia INAIL – Milano, febbraio 2016

Una proposta di ontologia per gli Open Data Inail sugli infortuni: dalla numerosità dei dati alla gestione della conoscenza

Patrizia Agnello, Silvia Maria Ansaldi¹

Sommario – Negli ultimi anni è stato avviato il processo di pubblicazione di “open data” da parte delle Pubbliche Amministrazioni anche a seguito delle norme in materia di trasparenza e di riutilizzo dell’informazione del settore pubblico.

Nel *Quaderno* n.1 di questa collana si descrive il “modello di lettura” della numerosità sugli *Open Data* Inail relativi agli infortuni sul lavoro; tale modello è indispensabile per “capire” i concetti a cui i dati corrispondono e le loro relazioni. Il vocabolario, che nasce per indirizzare gli utenti verso il linguaggio speciale Inail, è di fatto una linea guida documentale a cui accedere per conoscere il *significato* di uno specifico dato.

Questo *Quaderno* presenta la definizione di un “modello concettuale formalizzato” in grado di gestire i concetti, e le relazioni tra loro, contenuti nel vocabolario, e per mezzo del quale accedere agli *Open Data* con una chiave di lettura esplicita e più completa.

La scelta dell’ontologia offre vantaggi nella definizione del modello attraverso la determinazione di classi, relazioni e proprietà, nella classificazione automatica dei nuovi casi di infortunio inseriti e per l’interoperabilità con altri pertinenti domini di conoscenza disponibili. Sono stati adottati standard opportuni compatibili con le indicazioni relative al processo di valorizzazione del patrimonio pubblico per l’Amministrazione Digitale.

Su quest’ultima considerazione si inserisce un altro aspetto: capire se e come gli *Open Data* sugli infortuni possano essere utilizzati nei progetti di ricerca, ed il ruolo che può avere l’ontologia che li descrive per sviluppi futuri.

L’ontologia proposta non è un prototipo generale applicabile a qualunque tipo di *open data*, ma modella esclusivamente il dominio “infortuni sul lavoro” come descritto negli *Open Data* Inail attraverso il modello di lettura.

Parole chiave, frasi: open data, ontologia, infortuni, interoperabilità.

¹ Le autrici svolgono l’attività di ricerca presso il Dipartimento Innovazioni Tecnologiche e Sicurezza degli Impianti, Prodotti e Insediamenti Antropici dell’Inail. La loro esperienza sullo sviluppo di ontologie è maturata nell’ultimo decennio nell’ambito delle industrie a rischio d’incidente rilevante dove vi è l’obbligo di legge di produrre il Sistema di Gestione della Sicurezza. Per le piccole aziende, hanno sviluppato software basati su ontologie sia per il supporto alla decisione sia per il coinvolgimento attivo dei lavoratori nella cultura della sicurezza.

Si ringraziano M. De Felice, R. Mosca e M. Veltroni per il prezioso aiuto alla migliore comprensione della struttura degli *Open Data* e alla loro corretta interpretazione.

Commenti possono essere inviati a p.agnello@inail.it e s.ansaldi@inail.it

Indice degli argomenti

Premessa	7
Convenzioni adottate nel testo	8
Acronimi	8
Definizioni	9
1 Ontologia: dalla filosofia all'informatica	12
1.1 Strumenti informatici per le ontologie	14
1.2 Standard	14
1.3 Concetti base	15
2 Mappa semantica degli Open Data Inail sugli infortuni	16
2.1 Open Data Inail	17
2.2 Metodologia per la definizione di <i>ontoInfortuni</i>	17
2.3 Processo di definizione e formalizzazione del modello concettuale	18
2.3.1 Processo <i>bottom-up</i>	19
2.3.2 Processo <i>top-down</i>	20
3 L'ontologia <i>ontoInfortuni</i>	21
3.1 Architettura e convenzioni	22
3.2 Modello proposto per <i>ontoInfortuni</i>	24
3.3 Tipologiche	31
3.3.1 Modellazione di tipologiche semplici	31
3.3.2 Modellazione di tipologiche complesse	33
3.3.3 Modellazione di tipologiche parametriche	39
4 Esempi di navigazione, interrogazione, visualizzazione dell'ontologia	43
4.1 Modalità di navigazione ipertestuale: <i>OWLDoc</i>	45
4.2 Modalità di interrogazione: <i>SPARQLinfortuni</i>	49
4.3 Modalità di rappresentazione grafica di <i>ontoInfortuni</i>	51
5 <i>Open Data</i> e ontologie nelle altre Pubbliche Amministrazioni	53
5.1 Portale dei <i>Linked Open Data</i> (SPCData)	53
5.2 Open Data di INPS	55
5.3 Altri esempi di interesse	57
5.4 LOV, <i>Linked Open Vocabulary</i>	58
6 Considerazioni conclusive e possibili utilizzi	60
Riferimenti Bibliografici	61

Premessa - La produzione e gestione di grosse quantità di dati ed informazioni che spesso si sovrappongono, poiché originate da fonti differenti, in modalità non compatibili tra loro e con differenti finalità, e il contemporaneo avanzamento delle tecnologie dell'informazione e della comunicazione per una migliore gestione e pubblicizzazione di tali dati, hanno richiesto un intervento normativo che, soprattutto nei confronti delle Pubbliche Amministrazioni, permette di definire, organizzare e proporre dati di utilità pubblica secondo regole ben precise ed uniformi, nel rispetto dei vincoli di segretezza e privacy laddove ritenuti imprescindibili, per permettere l'accesso ai dati senza restrizioni non riconducibili a esplicite norme di legge.

Anche Inail ha aderito alle indicazioni dei decreti legislativi, in materia di riutilizzo dei documenti contenenti dati nel settore pubblico e di Codice dell'amministrazione digitale, pubblicando gli *Open Data* sugli infortuni nei formati previsti per la consultazione pubblica ed affiancati da un modello di lettura e da un vocabolario che ne permettano la corretta interpretazione.

Nella pagina di presentazione degli *Open Data* Inail si legge: "Nell'ambito del processo di valorizzazione del proprio patrimonio informativo, l'Inail mette a disposizione dei cittadini un set di dati pubblici, in formato aperto e senza restrizioni per il riutilizzo: un patrimonio che può rappresentare un'interessante opportunità per la sua valenza storico-statistica, ma soprattutto sociale e scientifica". Con l'esperienza maturata in questi anni sulla gestione e sul recupero della conoscenza attraverso l'utilizzo di ontologie, appositamente create nei diversi domini d'interesse, si è scelto di porsi come utenti esterni con tale esperienza che approdano agli *Open Data* degli Infortuni nel sito Inail con l'intenzione di realizzare una applicazione che li utilizzi. Si è così provato ad integrare in un modello concettuale formalizzato le informazioni rappresentate negli *Open Data* con la conoscenza contenuta nei documenti a corredo degli *Open Data* stessi, tenendo anche conto le regole di interoperabilità tra pubbliche amministrazioni.

L'obiettivo è quello di permettere una lettura più completa degli *Open Data* e di produrre una solida struttura per le ontologie tale da accogliere le sfide di interoperabilità e *linked open data* proposte per l'amministrazione digitale attraverso lo sviluppo di ontologie già collegabili a quei dati di valore più generale, gestiti ed aggiornati dalle amministrazioni che li producono, ed inoltre si è predisposto un modello di conoscenza sugli infortuni più robusto che permetta di condurre analisi per la prevenzione degli infortuni se utilizzato sui dati completi da cui gli *Open Data* sono estratti.

Il modello proposto nel *Quaderno* è stato chiamato *ontoInfortuni*.

Gli strumenti utilizzati nello sviluppo dell'ontologia sono quelli comunemente utilizzati in tale ambito, ma per rendere la proposta compatibile con quanto già presente nelle Pubbliche Amministrazioni ed enti di ricerca si è condotta preliminarmente un'analisi sul livello di sviluppo dei protocolli e sull'aderenza alle normative in materia.

La formalizzazione del modello permette, attraverso le ontologie, di definire diverse rappresentazioni, anche di tipo grafico, che meglio si prestano per illustrare ed applicare il modello.

Questo *Quaderno* si articola in 6 capitoli. Nel primo capitolo si introduce l'ontologia, con un breve cenno sull'evoluzione dell'uso di questo termine, e con l'illustrazione degli aspetti informatici propri di questa metodologia. Il capitolo 2 tratta lo sviluppo metodologico adottato per la creazione della mappa semantica relativa agli *Open Data* Inail sugli infortuni. Nel capitolo 3 è descritta *ontoInfortuni*, ovvero l'insieme di ontologie sviluppate per gli obiettivi sopra descritti; nel capitolo 4 alcuni esempi di navigazione ed interrogazione. Il quinto capitolo presenta una breve panoramica sullo stato dell'arte nelle altre Pubbliche Amministrazioni. Nel capitolo 6, infine, alcune considerazioni conclusive sull'ontologia proposta e sui possibili sviluppi delle attività di ricerca attraverso un utilizzo proficuo di *ontoInfortuni*.

Convenzioni adottate nel testo

Per gli elementi che compongono l'ontologia si è preferito usare, quando possibile, il nome in italiano ponendo tra parentesi tonde il corrispondente in inglese ai fini di una lettura agevole di alcune figure.

I metadati sono sempre riportati tra parentesi acute <...>.

Vocabolario: nel testo questo termine indica sempre ed esclusivamente il vocabolario presentato nel *Quaderno 1* [CDFMV-013] alle pagine 13 - 25.

Con l'espressione "modello di lettura" si indica il paragrafo 6.1 dello stesso *Quaderno 1* della collana *Quaderni di ricerca* dell'Inail.

Acronimi

CSV – Comma Separated Values

FOAF – Friend Of A Friend

HTML – HyperText Markup Language

LOD - Linked Open Data

LOV - Linked Open Vocabulary

OD – Open Data

OWL - Web Ontology Language

PA – Pubblica Amministrazione

RDF - Resource Description Framework

RDFS - RDF Schema

SDMX – Statistical Data and Metadata eXchange

SKOS – Simple Knowledge Organization System

SPC – Sistema Pubblico di Connettività e Cooperazione

SPARQL - Sparql Protocol And RDF Query Language

URI – Uniform Resource Identifier

W3C – World Wide Web Consortium

XKOS – eXtended Knowledge Organization System

XLS – estensione file Excel

XML – eXtensible Markup Language

Definizioni

Ateco: classificazione delle attività economiche.

Classe (*Class*): inteso come componente dell'ontologia, è l'insieme che contiene gli individui.

Comma Separated Values (CSV): formato di file di testo utilizzato per rappresentare informazioni, organizzate in una struttura tabellare.

Data Cube: è un vocabolario che consente di rappresentare dati multi-dimensionali in formato RDF.

Dataset: una collezione di dati, generalmente riguardanti una stessa organizzazione, che vengono erogati e gestiti congiuntamente.

DatatypeProperty: inteso come componente dell'ontologia, è la relazione binaria tra un individuo e un dato di tipo RDF.

Dominio (*Domain*): inteso come componente dell'ontologia, è il dominio (in senso matematico) di una relazione.

Dublin Core: è un sistema di metadati costituito da un nucleo di elementi essenziali per descrivere qualsiasi materiale digitale accessibile attraverso la rete informatica.

Friend Of A Friend (FOAF): è un'ontologia adatta a descrivere persone, con le loro attività e le relazioni con altre persone e oggetti.

HyperText Markup Language (HTML): linguaggio utilizzato per rappresentare pagine web e documenti in format ipertestuale.

Individuo (*Individual*): inteso come componente dell'ontologia, è un elemento o un oggetto specifico nel dominio.

Inferenza: è la capacità (espressa attraverso un processo logico) del *reasoner* di dedurre un'informazione (o concetto) a partire da altre informazioni.

Interoperabilità: in ambito informatico, la capacità di sistemi differenti e autonomi di cooperare e di scambiare informazioni in maniera automatica, sulla base di regole comunemente condivise. [AGID-014b]

Interoperabilità semantica: la capacità di elaborare informazioni da fonti esterne o secondarie senza perdere il reale significato delle informazioni stesse nel processo di elaborazione. [AGID-014b]

Istanza (*Instance*): inteso come componente dell'ontologia, è sinonimo di individuo.

Java: linguaggio di programmazione.

Jena: Apache Jena è una struttura Java open source per sviluppare applicazioni semantiche per il WEB e Linked Data.

Linked Open Data (LOD): sono dati aperti a cui è stata assegnata un'identità per renderli collegati tra loro e interoperabili.

Linked Open Vocabulary (LOV): sito contenente una vasta collezione di ontologie e vocabolari.

Mappa semantica: rappresentazione del modello concettuale della conoscenza, attraverso un grafo i cui nodi corrispondono ai concetti e gli archi alle relazioni.

Metadato: un'informazione che descrive un insieme di dati.

Modello concettuale: è un modello che definisce informazioni collegamenti tra i diversi concetti, anche attraverso un vocabolario.

ObjectProperty: inteso come componente dell'ontologia, è la relazione binaria tra due individui appartenenti a due classi.

Open Data: più specificatamente dati aperti del settore pubblico, l'insieme delle informazioni, pubbliche o no, accessibili e riutilizzabili.

Ontologia: in ambito informatico, una rappresentazione formale e condivisa dei concetti e delle mutue relazioni che caratterizzano un certo dominio di conoscenza.

Namespace: inteso come componente di un'ontologia, è una sequenza di caratteri che viene usata come prefisso per tutti gli elementi (classi, proprietà, individui) che compongono quell'ontologia, fondamentale per garantirne la non ambiguità.

PHP: linguaggio di programmazione.

Query: in ambito SPARQL, è la forma di interrogazione di dati RDF, contenenti le triple necessarie ad estrarre informazioni dal modello. Ogni tripla si compone di soggetto, predicato, oggetto.

Plug-in: in ambito informatico, un programma non autonomo ma che si avvale ed estende le funzionalità originarie di un sistema.

Range: inteso come componente dell'ontologia, è il codominio (in senso matematico) di una relazione.

Relazione (o proprietà): inteso come componente dell'ontologia, mette in corrispondenza due elementi dell'ontologia.

Resource Description Framework (RDF): è un linguaggio per rappresentare informazioni sul Web. [RDF]

Restrizione: inteso come componente dell'ontologia, è il requisito per determinare l'appartenenza di un individuo ad una classe. Ci sono vari tipi di restrizioni, quali: universale, esistenziale, di cardinalità, di valore.

RDF Schema (RDFS): è un'estensione di RDF che consente di rappresentare semplici schemi per i dati.

Reasoner: o classificatori, sono sistemi che offrono servizi di controllo e verifica di coerenza e consistenza, di inferenza (deduzione) operando sulle basi di conoscenza realizzate con le ontologie.

Simple Knowledge Organization System (SKOS): è un'ontologia adatta a rappresentare vocabolari, tassonomie, thesauri.

Sparql Protocol And RDF Query Language (SPARQL): è un linguaggio standard W3C di interrogazione dati RDF (*Resource Description Framework*) e costituisce una delle basi portanti dei linked open data e del web semantico, una metodologia sempre più utilizzata da provider di dati anche istituzionali. [SPARQL]

SPARQL Endpoint: è uno strumento software, conforme con il protocollo SPARQL, che abilita gli utenti (persone o macchine) ad interrogare una base di conoscenza.

SPCData: è lo spazio contenente i dati del Sistema Pubblico di Connettività e Cooperazione (SPC).

Statistical Data and Metadata eXchange (SDMX): è un linguaggio XML per lo scambio di dati e metadati statistici.

Tassonomia: in generale, la disciplina della classificazione. In ambito matematico, una struttura (gerarchica) ad albero di categorie appartenenti ad un dato gruppo di concetti.

Uniform Resource Identifier (URI): sequenza di caratteri che identifica univocamente una risorsa (es.: pagina WEB, documento, immagine).

Web Ontology Language (OWL): è un linguaggio di rappresentazione dell'ontologia, riconosciuto come standard dal Consorzio W3C.

World Wide Web Consortium (W3C): è un consorzio (comunità) internazionale che ha come obiettivo quello di rendere pubblico (aperto) il lavoro svolto insieme per sviluppare gli standard per il Web.

eXtensible Markup Language (XML): è un linguaggio di marcatura, standardizzato da W3C.

1 Ontologia: dalla filosofia all'informatica

“Ontologia”², tradotta con “fenomenologia dell'essere”, come branca della filosofia è la scienza empirica di ciò che è, dei tipi e strutture di oggetti, proprietà, eventi, processi e relazioni individuabili in ogni settore della realtà. Il termine ontologia è stato spesso usato come sinonimo di “metafisica” [Sm-999].

Attraverso il contributo dei filosofi del linguaggio, si è passati ad utilizzare il termine “ontologia” anche negli sviluppi informatici. L'esigenza è nata nel momento in cui la disponibilità di dati ed informazioni messi a disposizione dall'affermarsi della rete *internet* ha dato luogo alla necessità di creare delle regole di accesso e di discriminazione delle informazioni stesse.

Il percorso logico, avvenuto con il massivo utilizzo di *internet* nato all'inizio come rete militare americana, impiegato poi come rete universitaria per depositare e cercare informazioni, ha avuto negli anni come primo esito quello di raccogliere una enorme quantità di dati, documenti, informazioni, pagine web sui più diversi argomenti. Il momento successivo è stato il conseguente incremento di ricerche fatte sui dati raccolti con dei motori di ricerca basati sulla ricerca testuale di parole o frasi. Visti i tempi di attesa sempre più lunghi per ricevere la risposta alle interrogazioni, e l'enorme quantità di risultati ottenuti all'interno dei quali cercare quelli significativi rispetto al proprio interesse, si è pensato di “etichettare” tutto ciò che era stato raccolto attraverso i “metadati”, un insieme di informazioni appartenenti ad uno schema più generale e condiviso, che fornissero la descrizione dei dati.

La loro introduzione ha permesso non solo di cominciare a distinguere i domini di conoscenza, cui le informazioni estratte appartenevano, mappandoli attraverso i concetti propri del dominio stesso, ma ha consentito di rendere accessibile ai computer lo schema di riferimento dei dati, velocizzando la ricerca ed orientando la visualizzazione di quei risultati pertinenti [So-984].

Lo schema adottato per i metadati, perfezionato sulla base di un modello semantico e sviluppato, sulla base dei concetti e delle relazioni tra essi, dichiarate per un dominio specifico, è l'ontologia. L'obiettivo raggiunto quello di far parlare uomo e macchina attraverso un linguaggio “comune”.

L'ontologia diventa il possibile strumento per promuovere la condivisione e, vista come insieme di termini gerarchicamente strutturato che descrive un dominio, può essere usato come il fondamento per un sistema basato sulla conoscenza per quel dominio [SRKR-997].

Dopo i primi tentativi di formalizzare la conoscenza in uno schema unico che fosse interpretabile da un computer, col diffondersi dell'uso dei metadati e di conseguenza dei motori semantici, si è confermato che migliori risultati si potevano ottenere dividendo la conoscenza in domini di interesse per i quali creare specifiche tassonomie e relativi vocabolari controllati [Gu-995] [Gu-998] che diventano la base di sviluppo dell'ontologia, ovvero del modello concettuale formalizzato, di quel dominio.-

La prima definizione di ontologia data in ambito informatico è quella di Gruber del 1993: “*l'ontologia è una esplicita specificazione di una concettualizzazione*” [Gr-993], [Gr-995], [Gr]. L'ambito matematico da cui prende in prestito i formalismi è quello della teoria assiomatica del primo ordi-

² Pare che il termine “ontologia” venne introdotto in filosofia dal filosofo svizzero Jacob Lorhard (Lorhardus), nel suo *Ogdoas scholastica* del 1606. Sebbene mai usato da Aristotele viene attribuito, in tempi successivi, alla “scienza dell'essere in quanto essere” da lui proposta in Aristotele, *Metafisica*, libro I, cap. I, 1003 a, 21-26; e in Aristotele, *Metafisica*, IV 2, 1003 a, 32-34.

ne esprimibile attraverso la logica descrittiva, che costituisce la base di funzionamento del *reasoner* ai fini dell'inferenza [BCMGNP-003].

Dalla definizione originaria si passa a quella più applicativa ed estesa “*un'ontologia è una specificazione formale ed esplicita di una concettualizzazione condivisa*”. Questa frase, ulteriormente chiarita diventa: una rappresentazione esplicita *formale e condivisa* dei concetti e delle mutue relazioni che caratterizzano un certo dominio di conoscenza. I termini *formale e condivisa* sono fondamentali per descrivere le potenzialità di un'ontologia. Con il primo si lascia intendere l'utilizzo di formalismi standardizzati che oltre a consentirne la gestione della conoscenza attraverso un computer, garantiscono anche l'indipendenza da specifici strumenti informatici. Con il termine *condivisa* ci si riferisce ad altri due aspetti fondamentali di queste rappresentazioni: devono essere di consenso pubblico, soprattutto da parte degli esperti del dominio ai fini della loro validità, e devono poter essere riutilizzate, ovvero i concetti e le relazioni tra loro rappresentano uno schema di conoscenza indipendente dal singolo applicativo.

Proprio da quest'ultimo aspetto deriva il concetto di interoperabilità semantica [BCMM-014], ovvero di una struttura di conoscenza “comune” (condivisa sia dall'uomo sia dalla macchina) per una determinata tipologia di dati, che ne permetta la disponibilità (D.lgs. 7 marzo 2005, n. 82) ed il riuso (D.lgs. 24 gennaio 2006, n. 36), nel nostro caso, tra le Pubbliche Amministrazioni.

Negli ultimi anni sono stati pubblicati molti lavori scientifici relativi sia a sviluppi di ontologie, in diversi ambiti dell'ingegneria e della medicina, sia a metodi o applicazioni basati su di esse. Il primo e naturale uso delle ontologie nell'ingegneria della sicurezza è stato quello di sviluppare metodi per migliorare la ricerca nelle banche dati relative agli incidenti [SBFSC-006] e [BMSFC-009]. Alcuni studi hanno sviluppato ontologie per la gestione dei metodi di identificazione e analisi del rischio. [ZCZQC-009] illustra un'applicazione, basata su ontologie, che implementa il metodo HAZOP (*Hazard Operability*), [ERS-010] propone un'ontologia basata sul metodo FMEA (*Failure Mode and Effect Analysis*) per migliorare le prestazioni di sicurezza nell'ambito delle costruzioni. [TYDWTC-009] illustra un metodo, basato su ontologie, per la gestione del rischio per i lavori in appalto, [AMAG-012] presenta un'applicazione di supporto all'individuazione dei metodi di identificazione, analisi e valutazione dei rischi in ambito industriale. Le ontologie sono anche state utilizzate come base per sviluppare strumenti di supporto alle decisioni, per esempio, rivolte al monitoraggio e alla manutenzione delle attrezzature nelle industrie di processo [ECT-013], o all'individuazione dei dispositivi di protezione individuale da usare per la manutenzione negli stabilimenti a rischio di incidente rilevante [AAB-015]. Si sono inoltre sviluppate ontologie cosiddette superiori o generali (*upper ontology*) in grado di modellare e rappresentare un dominio applicativo vasto, che possono essere riutilizzate in sotto-domini più specifici. OntoCAPE è un esempio di *upper ontology*, rivolta al dominio di ingegneria di processo, per favorire la progettazione e gli sviluppi di strumenti applicativi software per la gestione della conoscenza relativa agli impianti industriali [MWM-009]. Infine, sono state sviluppate ontologie della sicurezza in specifici ambiti industriali, tenendo conto anche delle normative vigenti, in particolare relative alle attrezzature in pressione. [CGMBPA-008] illustra un esempio di ontologia di supporto alla progettazione di tali attrezzature e alla loro verifica secondo la normativa europea per le attrezzature in pressione. [BADs-013] propone un metodo per la gestione delle modalità di guasto e delle loro peculiarità, tenendo conto non solo della loro numerosità ma anche della conoscenza associata che, se opportunamente rappresentata, può essere di miglioramento per la gestione di tali attrezzature negli impianti industriali.

Andando all'aspetto concreto dello sviluppo realizzato, vale precisare a questo punto che prima

di cominciare con la progettazione di una ontologia bisogna porsi le seguenti domande: di cosa vogliamo parlare (il dominio); perché ne vogliamo parlare (l'obiettivo); dove troviamo le informazioni riutilizzabili già disponibili e quali mettiamo noi a disposizione (l'interoperabilità).

Questo costituisce lo "spazio dei problemi" per sviluppare il progetto dell'ontologia. Le soluzioni, appartenenti ad uno "spazio delle soluzioni" non sono così immediatamente individuabili ed inoltre devono rispondere sia alle caratteristiche descrittive del dominio sia all'obiettivo prefissato. Una ontologia ben progettata risponde a tali requisiti, ma è sempre influenzata dalle scelte di chi la progetta in termini di analisi del dominio rispetto agli obiettivi e di soluzioni adottate tra quelle possibili; si presenta quindi come soggettiva, pertanto non c'è unicità nello sviluppo.

1.1 Strumenti informatici per le ontologie

Un aspetto caratteristico delle ontologie sviluppate secondo gli standard OWL [OWL-012], il linguaggio formale di rappresentazione, è la disponibilità di strumenti informatici di *editing* per la definizione, e soprattutto di *reasoning*, per i controlli e le classificazioni dell'ontologia.

L'editore utilizzato in questo studio è Protégé [Ho-011], un *open-source* sviluppato dalla Stanford University (in particolare allo Standard Centre for Biomedical Informatics Research della Stanford University School of Medicine). Il sistema, il cui utilizzo è molto diffuso, ha il merito di avere un'interfaccia molto chiara e ben strutturata, di garantire la coerenza con OWL, di disporre di un *reasoner* affidabile e di potenti strumenti per il controllo e la verifica di validità dell'ontologia. Data la sua diffusione nella comunità scientifica, è ricco di *plug-in* per la visualizzazione e l'interrogazione dell'ontologia.

Il linguaggio OWL si differenzia da altri linguaggi, utilizzati per descrivere e rappresentare dati e informazioni sul *web* (per esempio HTML, XML), perché intende rappresentare la conoscenza (*knowledge*) presente in un certo dominio di interesse, composta di concetti e loro relazioni, e non vuole solo organizzare in maniera più o meno strutturata le informazioni.

I *reasoner* sono programmi che, a partire da un'ontologia, offrono una serie di servizi per ragionare sulla base di conoscenza, quali i controlli di validità, di coerenza e consistenza, di inferenza. Quest'ultimo è l'aspetto più interessante che forse più degli altri motiva l'utilizzo di ontologie, cioè la capacità automatica dello strumento di ricavare ed esprimere deduzioni sulla base delle modalità di definizioni delle classi, delle istanze e delle loro proprietà e restrizioni. In particolare i *reasoner* interni al software [Protégé] permettono di verificare la congruenza e le inferenze durante la fase di progettazione e sviluppo dell'ontologia. Nel corso della presentazione dell'ontologia sviluppata alcuni esempi di inferenze mettono in evidenza quali siano le proprietà che le hanno prodotte.

Negli esempi che abbiamo riportato per descrivere attraverso delle figure alcuni aspetti di *ontoInfortuni*, l'applicazione utilizzata è *OntoGraf* [OntoGraf], che fornisce una rappresentazione a grafo dell'ontologia o di parti di essa, e il *plug-in* SPARQL, particolarmente utile per verificare le *query* di interrogazione dell'ontologia.

1.2 Standard

Il documento "Linee guida nazionali per la valorizzazione del patrimonio pubblico (anno 2014)" [AGID-014b] illustra, tra l'altro, i formati tecnici e le ontologie di riferimento per poter rappresentare i modelli di dati e metadati. Qui di seguito forniamo l'elenco dei formati e delle ontologie/

vocabolari utilizzati o di interesse in questo progetto, rimandando al documento sopracitato per una più completa descrizione dei diversi standard.

Gli standard utilizzati in *ontoInfortuni* sono:

- RDF (*Resource Description Framework* [RDF-014a] [RDF-014b]) è uno dei formati in cui sono rappresentati gli *open data*, che consente di catturare la semantica dei dati. Ogni elemento viene considerato una risorsa, che se univocamente definita (nello specifico tramite un URI), si può rappresentare sul WEB, rendendola accessibile e fruibile da tutti.
- RDFS (RDF Schema) [RDF-014c], [RDF-014d] è un'estensione di RDF che consente la definizione di semplici schemi di dati, come le classi e sottoclassi, utili per rappresentare, per esempio, tassonomie e classificazioni.
- SPARQL (*Sparql Protocol And Rdf Query Language*) [SPA-008a], [SPA-008b], [SPA-013] è un protocollo standardizzato da W3C che consente l'interrogazione e la gestione dei dati di tipo RDF.
- OWL (*Web Ontology Language*) [OWL-004a], [OWL-004b], [OWL-012] è uno schema più evoluto di RDFS che permette di definire schemi più complessi, ulteriori formalismi, includendo semantica formale e logica descrittiva. OWL permette la definizione completa di un'ontologia.
- CSV (*Comma Separated Values*) [CSV-005] è un formato di file basato su file di testo utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o database) di una tabella di dati.

I formati standard utilizzati in questo progetto sono quelli normalmente utilizzati per gli *open data*.

1.3 Concetti base

Gli elementi che contraddistinguono un'ontologia espressa con il linguaggio OWL sono: gli individui (o istanze), le classi a cui gli individui appartengono, le proprietà e le relazioni fra le istanze. Gli individui o istanze (*individual* o *instance*) sono gli oggetti che caratterizzano il dominio da rappresentare. Con il termine classe (*class*) si definisce l'insieme che contiene un certo tipo di individui; quindi la classe deve essere formalmente (cioè matematicamente) descritta per precisare i requisiti di appartenenza degli individui alla classe. Le classi possono essere organizzate in modo gerarchico, con superclassi e sottoclassi, un esempio è il caso in cui il dominio di interesse sia rappresentabile in base a delle tassonomie. Ogni sottoclasse eredita le proprietà della classe padre. Esiste una classe globale che in qualche modo rappresenta il dominio di conoscenza nella sua interezza, ed in OWL è denominata classe *owl:Thing*, ovvero la radice unica da cui discendono tutte quelle definite dall'utente che risultano sue sottoclassi.

Le proprietà sono relazioni binarie (vale a dire relazioni tra due elementi), se ne distinguono di due tipi: relazioni con un dato (*DatatypeProperty*) e relazioni con un individuo (*ObjectProperty*). Più precisamente, *DatatypeProperty* corrisponde alla relazione tra individuo (*instance*) e tipo di dato contenuto negli schemi RDF o XML, per esempio *integer*, *string*, *date*; *ObjectProperty* è la relazione tra individui appartenenti a due classi.

Come per le classi, anche le proprietà si possono definire in modo gerarchico, come specializza-

zioni di proprietà esistenti. Ogni proprietà, intesa come relazione, ha un *dominio* (*domain*) e un *range* che possono essere specificati oppure no. Il *dominio* di una proprietà è la classe, o l'insieme delle classi, alle cui istanze di appartenenza si può applicare quella proprietà. Analogamente, il *range* di una proprietà è la classe, o insiemi di classi, che contengono gli individui che possono essere valori di quella proprietà.

Si possono, poi, assegnare delle caratteristiche matematiche alle relazioni, una *ObjectProperty* può essere, ad esempio: funzione, inversa, transitiva, simmetrica o riflessiva.

Un utilizzo delle proprietà è quello di definire le restrizioni di una classe, cioè i requisiti che determinano le condizioni di appartenenza di individui alla classe. Le restrizioni possono essere di vario tipo, quelle di tipo quantitativo sono caratterizzate da un quantificatore, da una proprietà e da un argomento, e sono: universale o esistenziale.

La restrizione *universale*, *owl:allValuesFrom* nel linguaggio OWL, rappresentata dal simbolo (\forall), identifica l'insieme di tutti gli individui che, data una certa proprietà, hanno *al più* relazioni con individui di una data classe, può essere letta anche come: "ha *solo* relazioni con istanze di quella classe".

La restrizione *esistenziale*, *owl:someValuesFrom* in OWL, rappresentato dal simbolo (\exists), identifica l'insieme di individui che, per una certa proprietà, hanno *almeno* relazioni con individui di una specifica classe, cioè può essere letta come "ha *almeno* una o qualcuna" istanza di quella classe.

Ci sono anche restrizioni di cardinalità (*owl:cardinality*) che indicano il numero di relazioni che un individuo può avere per una data proprietà.

La restrizione, poi, *owl:hasValue* permette di specificare classi sulla base dell'esistenza di un particolare valore della proprietà stessa.

Infine, tra le restrizioni applicabili alle *DatatypeProperty*, si sottolinea la possibilità di definire classi di individui che specificano l'appartenenza del *DatatypeProperty* ad un dato intervallo di valori numerici, utilizzando, per esempio, *xsd:minInclusive* o *xsd:maxInclusive*.

Citiamo anche la proprietà di equivalenza applicata alle classi, utilizzata quando si intende che due classi hanno gli stessi individui. Per esempio, utilizzando la proprietà *owl:equivalentClass* combinata con restrizioni si possono modellare condizioni necessarie e sufficienti di relazioni.

Un'altra proprietà utilizzata è quella di definire due o più classi disgiunte, per esempio *owl:disjointWith* garantisce che un individuo che appartiene ad una classe non può essere anche individuo di altre classi.

2 Mappa semantica degli Open Data Inail sugli infortuni

Obiettivo principale è quello di definire un modello concettuale formalizzato degli *Open Data* (OD) dell'Inail sugli infortuni sul lavoro, del relativo vocabolario e del modello di lettura. Considerati tali elementi di partenza, il lavoro svolto si è sviluppato seguendo una metodologia di tipo *reverse engineering*, che nel settore informatico indica il processo di analisi di un sistema (*software*) che porta come risultato ad una sua descrizione formale intesa come insieme di modelli, teorie, algoritmi.

Partendo quindi dagli OD e dalle loro descrizioni e tenendo conto delle peculiarità che caratterizzano un'ontologia, si è arrivati a fornire una rappresentazione semantica degli *Open Data* degli infortuni. Al termine del processo di definizione di tale ontologia, si è potuta ripercorrere la mappa semantica ritrovando tutti i concetti contenuti nel vocabolario e nel modello di lettura, se presenti negli OD Inail, e le relazioni tra essi, con la differenza che, al posto di una descrizione testuale in

linguaggio naturale, i concetti sono stati definiti mediante formalismi matematici (interpretabile in maniera automatica da un computer).

2.1 Open Data Inail

Gli OD Inail relativi agli infortuni costituiscono il punto di partenza di questo studio in continuità con quanto presentato nel *Quaderno1* [CDFMV-013], indispensabile riferimento per seguire lo sviluppo del lavoro.

Gli OD sono disponibili sul portale dell'Istituto, nell'apposita sezione di Open Data [Inail-013], scaricabili in diversi formati standard (CSV, RDF, XML), e sono organizzati in modo che ciascun record corrisponda ad una denuncia di infortunio e sia composto da una serie di attributi.

Gli OD Inail sugli infortuni, la cui peculiarità principale è proprio quella di essere di granularità fine, ovvero un singolo dato è rappresentato in un singolo record senza alcuna elaborazione o aggregazione, sono divisi per regione e con cadenza semestrale e mensile. *On line* viene fornita la legenda associata agli OD, riportata dal cap. 6.1 del *Quaderno 1*, in cui gli attributi vengono raggruppati in 7 tipi. A ciascun attributo si assegnano una definizione ed una descrizione testuale. Nella pagina dedicata alle operazioni di *download* dei file contenenti gli OD, si forniscono informazioni più specifiche relative ai dati, in particolare per ciascun attributo viene indicato il nome, la descrizione e il tipo informatico di dato (es.: *date, string, integer,...*).

Sotto la voce *Tipologiche* sono raccolte le tabelle di riferimento utilizzate in modo trasversale dagli OD, anch'esse disponibili nei diversi formati standard. Analogamente agli OD, per ciascuna tipologica, oltre che essere presente una descrizione generale, sono definiti gli attributi (nome, descrizione, tipo di dato), rimandando al vocabolario (e thesaurus) per l'elenco e la descrizione dei dati che essa può contenere.

La maggior parte delle tipologiche si riferisce a termini specifici del mondo Inail relativamente agli infortuni (per esempio *Gestione tariffaria, Definizione amministrativa evento*); altre invece sono di carattere più generale e si possono riferire a quelle messe a disposizione anche da altre PA (per esempio *Classificazione Ateco, Province*).

Sul portale Inail, insieme con gli OD, sono disponibili delle tabelle, presenti nel formato *pdf*, che rappresentano i dati aggregati per determinare alcune statistiche sulla numerosità degli infortuni, ragionevolmente ritenute di maggior interesse per l'utente degli *Open Data*. Le tabelle sono organizzate per livelli e per linee tematiche³, e vengono qui citate poiché, quelle derivate da soli dati presenti negli OD, sono state riprodotte attraverso *query* in uno SPARQL Endpoint, appositamente sviluppato in questo progetto, per verificare che tutti i concetti che concorrono a comporre siano stati modellati in *ontoInfortuni*, come vedremo nel capitolo 4.

2.2 Metodologia per la definizione di *ontoInfortuni*

Come detto, nella metodologia adottata si è tenuto conto del documento [AGID-014b], che ha come obiettivo quello di definire una strategia nazionale per indirizzare le PA verso un processo di produzione dei dati standardizzato ed interoperabile. Di particolare interesse sono i capitoli

³ Per le definizioni si rimanda all'allegato 1 del *Quaderno 1*, alle pagine 41-46 per le tabelle "annuali" e alle pagine 100-102 per le tabelle di "periodo"; oppure agli omonimi link nel portale degli *Open Data* INAIL.

riguardanti le ontologie ed i vocabolari disponibili, condivisi a livello europeo, che possono essere utilizzati. Si sottolinea che l'utilizzo di questi vocabolari nella descrizione degli *open data* può far sì che “*all'interno della PA, dataset differenti ma di una stessa tipologia abbiano uno schema comune e siano pertanto rappresentati in maniera omogenea*”⁴.

Inoltre, si evidenzia la distinzione tra dati *trasversali* e dati *verticali*, i primi sono dati comuni a tutte le amministrazioni, centrali e locali, gli altri sono invece prodotti e gestiti da specifiche amministrazioni. Dall'analisi degli OD Inail, si evince che sono presenti entrambi i tipi di dati, tuttavia per quelli trasversali, quali ad esempio settore economico Ateco, Provincia o Nazione di nascita dell'infortunato, si è preferita una rappresentazione sviluppata ad hoc per gli specifici riferimenti negli *open data*.

Nel processo di costruzione di *ontoInfortuni*, si evidenziano le scelte fatte e le difficoltà incontrate che, in particolare all'inizio del lavoro, hanno spesso fatto cambiare, e a volte anche abbandonare, le strade inizialmente intraprese.

2.3 Processo di definizione e formalizzazione del modello concettuale

Non esiste un percorso “corretto” ed univoco oppure una metodologia privilegiata per definire un'ontologia; il processo di costruzione dipende dalle caratteristiche del dominio che si vuole rappresentare [NMG-001, §3] e dall'obiettivo che si vuole raggiungere con tale sviluppo. Quando è possibile rappresentare il dominio in tassonomie, l'ontologia può svilupparsi attraverso la costruzione di una corrispondente rappresentazione gerarchica di classi, alle quali vengono associate relazioni con opportune proprietà.

Gli esperti di ontologie [NMG-001, §3] indicano che il processo di definizione di un'ontologia è sempre molto complesso ed iterativo e, proprio perché non esiste un metodo migliore di altri, suggeriscono di utilizzare, combinandoli fra di loro, i due approcci: *bottom-up* e *top-down*.

L'attività di costruzione dell'ontologia relativa agli infortuni si è sviluppata in entrambe le direzioni. Infatti, partendo dai dati (tipologiche e *Open Data*) e dalle singole definizioni (vocabolario) si sono individuati, con approccio *bottom-up*, i concetti da modellare e le istanze da rappresentare. Studiando diverse scelte di modellazione, alcune si presentavano alternative fra di loro, in particolare nella scelta delle classi e delle relazioni, ma comunque tutte utili, in una analisi preliminare, per poter valutare la strategia più confacente agli obiettivi posti.

Al termine di questa analisi *bottom-up* restava, tuttavia, ancora non completamente esplicita la visione di insieme del problema, sicuramente chiara agli addetti ai lavori, ma per le autrici di questo *Quaderno* diventava estremamente importante individuare una struttura logica generale, individuare cioè i concetti cardine attorno ai quali ruotano tutti gli altri.

A tal proposito è stata esplicativa la classificazione dei dati elementari per la lettura annuale, indicata nel capitolo 6.1 del *Quaderno1*, che evidenzia quali siano gli elementi principali di riferimento: caso d'infortunio, infortunato, datore di lavoro.

Tutti e tre i concetti sono presenti negli OD, mediante degli identificativi, mentre l'unica definizione presente nel vocabolario è quella relativa al “caso d'infortunio”. Gli altri due concetti non sono definiti in chiaro nel vocabolario perché la loro presenza è implicita nel dominio descritto in quanto indispensabili per “creare” l'evento infortunio.

⁴ [AGID-014b] pag. 67, §6.3

Mediante l'approccio *top-down* si è perciò modellata l'ontologia a partire da questi tre concetti fondamentali, definendo le tre rispettive classi e le relazioni che li legano tra loro e alle tipologiche.

2.3.1 Processo bottom-up

Il vocabolario relativo agli infortuni è composto di 69 lemmi, di ciascuno si dà la definizione e si indicano le possibili relazioni (ascendenti o discendenti) con altri lemmi⁵. Sulla base di queste caratteristiche, in un primo momento, si è creduto che fosse sufficiente (ed esaustivo) rappresentare ogni concetto contenuto nel vocabolario come *classe* e trattare le *relazioni* di ascendenza e discendenza come relazioni gerarchiche.

Da un'analisi più dettagliata delle definizioni contenute nel vocabolario e dal confronto con i campi presenti nei tracciati record degli OD è emerso che alcuni lemmi, sebbene presenti nel vocabolario, non figurano, invece, negli OD poiché nel "linguaggio speciale" Inail che si è definito sono presenti tutte le voci che permettono la lettura, oltre che dei dati elementari, anche delle tabelle di dati aggregati, o che chiariscono alcuni aspetti gestionali e amministrativi che non sono esplicitamente presenti nei record del *dataset* sugli infortuni (ad esempio *classe d'età* o *giorni di inabilità* entrambi utilizzati nelle tabelle di dati aggregati). Non essendo tali lemmi direttamente rappresentati negli OD non sono stati modellati nell'ontologia.

Riguardo i lemmi che sono rappresentati negli OD, di fatto essi corrispondono sia ad un concetto (*classe*) sia ad una relazione (*proprietà*) con altri elementi.

Un esempio è il lemma *indennizzo* che viene definito nel vocabolario come segue:

“è la prestazione economica che l'Inail corrisponde agli infortunati; sono 4 le modalità principali con cui viene erogata la prestazione: *in temporanea, in capitale, in rendita diretta e in rendita a superstiti*”

Il lemma *indennizzo* è messo in relazione (discendente) con i lemmi *in temporanea, in capitale, in rendita diretta, in rendita a superstiti*, ciò è come dire, nella modellazione dell'ontologia, che la classe *indennizzo* si articola in quattro sottoclassi, questo soddisfa la seconda parte della definizione sopra indicata. La prima parte della definizione indica che il concetto *indennizzo* è in relazione con l'*infortunato*. D'altra parte esaminando le descrizioni degli attributi contenuti negli OD si legge che il metadato *indennizzo* è:

“il codice della tipologia di indennizzo
(se *in temporanea, in capitale, in rendita diretta e in rendita a superstiti*)”

Infine, esaminando direttamente gli OD, si osserva che al metadato <*Indennizzo*> si associa uno dei seguenti codici: *TE, CA, RD, RS, NE*.

Riassumendo, il concetto *indennizzo* assume diversi ruoli (classe, relazione, istanza) a seconda del contesto in cui viene considerato (descrizione generale, rapporto con l'infortunato, codice). Ciò risulta chiaro se espresso in linguaggio naturale, ma dal punto di vista di una formalizzazione del modello concettuale l'ambiguità generata dai diversi ruoli di tale concetto deve essere risolta.

⁵ Da pag. 13 a pag. 24 di [CDFMV-013].

Da questo tipo di analisi emerge, quindi, che non è sufficiente tener conto solo del vocabolario, ma si deve avere presente anche la descrizione dei metadati contenuti negli OD, e solo dal confronto di entrambe le definizioni si può decidere come modellare i concetti.

Un esempio di tale affermazione è rappresentato dall'*età*. Il concetto viene definito nel vocabolario come segue:

“è l'età dell'infortunato alla *data di accadimento* dell'infortunio”

Definita allo stesso modo nella descrizione dei metadati, corrisponde ad un dato di tipo intero che, anche se dal punto di vista logico sembra essere in relazione con l'infortunato, di fatto rappresenta l'età dell'infortunato al momento (*data di accadimento*) dell'infortunio, ma l'effettiva relazione è tra il valore che rappresenta l'età e la data di accadimento dell'infortunio rispetto a cui è stato calcolato, pertanto è stato modellato come *DatatypeProperty*, cioè un attributo di tipo *integer*, da assegnare all'*infortunio*.

2.3.2 Processo top-down

Per avere una visione generale degli OD relativi agli infortuni, riprendiamo quanto detto nel §2.1 riguardo la legenda degli attributi inserita nel capitolo 6.1 del *Quaderno1*, e per chiarezza di trattazione, riportiamo in tabella 1 i sette tipi sotto cui essi sono raggruppati:

Tabella 1 - raggruppamento degli attributi in 7 tipi

A – localizzazione temporale dell' infortunio
B – localizzazione geografica dell' infortunio
C – caratteristiche dell' infortunato
D – modalità dell' infortunio
E – caratteristiche amministrative dell' infortunio
F – caratteristiche medico-legali dell' infortunio
G – caratteristiche del datore di lavoro

In grassetto sono stati evidenziati i concetti fondamentali di riferimento, senza i quali si snaturebbero gli stessi OD: infortunio, infortunato, datore di lavoro. Volendo formalizzare tale rappresentazione, ricordando che gli ultimi due termini sono presenti nei record ma non nel vocabolario, questi devono essere modellati esplicitamente, difatti, come detto, negli OD sono presenti e sono rappresentati univocamente tramite gli identificativi.

Il “caso d'infortunio” risulta essere il concetto imprescindibile in questo dominio; è indipendente dal riconoscimento o meno dell'infortunio come accaduto sul lavoro, aggiunge un nuovo record nel *dataset* degli infortuni nel momento in cui viene registrata la denuncia, il suo identificativo è unico in tutto il *dataset*. Il concetto di “infortunato” è obbligatoriamente presente ma “discende” dal caso d'infortunio, è inteso come oggetto del fatto accaduto; l'identificativo ad esso associato è univoco rispetto alla persona ma può avere una molteplicità di ricorrenze nel *dataset* se a quell'identificativo sono accaduti più infortuni. Il concetto di “datore di lavoro”, infine, può non essere presente, ma quando lo è l'identificativo associato è anch'esso univoco e può comparire più volte nel *dataset* se si sono verificati più infortuni collegati allo stesso datore di lavoro.

3 L'ontologia ontoInfortuni

In questo capitolo si presenta *ontoInfortuni*, il modello che meglio risponde ai requisiti individuati con l'analisi dei dati e dei documenti per rappresentare gli OD sugli infortuni. Si illustrano il tipo di organizzazione utilizzata per le ontologie che compongono *ontoInfortuni* e le convenzioni adottate per modellarle.

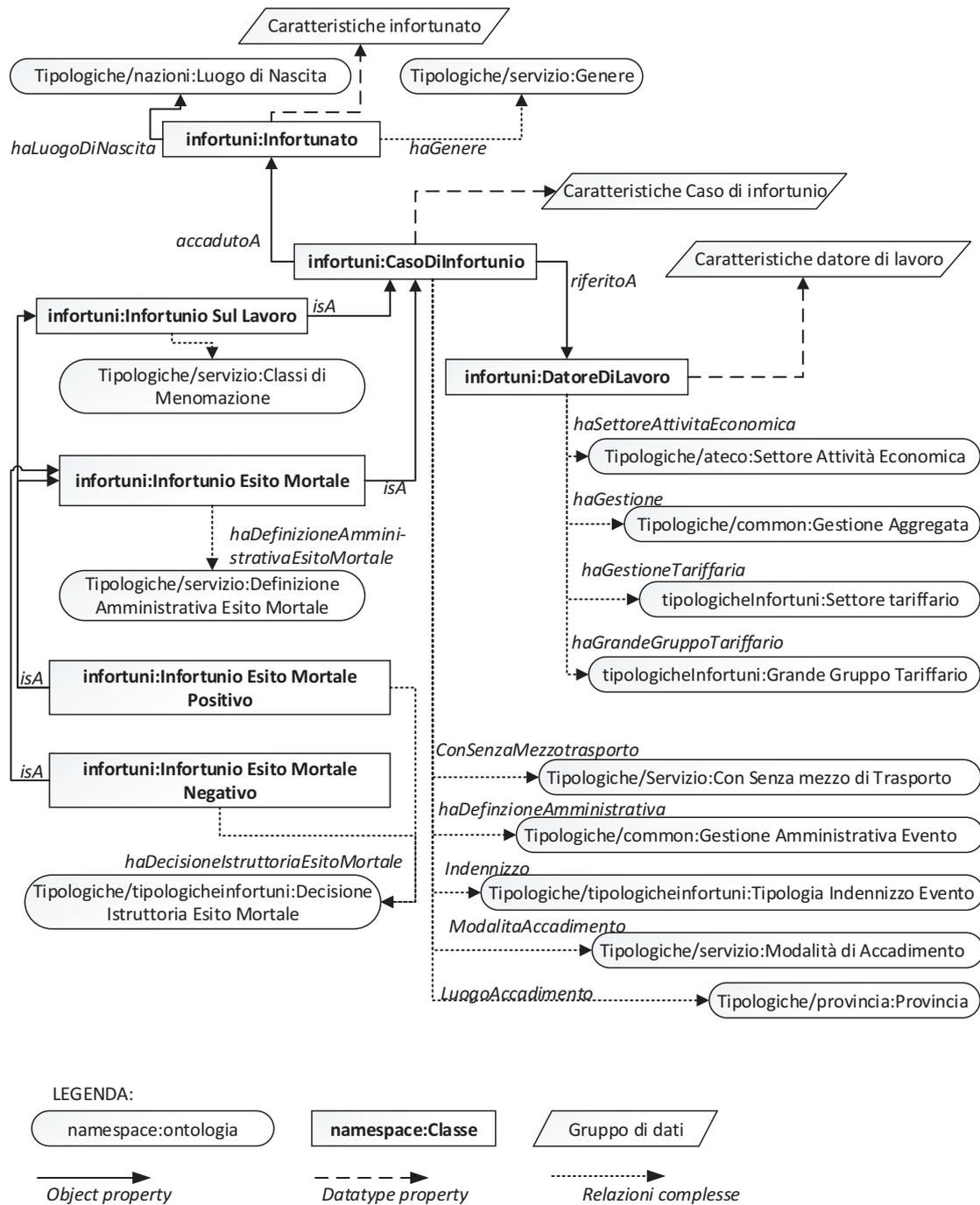


Figura 1 - Diagramma di *ontoInfortuni*

Il diagramma in figura 1, oltre a dare la visione d'insieme dello sviluppo realizzato, in una modalità più consueta e di agevole lettura, può essere utilizzato come riferimento sia per collocare quanto appena detto sulle caratteristiche degli OD modellati, sia per seguire le spiegazioni successive.

In fondo al diagramma si trova una legenda che permette di ricollegare agli elementi propri del linguaggio software di sviluppo quelli riconducibili agli analoghi negli OD sugli infortuni grazie alle convenzioni sui nomi adottate.

Occorre precisare che lo sviluppo che verrà illustrato si occupa solo ed esclusivamente di quei concetti, richiamati dalla scelta degli autori del *Quaderno1*, che rappresentano un sottoinsieme significativo di informazioni rispetto al *dataset* completo sugli infortuni, da cui gli *Open Data* vengono estratti.

Per gli scopi istituzionali tale *dataset* si compone anche di elementi di carattere amministrativo, giuridico, economico, medico che non sono negli OD e quindi non sono oggetto di questo *Quaderno*. Inoltre, quelli appena citati, sono dei domini complessi, che meriterebbero una modellazione indipendente, e che, attraverso concetti chiave, possono correlarsi ed integrare *ontoInfortuni*.

Si potrà pensare di svilupparli in futuro, sfruttando appieno la capacità delle ontologie di creare dei percorsi di conoscenza specifici che colleghino informazioni appartenenti ad ambiti differenti e che sono in grado di classificare e collocare nella corretta posizione le nuove informazioni e quindi i nuovi casi d'infortunio che si aggiungono.

3.1 Architettura e convenzioni

In figura 2 è illustrata l'architettura creata tenendo conto dei risultati dei due processi *top-down* e *bottom-up* e delle altre considerazioni descritte. Si sono individuati sette gruppi al cui interno sono state definite specifiche ontologie. L'ontologia centrale è quella portante, le altre relative alle tipologiche concorrono alla completa caratterizzazione, secondo i criteri individuati, del dominio degli infortuni.

Si è ritenuto, infatti, di dover rappresentare in modo distinto, vale a dire con modelli separati, le tipologiche definite per gli OD. Tale scelta offre il vantaggio di renderle indipendenti e riutilizzabili in altri contesti; volendo, per esempio, definire una mappa semantica per gli *open data* relativi alle malattie professionali, si possono usare le stesse ontologie laddove vengano adottate le stesse tipologiche.

I diversi colori evidenziano quale sia la PA responsabile dei dati, distinguendo tra tipologiche proprie del mondo Inail create per esprimere dati di propria responsabilità, almeno dal punto di vista della proprietà del dato e della sua certificazione, e delle altre PA: in blu Inail, in rosso Istat, in giallo e in verde sono indicate le tipologiche di carattere geografico i cui riferimenti sono nel portale SPCData nella sezione dei *linked open data*.

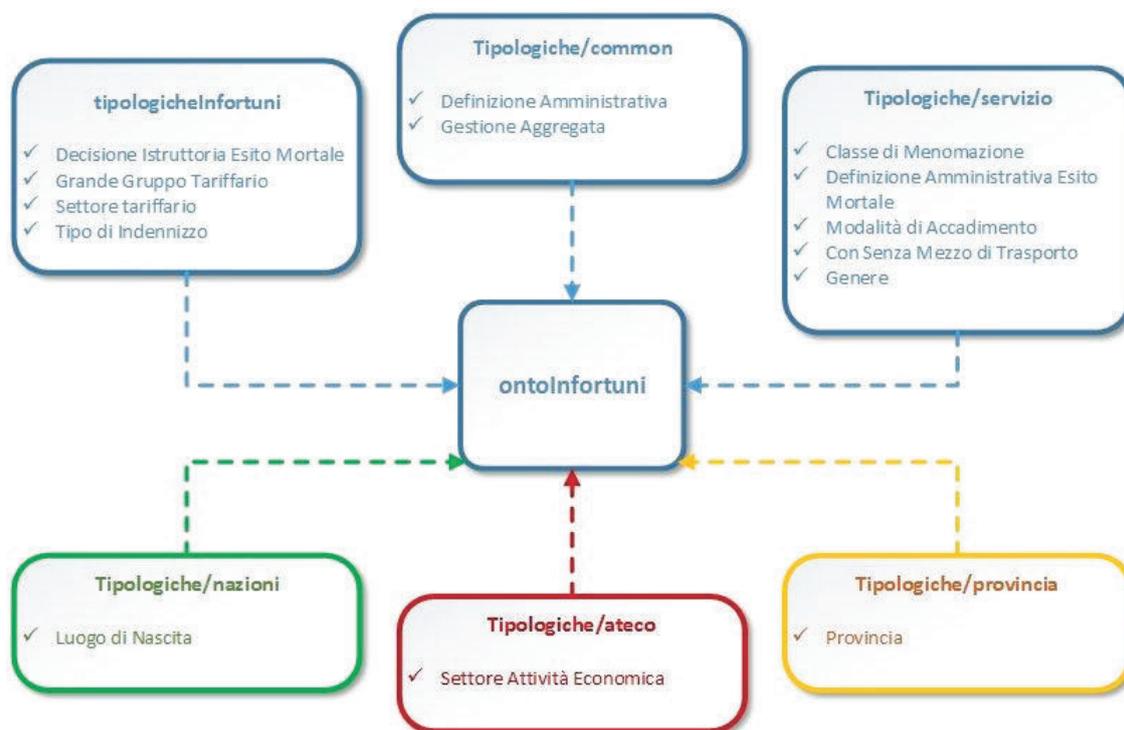


Figura 2 - Architettura di *ontoInfortuni*

Le tipologie nei riquadri blu, quelle Inail, sono raggruppate in base ai seguenti criteri:

- *tipologicheinfortuni*, riferiscono a concetti propri del modello di infortuni, quali: Decisione Istruttoria Esito Mortale, Grande Gruppo Tariffario, Settore Tariffario, Tipologia di Indennizzo;
- *common*, riferiscono a concetti presenti anche nel modello delle Malattie Professionali (non oggetto di questo lavoro) e che quindi potrebbero essere utilizzati nella relativa ontologia. Esse sono: Definizione Amministrativa, Gestione Aggregata;
- *servizio*, sono quelle modellate sulla base delle descrizioni contenute nel vocabolario, ma che non sono state esplicitamente rappresentate in tabelle a supporto degli OD. Esse sono: Definizione Amministrativa Esito Mortale, Genere, Modalità di Accadimento, Con Senza Mezzo di Trasporto, Classe di Menomazione.

Le altre tipologie sono state definite in modelli separati poiché, nell'ottica di una reale digitalizzazione delle Pubbliche Amministrazioni, potrebbero essere sostituite con quelle di riferimento pubblicate sui siti *open data* della PA.

L'ontologia centrale relativa agli infortuni è *ontoInfortuni*, in cui sono modellate le classi definite fondamentali, ovvero caso d'infortunio, infortunato, e datore di lavoro, e tutte le relazioni che le connettono tra loro e con le tipologie.

La seguente tabella 2 illustra il nome assegnato a ciascuna ontologia, il *namespace* completo, il nome del relativo file OWL.

Tabella 2 - Ontologie identificate con prefissi, *Namespace*, nome file

Prefix	Namespace	File (.owl)
infortuni	/RicercaInail.ontologie/OpenData/ontoInfortuni#	ontoInfortuni
ateco	/RicercaInail.ontologie/OpenData/Tipologiche/ateco#	ateco
provincia	/RicercaInail.ontologie/OpenData/Tipologiche/provincia#	provincia
tipologicheinfortuni	/RicercaInail.ontologie/OpenData/Tipologiche/tipologicheinfortuni#	tipologicheinfortuni
common	/RicercaInail.ontologie/OpenData/Tipologiche/common/inail#	common
servizio	/RicercaInail.ontologie/OpenData/Tipologiche/servizio#	servizio
nazioni	/RicercaInail.ontologie/OpenData/Tipologiche/nazioni#	nazioni

Per definire le ontologie si sono adottate delle convenzioni di tipo sintattico quali: il nome di una classe inizia sempre con la lettera maiuscola, le relazioni (*DatatypeProperty* e *ObjectProperty*) sono espresse, nella maggior parte dei casi, come verbo e con la lettera iniziale minuscola. Per la scelta dei nomi da attribuire agli elementi dell'ontologia si è deciso di adottare, per quanto possibile e se non risultano ambigui, nomi uguali ai relativi metadati presenti negli OD o ai lemmi descritti nel vocabolario.

Per evitare le ambiguità nella corrispondenza tra gli elementi dell'ontologia e i concetti espressi nel vocabolario e negli *open data*, sono state introdotte delle annotazioni che riportano esattamente le frasi dei documenti di riferimento.

Le annotazioni utilizzate sono dei seguenti tipi:

- *label*: indica il nome dell'elemento per esteso (sempre presente);
- *comment*: contiene la descrizione, laddove presente, del concetto secondo quanto riportato nel vocabolario del *Quaderno 1*, in cui si fornisce la descrizione dei dati elementari per la lettura annuale;
- *description*: è la descrizione del metadato riportata sul portale Inail;
- *notation*: corrisponde al valore di metadato espresso negli *open data*;
- *seeAlso*: indica i riferimenti normativi quando sono esplicitati nel *Quaderno 1*.

La definizione di *ObjectProperty* mette in relazione elementi appartenenti a classi diverse; quando è chiaro quali siano le classi *dominio* e *range*, queste sono definite esplicitamente. Analogamente per quanto riguarda le *DatatypeProperty* si indica il dominio e il tipo di dato. Infine, si è deciso di considerare come *istanze* (o *individui*) tutti i valori indicati nelle tipologiche.

3.2 Modello proposto per *ontoInfortuni*

Lo sviluppo di *ontoInfortuni* parte dalla modellazione delle tre classi di riferimento (caso di infortunio, infortunato, datore di lavoro) e dalla definizione delle relazioni (*ObjectProperty* e *DatatypeProperty*), applicabili a ciascun elemento delle classi rispetto alle tipologiche⁶.

⁶ Sulle tipologiche si trova in dettaglio, nel paragrafo 3.3, la descrizione sulle convenzioni adottate e sulla loro formalizzazione.

Il *CasoDiInfortunio* è la classe indispensabile per la definizione degli infortuni. Questa classe è caratterizzata da una relazione di equivalenza sulla *DatatypeProperty* corrispondente all'*IdentificativoCaso*.

$$Individual \in CasoDiInfortunio \Leftrightarrow IdentificativoCaso \text{ not Null}$$

Dalle caratteristiche relative al caso d'infortunio elencate nel *Quaderno1* [pag. 32] si evince che alcune sono *DatatypeProperty* altre invece sono relazioni con le tipologiche (*ObjectProperty*). Relazioni *DatatypeProperty* sono quelle che corrispondono alle "localizzazioni temporali" dell'infortunio, ossia le date; alle "caratteristiche medico-legali" dell'infortunio, il grado di menomazione e i giorni indennizzati; all'età dell'infortunato alla data di accadimento dell'infortunio. Tutte queste relazioni prendono il nome direttamente dal metadato utilizzato negli OD, con la restrizione sulla cardinalità pari a 1, possono quindi avere un solo valore.

Le relazioni *ObjectProperty* mettono in corrispondenza individui della classe *CasoDiInfortunio* con alcune tipologiche (tabella 3), come dedotto da alcuni dei tipi (B, D, E) del modello di lettura:

Tabella 3 - Corrispondenza tra tipo di dato e relazione con la tipologica

Tipi di dato	Relazioni e restrizioni
Luogo di accadimento	\exists <i>LuogoAccadimento Class_Provincia</i>
Modalità di accadimento	\exists <i>ModalitaAccadimento MA_ModalitaAccadimento_Tipo</i>
Con/senza mezzo di trasporto	\exists <i>ConSenzaMezzoTrasporto TR_ConSenzaMezzoTrasporto_Tipo</i>
Definizione amministrativa	\exists <i>haDefinizioneAmministrativa</i> <i>DA_DefinizioneAmministrativaEvento_Tipo</i>
Indennizzo	\exists <i>Indennizzo TI_TipologiaIndennizzoEvento_Tipo</i>

Come indicato nel vocabolario, il caso di infortunio è l'infortunio registrato, pertanto è stato modellato come radice di una breve struttura gerarchica, a grafo e non ad albero come intuitivamente potrebbe apparire, illustrata in figura 3.

La classe *CasoDiInfortunio* ha due sottoclassi: *InfortunioEsitoMortale* e *InfortunioSulLavoro*, che ereditano le relazioni definite nella classe padre.

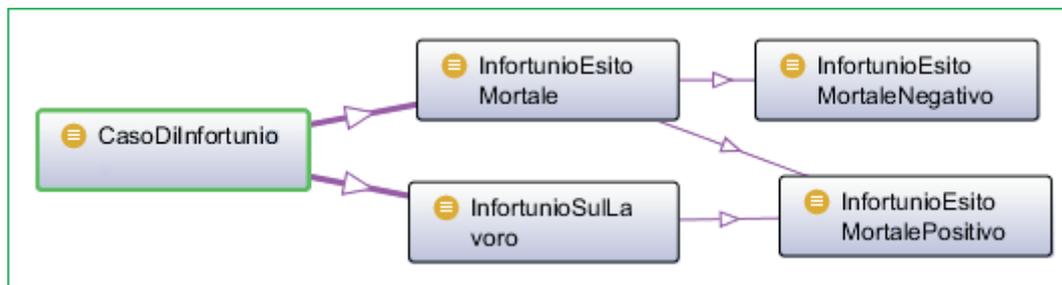


Figura 3 - Struttura gerarchica a grafo del caso di infortunio

La classe *InfortunioSulLavoro* è stata modellata come classe equivalente al concetto che abbia definizione amministrativa positiva, cioè:

$$\text{Individual} \in \text{InfortunioSulLavoro} \Leftrightarrow \text{haDefinizioneAmministrativa hasValue } \{DA_P\}$$

Dove *DA_P* è l'istanza corrispondente al codice positivo della definizione amministrativa. Inoltre sono state aggiunte due relazioni di tipo *DatatypeProperty*, quali *GiorniIndennizzati* e *GradoMenzione*, sempre con la restrizione sulla cardinalità ad un solo valore. Si noti che queste due relazioni sono definite solo per individui che appartengono a questa classe.

La classe *InfortunioEsitoMortale* intende raggruppare tutti quegli individui che hanno una data di morte, questo concetto è modellato come segue:

$$\text{Individual} \in \text{InfortunioEsitoMortale} \Leftrightarrow \text{DataMorte not Null}$$

Inoltre gli individui che appartengono a questa classe hanno relazione con la classe di definizione amministrativa esito mortale di tipo universale:

$$\forall \text{haDefinizioneAmministrativaEsitoMortale} \\ \text{EM_DefinizioneAmministrativaEsitoMortale_Tipo}$$

A sua volta la classe *InfortunioEsitoMortale* è ancora stata suddivisa in *InfortunioEsitoMortaleNegativo* e *InfortunioEsitoMortalePositivo*, in base agli individui relativi alla classe *DecisioneIstruttoriaEsitoMortale*. In particolare la classe *InfortunioEsitoMortaleNegativo* è modellata come:

$$\text{Individual} \in \text{InfortunioEsitoMortaleNegativo} \Leftrightarrow \\ \exists \text{haDecisioneIstruttoriaEsitoMortale } (\{DI_21, DI_22, DI_24, DI_244, \\ DI_27, DI_28, DI_300\})$$

Ciò significa che la condizione necessaria e sufficiente perché un'istanza appartenga alla classe *InfortunioEsitoMortaleNegativo* è che abbia “causa di negatività per l'esito mortale”, vale a dire che il valore della proprietà *haDecisioneIstruttoriaEsitoMortale* sia uno di quelli elencati nella restrizione.

Infine, è stata assegnata la seguente restrizione alla proprietà *haDefinizioneAmministrativaEsitoMortale*:

$$\forall \text{haDefinizioneAmministrativaEsitoMortale } (\{EM_N\})$$

Dove *EM_N* è l'istanza corrispondente al codice che indica la definizione amministrativa dell'esito mortale negativa.

In modo del tutto analogo si è definita la classe *InfortunioEsitoMortalePositivo*:

$$\text{Individual} \in \text{InfortunioEsitoMortalePositivo} \Leftrightarrow \\ \exists \text{haDecisioneIstruttoriaEsitoMortale } (\{DI_15, DI_16, DI_17, DI_18\})$$

Con la restrizione sul codice positivo:

$$\forall \text{haDefinizioneAmministrativaEsitoMortale } (\{EM_P\})$$

Infine, le classi (*InfortunioEsitoMortalePositivo* e *InfortunioEsitoMortaleNegativo*) sono state definite disgiunte fra di loro mediante la proprietà *owl:disjointWith*.

La definizione della condizione necessaria e sufficiente relativa alle classi di infortunio di esito mortale, negativo o positivo, sulla base del valore assegnato alla causa di negatività o positività è solo una delle possibili scelte. Infatti, un'altra possibilità è quella di definire la condizione necessaria e sufficiente rispetto alla definizione amministrativa, nel caso di definizione negativa risulterebbe:

$$\begin{aligned} & \text{Individual} \in \text{InfortunioEsitoMortaleNegativo} \Leftrightarrow \\ & \text{haDefinizioneAmministrativaEsitoMortale hasValue } (\{EM_N\}) \end{aligned}$$

Assegnando alla classe *InfortunioEsitoMortaleNegativo* la seguente restrizione:

$$\begin{aligned} \exists \text{haDecisioneIstruttoriaEsitoMortale } (\{DI_21, DI_22, DI_24, DI_244, \\ DI_27, DI_28, DI_300\}) \end{aligned}$$

In modo del tutto analogo, la classe *InfortunioEsitoMortalePositivo* si può modellare adottando la stessa strategia. Dal punto di vista del modello le due scelte sembrano essere equivalenti, d'altra parte nel vocabolario non è emerso nulla che facesse prediligere una scelta piuttosto che l'altra. Si può notare, quindi, che le classi *InfortunioEsitoMortale* e *InfortunioSulLavoro* sono due classi, figlie della stessa classe *CasoDiInfortunio*, che non possono essere disgiunte, infatti, ci sono istanze di casi mortali che sono riconosciuti come infortuni sul lavoro, cioè appartengono ad entrambe le classi.

La figura 4 illustra il grafo relativo al caso di infortunio e le relazioni illustrate in precedenza rispetto alle tipologiche.

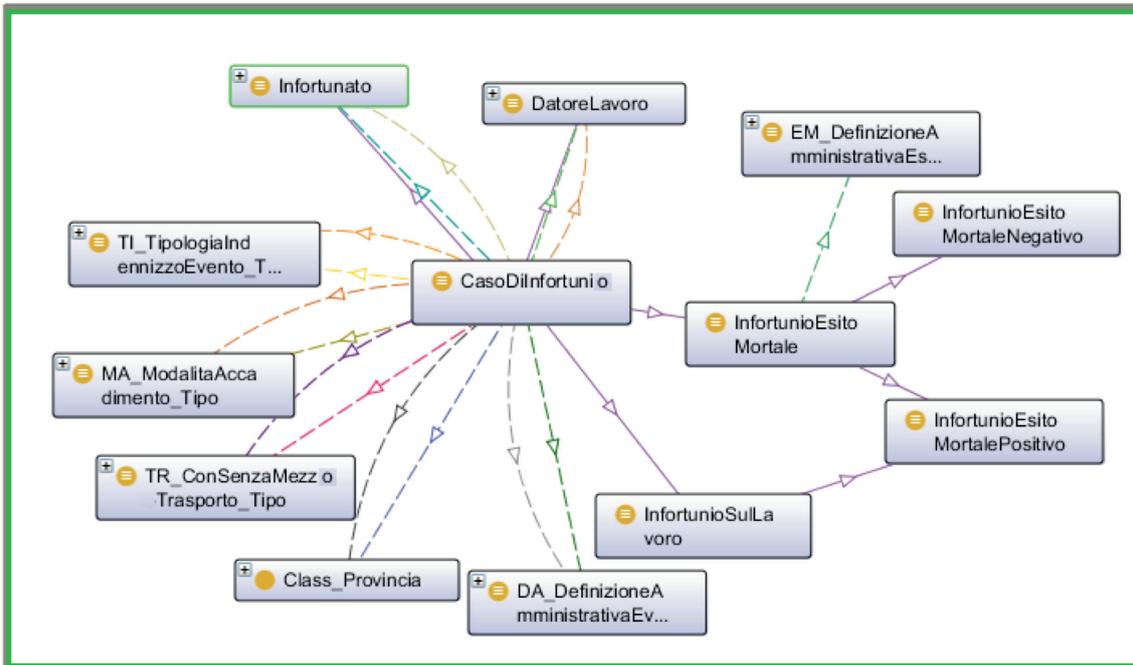


Figura 4 - Grafo di rappresentazione del *CasoDiInfortunio*

Come è stato evidenziato in tabella 1 (nel paragrafo 2.3.2), sono tre i concetti fondamentali sui quali si impenna e sviluppa l'ontologia degli infortuni, ma le relazioni che li collegano non sono dichiarate.

Nell'ontologia invece le relazioni tra elementi sono state esplicitate ed il fulcro è il caso di infortunio. Gli altri due concetti, datore di lavoro e infortunato, rappresentano, rispettivamente, parte del contesto a cui si fa riferimento e la persona a cui l'infortunio è accaduto. Per definire il modo in cui si rapportano, sono state modellate due relazioni, entrambe con dominio *CasoDiInfortunio*: la relazione *referitoA* con range *DatoreLavoro*, e la relazione *accadutoA* con range *Infortunato*. Queste due relazioni sono rappresentate in evidenza nella figura 5.

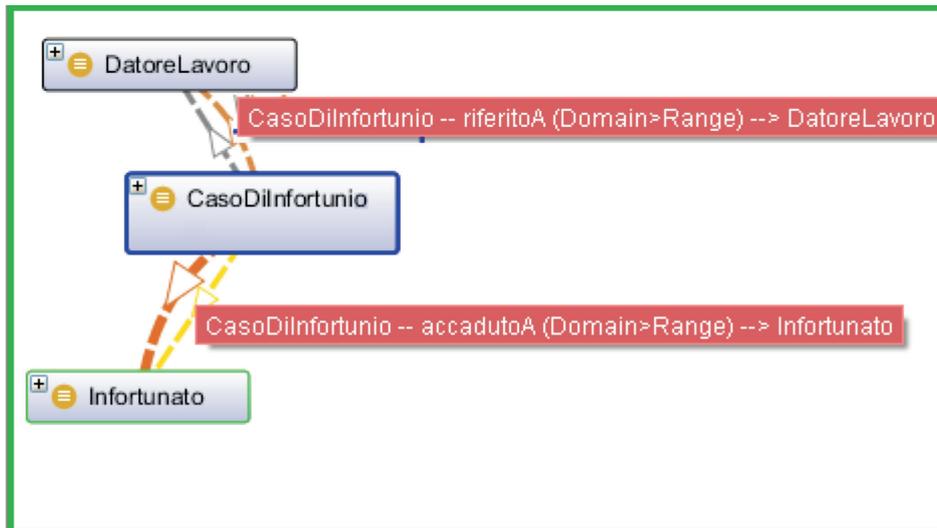


Figura 5 - Relazioni tra le tre classi principali.

Con questa scelta un'istanza della classe *DatoreLavoro* sarà univoca, in base al suo codice identificativo, indipendentemente dai casi di infortunio a cui sarà associata (*referitoA*). Analogamente l'individuo *Infortunato* è anch'esso identificato univocamente da un codice che sarà messo in relazione con il caso di infortunio (*accadutoA*). In tal modo si possono riscontrare più casi di infortunio che riferiscono ad un solo datore di lavoro, oppure che sono accaduti allo stesso lavoratore (infortunato). Anche utilizzando gli OD si ritrova lo stesso risultato, con la differenza che viene dato come insieme di record in cui verificare l'uguaglianza del codice, mentre nel caso di *ontoInfortuni* si tratta del riferimento ad un solo individuo.

Un quadro d'insieme che illustra l'ontologia rappresentata graficamente attraverso le classi principali e le relative relazioni è nella figura 6.

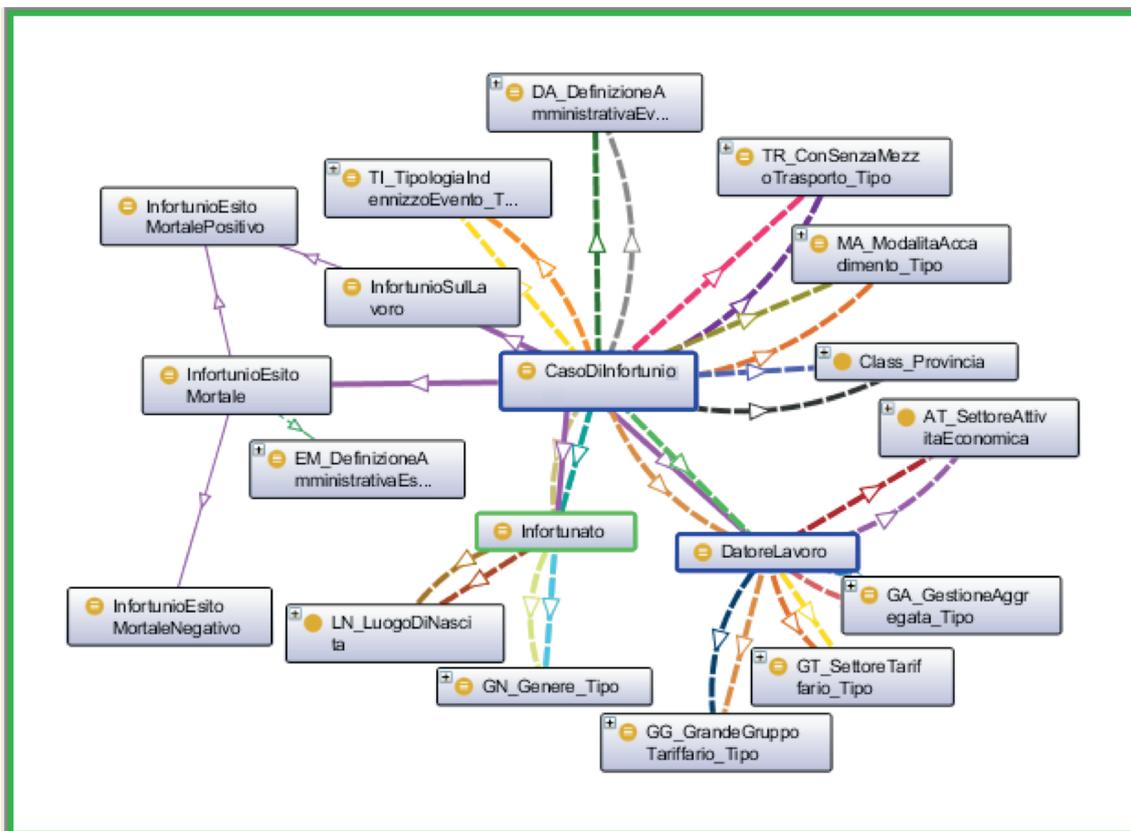


Figura 6 - *ontoInfortuni* completa di tutte le classi principali

Il concetto successivo considerato nel modello è *Infortunato*; si tratta di una classe in cui ciascun elemento è caratterizzato univocamente da un identificativo (*IdentificativoInfortunato*), mediante la seguente relazione di equivalenza:

$$Individual \in Infortunato \Leftrightarrow IdentificativoInfortunato \text{ not Null}$$

Questo significa che ciascun individuo di *Infortunato* ha un valore per la proprietà *IdentificativoInfortunato*, viceversa qualunque istanza abbia valore con la proprietà *IdentificativoInfortunato* apparterrà alla classe *Infortunato*.

L'infortunato è messo in relazione con le classi che descrivono il genere e il luogo di nascita, definite nelle relative tipologiche (vedi figura 7) come segue:

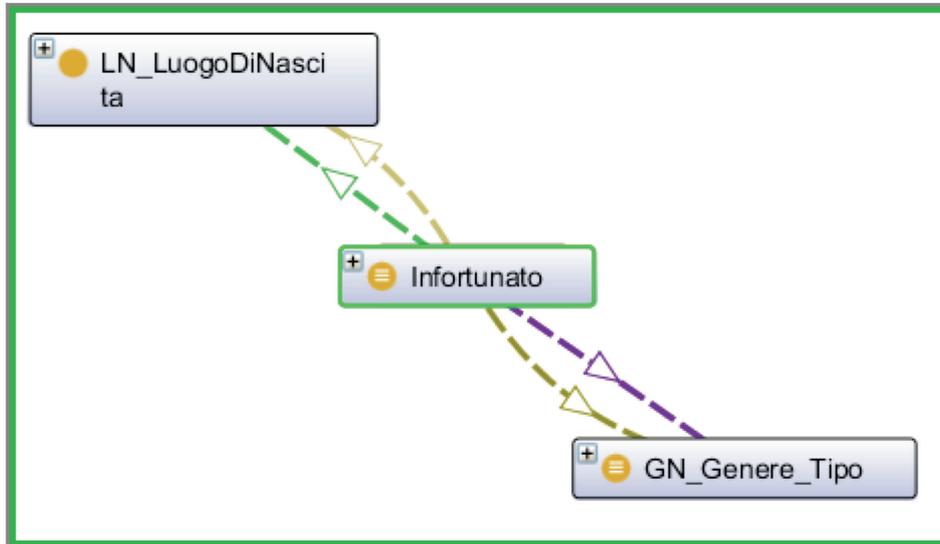


Figura 7- La classe *Infortunato* in relazione con le tipologiche *Luogo di nascita* e *Genere*

Analogamente si definisce la classe *DatoreLavoro*, come illustrato in figura 8, anch'essa caratterizzata da un identificativo (*IdentificativoDatoreLavoro*), che, quando presente, ha analogha relazione di equivalenza:

$$\text{Individual} \in \text{DatoreLavoro} \Leftrightarrow \text{IdentificativoDatoreLavoro not Null}$$

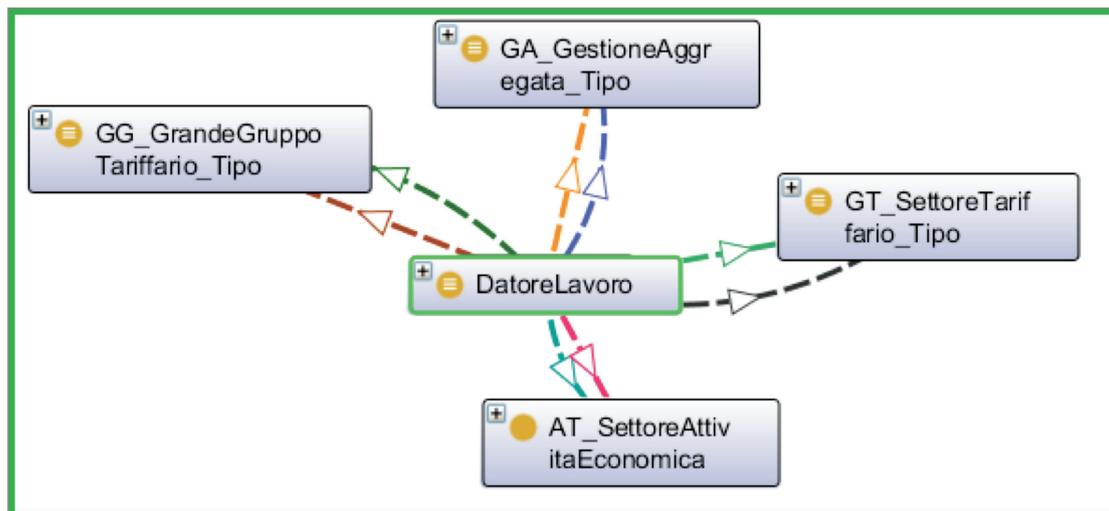


Figura 8 - La classe *DatoreLavoro* è in relazione con le tipologiche

Gli individui della classe *DatoreLavoro* hanno relazioni con le tipologiche: Grande Gruppo Tariffario, Settore Tariffario, Gestione Aggregata, codice Ateco o Settore Attività Economica. In particolare:

(\exists *haGestione* *GA_GestioneAggregata_Tipo*)

(\exists *haGestioneTariffaria GT_SettoTariffario_Tipo*)
 (\exists *haGrandeGruppoTariffario GT_GrandeGruppoTariffario_Tipo*)
 (\exists *haSettoAttivitaEconomica AT_SettoAttivitaEconomica_Tipo*)
 (*Cardinality(PosizioneAssicurativaTerritoriale = 1)*)

3.3 Tipologiche

Insieme agli OD, sono state pubblicate, negli stessi formati standard, le *Tipologiche*, cioè tabelle a corredo che contengono l'esplicitazione e la descrizione dei codici corrispondenti a specifici concetti utilizzati negli OD.

Anche per le tipologiche il processo adottato è stato quello indicato in precedenza, ciascuna di esse è stata modellata come singola rappresentazione tabellare (modalità *bottom-up*), poi si è individuato a quale concetto fondamentale si relazionasse (approccio *top-down*).

Analizzando le tipologiche si è visto che alcune sono tabelle semplici, cioè corrispondenti ad un solo concetto che assume diversi valori. Altre tipologiche si presentano invece più complesse e hanno richiesto un maggiore approfondimento e l'adozione di soluzioni differenti.

Il nome delle istanze relative alle tipologiche viene composto da un acronimo di due lettere che individua la tipologica (in tabella 4), più il valore rappresentato, per esempio con *GG_1* si intende il codice 1 della tipologica Grande Gruppo Tariffario (GG).

Tabella 4 - Elenco dei codici adottati in riferimento alle tipologiche

Acronimo	Tipologica
AT	Setto Attività Economica (Ateco)
CM	Classe di Menomazione
DA	Definizione Amministrativa
DI	Decisione Istruttoria Esito Mortale
EM	Definizione Amministrativa Esito Mortale
GA	Gestione Aggregata
GG	Grande Gruppo Tariffario
GN	Genere
GT	Setto Tariffario
LN	Luogo di Nascita
MA	Modalità Accadimento
PR	Province
TI	Tipologica Indennizzo Evento
TR	Con Senza Mezzo Trasporto

3.3.1 Modellazione di tipologiche semplici

Tra le Tipologiche vi sono delle tabelle considerate semplici da modellare poiché corrispondono ad un solo *concetto* con due proprietà, il codice e la descrizione. Ciascun elemento della tipologica è stato rappresentato perciò come una istanza della classe tipologica di riferimento.

Sul portale Inail, al *link* degli *open data*, alla voce "Grande gruppo tariffario in cui ricade l'infortunio" delle Tipologiche, è disponibile la tabella relativa alla "*decodifica del grande gruppo tariffario*"

relativo alla PAT su cui ricade l'infortunio", ogni record contiene due metadati, uno relativo al codice <GrandeGruppoTariffario>, l'altro alla sua descrizione <DescrGrandeGruppo>.

Ogni tipologica, rappresentata da una tabella di questo tipo, è stata modellata come una classe, il cui nome è composto da: due lettere che corrispondono all'acronimo (come in tabella 4), il nome della tipologica, la parola finale *Tipo* per sottolineare che si tratta della corrispondenza con una tipologica.

La tipologica relativa al Grande gruppo tariffario è stata perciò modellata come classe *GG_GrandeGruppoTariffario_Tipo*.

Naturalmente questa è una scelta arbitraria delle autrici, ma è stata introdotta per rendere più chiara e immediata la lettura dell'ontologia.

Sono state definite due *DatatypeProperty*, una corrispondente al codice e l'altra alla descrizione, entrambi i dati sono di tipo *string*. In una versione preliminare erano state definite due generiche proprietà, *haCodice* e *haDescrizione*, che potevano essere utilizzate da tutte le classi corrispondenti a tipologiche di questo tipo. Dal punto di vista della definizione di un'ontologia, questa poteva essere considerata come soluzione più adeguata, ma si è preferito evidenziare il collegamento che queste relazioni hanno con i metadati degli OD, perciò sono state definite relazioni specifiche per ciascuna tipologica.

Riprendendo l'esempio del Grande gruppo tariffario, per la classe *GG_GrandeGruppoTariffario_Tipo* sono definite due *DatatypeProperty* con la restrizione di tipo esistenziale.

I nomi scelti per le proprietà, in assenza di ambiguità, coincidono con i nomi dei metadati delle tipologiche, <GrandeGruppoTariffario> e <DescrGrandeGruppoTariffario>.

Nel formalismo matematico le restrizioni si esprimono con:

$$\begin{aligned} &\exists \text{ GrandeGruppoTariffario string} \\ &\exists \text{ DescrGrandeGruppoTariffario string} \end{aligned}$$

La figura 9 mette a confronto, a sinistra il formato tabellare della tipologica, così come si recupera dal portale Inail, a destra il grafo che rappresenta la classe (rettangolo piccolo verde) e le sue istanze (rettangoli viola).

Ogni *istanza* corrisponde ad un record della tabella della tipologica, ed è perciò composta di due *DatatypeProperty*: il valore del parametro e la sua descrizione.

Nell'area su fondo giallo in basso a destra si possono leggere le caratteristiche dell'istanza *GG_6*, i valori assegnati alle sue proprietà, cioè il codice "6" e la descrizione "Metalli e Macchinari".

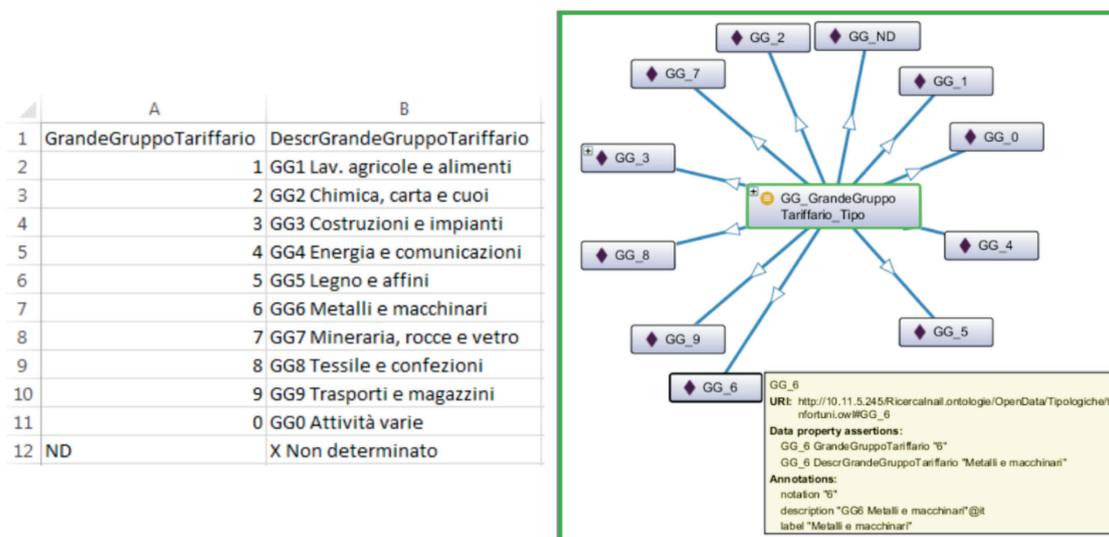


Figura 9 - Comparazione tra la tipologica e la sua rappresentazione a grafo (classe e istanze)

Infine, la classe *GA_GrandeGruppoTariffario_Tipo* è stata definita come *equivalente* alla lista specifica delle sue istanze (*enumerated class*).

La metodologia illustrata è stata applicata alle tipologiche in tabella 5 che hanno caratteristiche simili a quella descritta in dettaglio.

Tabella 5 - Tipologiche semplici e nome della classe corrispondente

Tipologica	Classe
Definizione Amministrativa	DA_DefinizioneAmministrativa_Tipo
Decisione Istruttoria Esito Mortale	DI_DecisioneIstruttoriaEsitoMortale_Tipo
Definizione Amministrativa Esito Mortale	EM_DefinizioneAmministrativaEsitoMortale_Tipo
Gestione Aggregata	GA_GestioneAggregata_Tipo
Grande Gruppo Tariffario	GG_GrandeGruppoTariffario_Tipo
Genere	GN_Genere_Tipo
Settore Tariffario	GT_SettoreTariffario_Tipo
Tipologica Indennizzo Evento	TI_TipologicaIndennizzoEvento_Tipo
Con Senza Mezzo trasporto	TR_ConSenzaMezzoTrasporto_Tipo
Modalità di Accadimento	MA_ModalitaAccadimento_Tipo

3.3.2 Modellazione di tipologiche complesse

Per comprendere cosa si intende per tipologiche complesse, è utile l'esempio della tipologica delle *Province* italiane. Ogni provincia è rappresentata da un identificativo numerico univoco e da una lista di attributi che, oltre a indicare il nome e la sigla della provincia stessa, definiscono la regione e la macro regione di appartenenza.

Una possibilità di modellazione è quella di considerare una sola classe e rappresentare ciascun campo della tabella (o metadato) come un attributo (*DatatypeProperty*) di tipo *integer* o *string* in base al tipo di valore che può assumere. Le istanze di questa classe diventano dunque i singoli

record contenuti nella tipologica. La figura 10 illustra la descrizione della provincia *Aosta* secondo questa modalità di costruzione.

Data property assertions +		
DescrRegione	"Valle D'Aosta"^^string	? @ X O
CodRegione	2	? @ X O
CodMacroRegione	1	? @ X O
CodSiglaProvincia	"AO"^^string	? @ X O
Provincia	"007"^^string	? @ X O
DescrMacroRegione	"Nord - Ovest"^^string	? @ X O
DescrProvincia	"Aosta - Aoste"^^string	? @ X O

Figura 10 - Esempio di descrizione della Provincia *Aosta*

Tale procedimento di modellazione riflette esattamente la rappresentazione tabellare della tipologica, tuttavia presenta degli inconvenienti nella sua gestione, andando anche contro alcuni principi di base dell'ontologia, ad esempio quello di univocità degli elementi. Infatti, in questo caso le informazioni relative a regioni e macro regioni sono ripetute per tutte le province di appartenenza, con possibilità di commettere errori, se si popola l'ontologia manualmente, e di rendere più complesse e articolate le interrogazioni.

Si è deciso di modellare il concetto di *provincia* esplicitando e modellando anche i concetti di *regione* e *macroregione* e definendo le relazioni tra questi elementi. Perciò ogni *provincia* ha i suoi attributi (*DatatypeProperty*) di tipo *integer* o *string* per definire il codice, la sigla e la denominazione, e una relazione (*ObjectProperty*) che la collega con l'istanza della regione di appartenenza, in questo modo la relazione tra provincia e regione è modellata tra istanze e non tra codici. La stessa strategia si applica nella relazione tra le regioni e le macroregioni.

La classe *Class_Provincia* (si è aggiunto il prefisso *Class_* per maggiore chiarezza dal momento che il termine *Provincia* è usato come nome della relazione) è definita con le seguenti restrizioni applicate alle proprietà *DatatypeProperty*:

$$\text{Cardinality}(\text{CodSiglaProvincia}) = 1$$

$$\text{Cardinality}(\text{DescrProvincia}) = 1$$

$$\text{Cardinality}(\text{Provincia}) = 1$$

Dove:

- *CodSiglaProvincia*, *DescrProvincia*, *Provincia* sono *DatatypeProperty* e indicano rispettivamente la sigla, la descrizione e il codice della provincia;
- *Cardinality* è la cardinalità della relazione, nel nostro caso uguale a 1
- *String* indica che il dato equivale ad una sequenza di caratteri.

È stata inoltre definita la relazione *inRegione* di tipo *ObjectProperty*, che lega istanze di *Class_Provincia* con individui della classe *Regione*:

$$\text{Cardinality}(\text{inRegione}) = 1$$

$$\text{Domain}(\text{inRegione}) = \text{Class_Provincia}$$

$$\text{Range}(\text{inRegione}) = \text{Regione}$$

$$\text{inRegione Inverse}(\text{haProvince})$$

La relazione *inRegione* è stata definita esplicitando il *dominio* e il *range*, la restrizione di tipo cardinale applicata alla classe delle province sta ad indicare che una provincia può appartenere ad una sola regione, si è poi definita come inversa della relazione *inRegione*, la relazione *haProvince*.

La classe *Regione* è stata modellata definendo le seguenti relazioni con restrizioni:

- \exists *haProvince* *Class_Provincia*
- \forall *haProvince* *Class_Provincia*
- Cardinality(DescrRegione) = 1*
- Cardinality(inMacroRegione) = 1*
- inMacroRegione Inverse(haRegioni)*

Ci soffermiamo solo sulle prime due relazioni, dal momento che le altre hanno spiegazione analoga al caso precedente. Con la prima relazione (esistenziale) si intende che ciascuna regione ha almeno una provincia, (cioè esiste almeno un individuo per la relazione *haProvince*); la seconda relazione (universale) indica invece che ogni volta che la relazione *haProvince* ha un valore per un individuo della classe *Regione*, l'istanza di riferimento appartiene alla classe *Class_Provincia*.

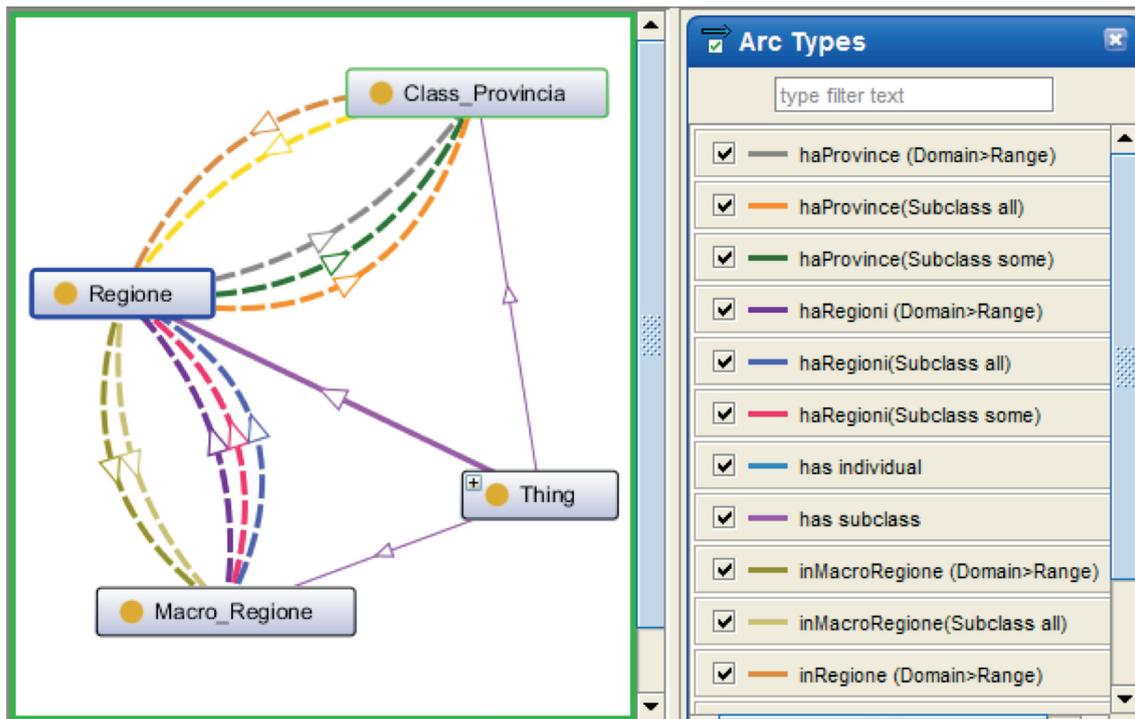


Figura 11 - Rappresentazione del grafo: regione, macro regione, provincia

La classe *Macro_Regione* è modellata in modo del tutto analogo al caso precedente. La figura 11 illustra il grafo risultante, la legenda di destra spiega come individuare, per colore, le relazioni e le restrizioni.

Il vantaggio di questo approccio è che i concetti sono modellati in modo distinto e completo e le istanze sono messe tra loro in relazione; inoltre utilizzando le proprietà relazionali (nell'esempio quella di *inversione*) si possono sfruttare le capacità di deduzione ed inferenza tipiche dei *reasoner* ontologici.

In figura 12 si illustra un esempio del risultato dell'azione del *reasoner*, per l'individuo *Abruzzo*. Nel riquadro *Annotations*, in alto nella figura 12, si può leggere oltre al nome della regione, anche il riferimento ad una pagina web, infatti si è ritenuto utile predisporre il *link* con il sito di SPCData relativo all'Italia e metterlo in corrispondenza con le Regioni e le Province, secondo il criterio dei *linked open data*.

Nella parte in basso, si trovano le proprietà inferite, che sono quelle racchiuse nei rettangolini tratteggiati con fondo giallo, ovvero il *reasoner* deduce che l'istanza *Abruzzo* appartiene alla classe *Regione*, come illustrato nel riquadro a sinistra, sebbene la proprietà non sia stata definita in fase di creazione dell'istanza, e che ha 4 province rappresentate dagli individui *PR_066*, *PR_067*, *PR_068*, *PR_069* elencate nel riquadro a destra.

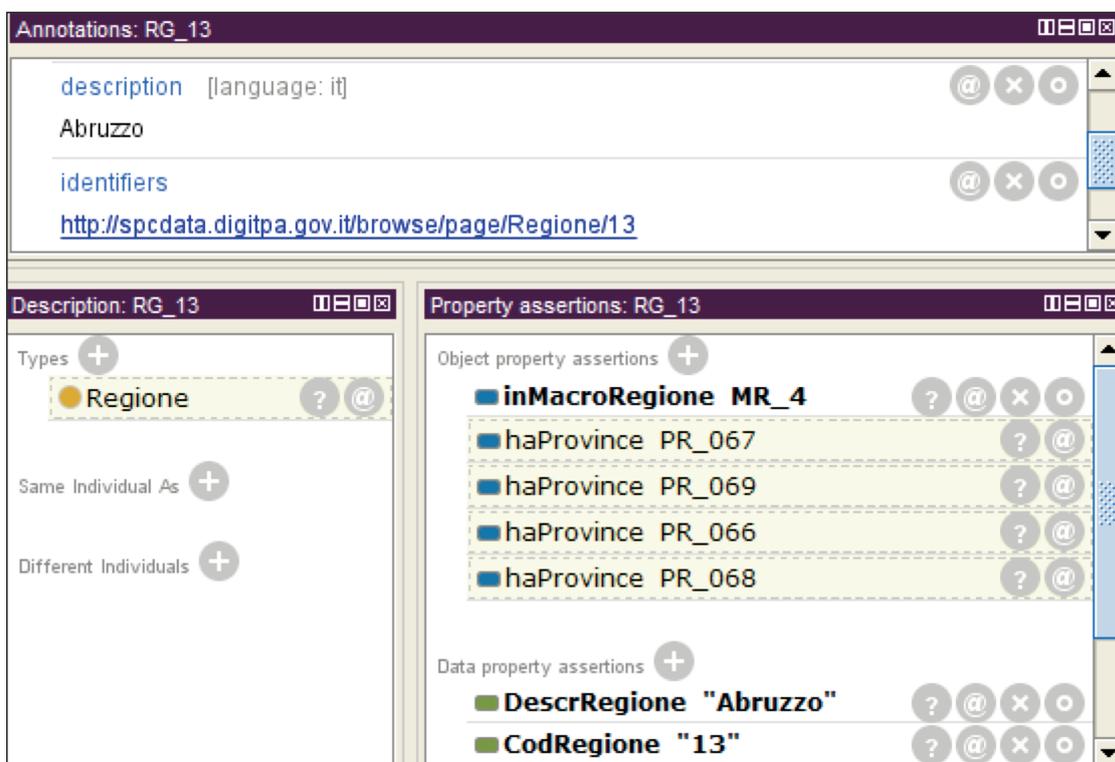


Figura 12 - Risultato di inferenza rispetto all'istanza *RG_13* (*Abruzzo*)

Protégé dispone di funzionalità che mostrano quali siano le regole adottate che portano alla deduzione; esse si visualizzano selezionando il "?" associato all'inferenza, nella figura 13 illustriamo solo alcune di esse.

Figura 13 - Regole delle inferenze mostrate in precedenza

Tra le tipologiche complesse ci sono quelle che contengono informazioni generali, prodotte e gestite da altre PA, esse fanno riferimento a:

- *luogo di nascita* dell'infortunato corrispondente al codice della Nazione di nascita;
- *luogo di accadimento* dell'infortunio che indica la Provincia di accadimento;
- *settore di attività economica* che fa riferimento al codice Ateco dell'azienda dove è avvenuto l'infortunio.

La tipologica Luogo di nascita è una tabella complessa composta di 7 campi, quelli che abbiamo preso in considerazione sono: il codice della nazione <LuogoNascita>, la sua denominazione <DescrNazioneNascita>, l'appartenenza oppure no all'Unione Europea <DescrGruppoPaese> e <FlagAppartenenzaUE>.

La soluzione adottata è quella di definire due classi corrispondenti: una per definire il luogo di nascita (*LN_LuogoDiNascita*), l'altra relativa all'appartenenza o meno all'UE (*LN_GruppoPaese*).

Quest'ultima è una classe con due *DatatypeProperty*:

- la denominazione del gruppo (*DescrGruppoPaese*)
- la sua appartenenza all'UE (valore booleano *FlagAppartenenzaUE*).

La classe *LN_LuogoDiNascita* ha due *DatatypeProperty*:

- il codice, corrispondente al metadato *LuogoNascita*;
- la denominazione della nazione, cioè il metadato *DescrNazioneNascita*;

e un *ObjectProperty* (*haAppartenenzaGruppo*) che mette in corrispondenza un'istanza di questa classe con un individuo della classe *LN_GruppoPaese*.

I seguenti esempi (figure 14 e 15) illustrano i due casi: nazione europea ed extraeuropea.

Figura 14 - Esempio di istanza *Belgio*Figura 15 - Esempio di istanza *Marocco*

Un discorso molto simile si applica alla tipologica relativa al settore di attività economica definita attraverso il codice Ateco.

La tipologica relativa al settore economico è anch'essa rappresentata come tabella organizzata in quattro campi: i primi due relativi al codice e alla descrizione del codice Ateco di secondo livello, i restanti analoghi per quello di primo livello. Il modello che abbiamo definito si compone di due classi corrispondenti al primo livello (*AtecoLiv1*) e al secondo livello (*SettoreAttivitaEconomica*) del codice Ateco, ciascuna classe ha due *DatatypeProperty*. Per la classe *AtecoLiv1* sono:

- il codice settore di attività economica di livello 1 (*CodAtecoLiv1*)
- la descrizione corrispondente (*DescrAtecoLiv1*).

Per la classe *SettoreAttivitaEconomica* abbiamo i seguenti *DatatypeProperty*:

- il codice settore di attività economica di livello 2 (*SettoreAttivitaEconomica*)
- la descrizione corrispondente (*DescrAteco*)

Inoltre sono state introdotte due *ObjectProperty* che mettono in corrispondenza gli elementi delle due classi:

- *haCodiceLiv1*, che mette in corrispondenza un'istanza della classe *AtecoLiv1* con la classe *SettoreAttivitaEconomica*;
- *haSettoriEconomici*, che mette in corrispondenza un'istanza della classe *SettoreAttivitaEconomica* con la classe *AtecoLiv1*.

Le due relazioni sono inoltre definite l'una inversa dell'altra, con interessanti conseguenze.

In figura 16 si illustra l'esempio di rappresentazione dell'istanza *AT_C25*.



Figura 16 - Le proprietà dell'istanza AT_C25

Nella figura 17 si evidenzia, invece, l'inferenza generata automaticamente dal *reasoner*, che deduce l'appartenenza dei settori economici (di secondo livello) al codice Ateco di primo livello "C".



Figura 17 - L'inferenza attribuisce i codici di livello 2 AT_C25, AT_C26 al codice di livello 1 AT_C.

Questa rappresentazione è, a nostro parere, quella più vicina dal punto di vista concettuale alla corrispondente tipologica messa a disposizione da Inail.

Tuttavia, nell'ottica di un'effettiva interoperabilità con gli OD delle altre Pubbliche Amministrazioni, sarebbe utile prendere in considerazione la struttura disponibile sul sito ISTAT "Classificazione delle attività economiche" [ATECO-007].

3.3.3 Modellazione di tipologiche parametriche

Dall'analisi è emerso che alcuni concetti non sono direttamente rappresentati né negli OD né nelle tipologiche; è il caso, ad esempio, delle classi di menomazione. Nella descrizione di questa tipologica non c'è la diretta corrispondenza tra il metadato contenuto negli OD (*<GradoMenomazione>*) e un valore in tabella, ma l'appartenenza di un individuo ad una classe di menomazione è determinata dalla verifica che il parametro appartenga ad un dato intervallo di valori.

Nel vocabolario, la *classe di menomazione* è "un elemento della partizione dell'intervallo di variabilità del grado di menomazione *"p"*" inoltre "sono state definite convenzionalmente sei classi"; dall'esito del confronto tra il valore del parametro *"p"* e l'intervallo di valori entro cui ricade si determina la classe di menomazione a cui appartiene l'infortunato.

Negli OD viene registrato solo il valore *"p"* del *grado di menomazione*, quindi chi volesse operare sugli *open data* per conoscere la classe di menomazione a cui gli infortunati appartengono, deve prevedere un piccolo controllo aggiuntivo sulla base della legenda contenuta nel vocabolario.

L'ontologia, in particolare attraverso l'ultima versione OWL [OWL-012], offre un modo elegante per modellare questi concetti utilizzando le proprietà di equivalenza e le restrizioni applicate alle *DatatypeProperty* per limitare i valori di appartenenza di un attributo ad uno specifico intervallo. La soluzione adottata è stata quella di definire due classi: una relativa ai codici, *CM_CodiceClasse-*

DiMenomazione, l'altra, *CM_ClasseDiMenomazione*, come classe radice della tassonomia relativa alle classi di menomazione.

La classe *CM_CodiceClasseDiMenomazione* è modellata seguendo le indicazioni per le tipologie semplici illustrate nel paragrafo 3.3.1, ovvero con due *DatatypeProperty* corrispondenti rispettivamente al valore (*haCodice*) e alla sua descrizione (*haDescrizione*). In figura 18 è rappresentato il grafo della classe con i suoi individui.

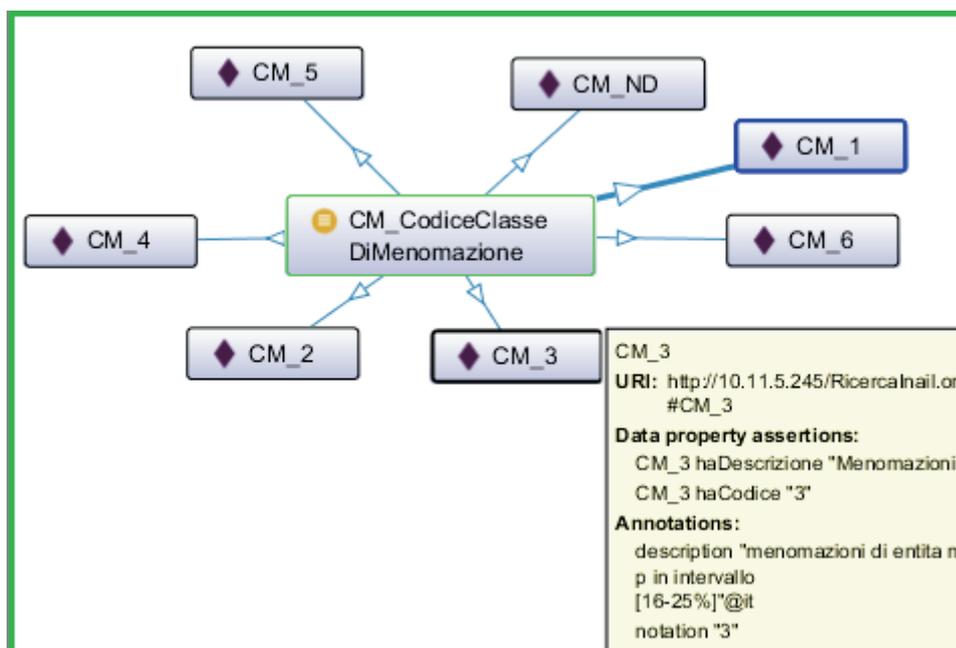


Figura 18 - Rappresentazione della classe relativa ai codici di menomazione

La classe *CM_ClasseDiMenomazione* è rappresentata da una struttura gerarchica di sette sottoclassi (figura 19) corrispondenti alle sei classi di menomazione descritte nel vocabolario più quella che raggruppa infortunati ai quali non è stato riconosciuto, o ancora definito, alcun grado di menomazione.

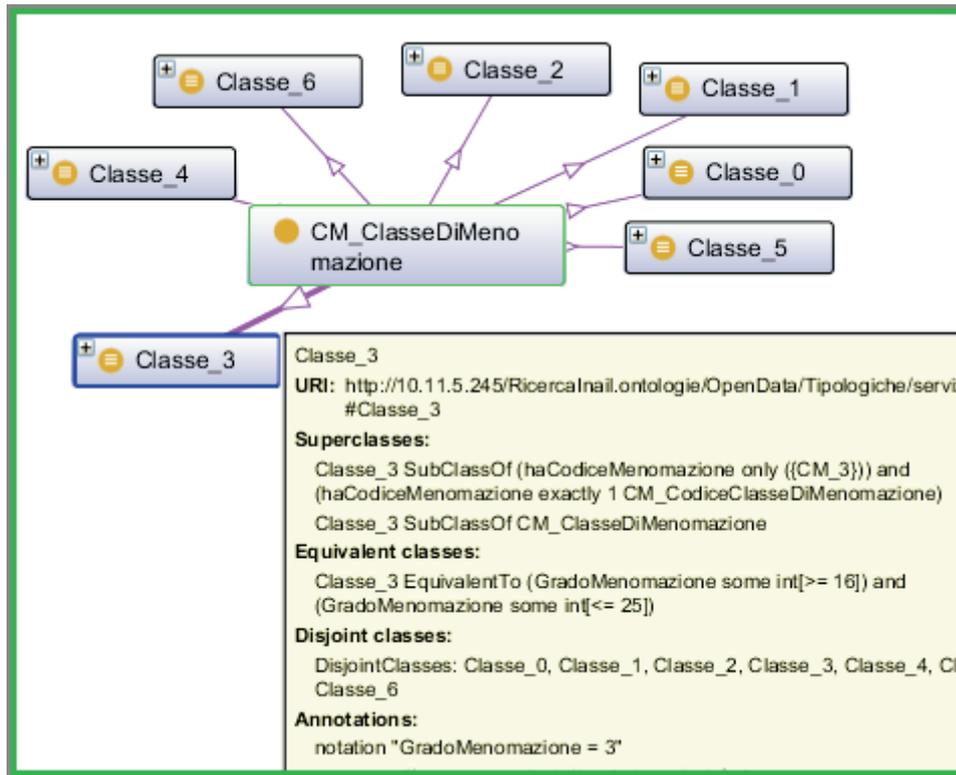


Figura 19 - Classe di Menomazione e sue sottoclassi

Ciascuna delle classi, per come sono state definite nel *Quaderno 1*, è una partizione dell'intervallo di numeri interi [0-100], che rappresenta il dominio in cui il parametro p (grado di menomazione) può assumere valore. Infatti:

$$\begin{aligned}
 & \text{sia } I = \{0, \dots, 6\}, P_i = [n_i, m_i], \forall i \in I, n_i < m_i \\
 & P_i \text{ partizione dell'intervallo di numeri interi } [0, 100] \\
 & P_i \cap P_j = \emptyset \text{ e } \cup_i P_i = [0, 100]
 \end{aligned}$$

Ciascuna sottoclasse è stata perciò modellata come equivalente alla classe a cui vengono applicate la proprietà e la restrizione seguenti:

$$(\exists \text{ GradoMenomazione integer } [\geq n_i]) \text{ and } (\exists \text{ GradoMenomazione integer } [\leq m_i])$$

Dove con i termini n_i e m_i si intendono gli estremi dell'intervallo i -esimo come definiti nel vocabolario, entro cui varia il parametro p , cioè il grado di menomazione assegnato da Inail al caso di infortunio.

La definizione di equivalenza (condizione necessaria e sufficiente) vuol dire che gli individui di ciascuna sottoclasse hanno il grado di menomazione p che ricade nell'intervallo di definizione, e viceversa se un'istanza ha il grado di menomazione p che cade nell'intervallo allora è anche istanza di quella sottoclasse.

$$\text{Individual} \in \text{Classe}_i \Leftrightarrow p \in [n_i, m_i]$$

Dove:

$p = \text{GradoDiMenomazione}(\text{Individual})$

\in indica appartenenza di un elemento ad una classe.

Inoltre, ciascuna sottoclasse viene messa in corrispondenza con il proprio codice di menomazione mediante la proprietà *haCodiceMenomazione* e le seguenti restrizioni:

$$(\forall \text{haCodiceMenomazione} (\{CM_n\})) \text{ and} \\ (\text{Cardinality}(\text{haCodiceMenomazione}) = 1)$$

Le sottoclassi sono, infine, state dichiarate disgiunte fra loro, in questo modo un individuo può appartenere ad una sola di esse.

Per completezza dell'ontologia, come detto, è stata definita una classe aggiuntiva rispetto alle sei definite nel vocabolario, la *Classe_0*, che indica che l'individuo non appartiene ad alcuna classe di menomazione, ovvero verifica la condizione *GradoMenomazione* ≤ 0 .



Figura 20 - Inferenza del caso di infortunio con *GradoMenomazione* = 2

Questa modellazione offre il vantaggio che gli infortuni possano essere automaticamente classificati dal *reasoner* come appartenenti ad una certa classe di menomazione, nell'esempio in figura 20 si vede che l'infortunio con grado di menomazione $p = 2$ viene assegnato dal *reasoner* come individuo (*inferito*) della Classe 1 di Menomazione.

4 Esempi di navigazione, interrogazione, visualizzazione dell'ontologia

In questo capitolo vengono presentate alcune modalità di navigazione, interrogazione e visualizzazione di *ontoInfortuni* per evidenziare la flessibilità del modello, per verificare la completezza e la coerenza con gli OD e visualizzare il modello da un punto di vista differente rispetto a quello di sviluppo.

Occorre ribadire che, poiché il modello è costruito sugli OD, ritroviamo esatta corrispondenza tra i campi che compongono un record e gli elementi dell'ontologia. D'altra parte però, *ontoInfortuni* non ha un link diretto con i *dataset* regionali dell'OD ma solo con la loro struttura logica attraverso i metadati, pertanto per gli esempi in questo capitolo, sono stati caricati manualmente dei record nell'ontologia.

L'inserimento di una denuncia di infortunio nell'ontologia viene fatto a partire da un record degli OD, creando le tre istanze corrispondenti al caso di infortunio, al datore di lavoro e all'infortunato e aggiungendo a ciascuna le corrispondenti proprietà.

In tabella 6 è illustrato un esempio di un record estratto dall'OD della Valle d'Aosta in formato RDF, in blu sono stati evidenziati i metadati relativi al caso di infortunio, in rosso quelli relativi al datore di lavoro ed in viola quelli dell'infortunato.

Tabella 6 - Record in formato RDF estratto dall'OD della Valle d'Aosta

```

<DataRilevazione rdf:datatype='xs:date'>30/04/2014</DataRilevazione>
<DataProtocollo rdf:datatype='xs:date'>29/07/2011</DataProtocollo>
<DataAccadimento rdf:datatype='xs:date'>28/07/2011</DataAccadimento>
<DataDefinizione rdf:datatype='xs:date'>12/10/2011</DataDefinizione>
<DataMorte rdf:datatype='xs:date'>28/07/2011</DataMorte>
<LuogoAccadimento rdf:resource='Provincia.rdf/007'/>
<IdentificativoInfortunato rdf:datatype='xs:string'>5224280</IdentificativoInfortunato>
<Genere rdf:datatype='xs:string'>M</Genere>
<Eta rdf:datatype='xs:integer'>28</Eta>
<LuogoNascita rdf:resource='LuogoNascita.rdf/ITAL'/>
<ModalitaAccadimento rdf:datatype='xs:string'>N</ModalitaAccadimento>
<ConSenzaMezzoTrasporto rdf:datatype='xs:string'>N</ConSenzaMezzoTrasporto>
<IdentificativoCaso rdf:datatype='xs:string'>14071252</IdentificativoCaso>
<DefinizioneAmministrativa rdf:resource='DefinizioneAmministrativa.rdf/P'/>
<DefinizioneAmministrativaEsitoMortale
rdf:datatype='xs:string'>P</DefinizioneAmministrativaEsitoMortale>
<Indennizzo rdf:resource='TipologiaIndennizzo.rdf/RS'/>
<DecisioneIstruttoriaEsitoMortale rdf:resource='DecisioneIstruttoriaEsitoMortale.rdf/015'/>
<GradoMenomazione rdf:datatype='xs:integer'>-1</GradoMenomazione>
<GiorniIndennizzati rdf:datatype='xs:integer'>0</GiorniIndennizzati>
<IdentificativoDatoreLavoro rdf:datatype='xs:string'>3673905</IdentificativoDatoreLavoro>
<PosizioneAssicurativaTerritoriale
rdf:datatype='xs:string'>3713534</PosizioneAssicurativaTerritoriale>
<SettoreAttivitaEconomica rdf:resource='SettoreAttivitaEconomica.rdf/F 41'/>
<Gestione rdf:resource='Gestione.rdf/I'/>
<GestioneTariffaria rdf:resource='GestioneTariffaria.rdf/1'/>
<GrandeGruppoTariffario rdf:resource='GrandeGruppoTariffario.rdf/3'/>

```

Nella figura 21 (A, B, C) sono illustrate le istanze inserite nell'ontologia, dove si nota che gli attributi di tipo *DatatypeProperty* (ad esempio le date, gli identificativi, l'età, etc.) sono gli stessi presenti nel record della tabella 6. Riguardo i codici (ad esempio "ITAL" o "007" per la provincia), nel record dell'*open data* sono valori che riferiscono ad elementi delle tipologiche (luogo nascita, luogo accadimento); in *ontoInfortuni*, invece, sono gli individui che appartengono alla corrispondente classe tipologica, secondo la sintassi descritta nel paragrafo 3.3.

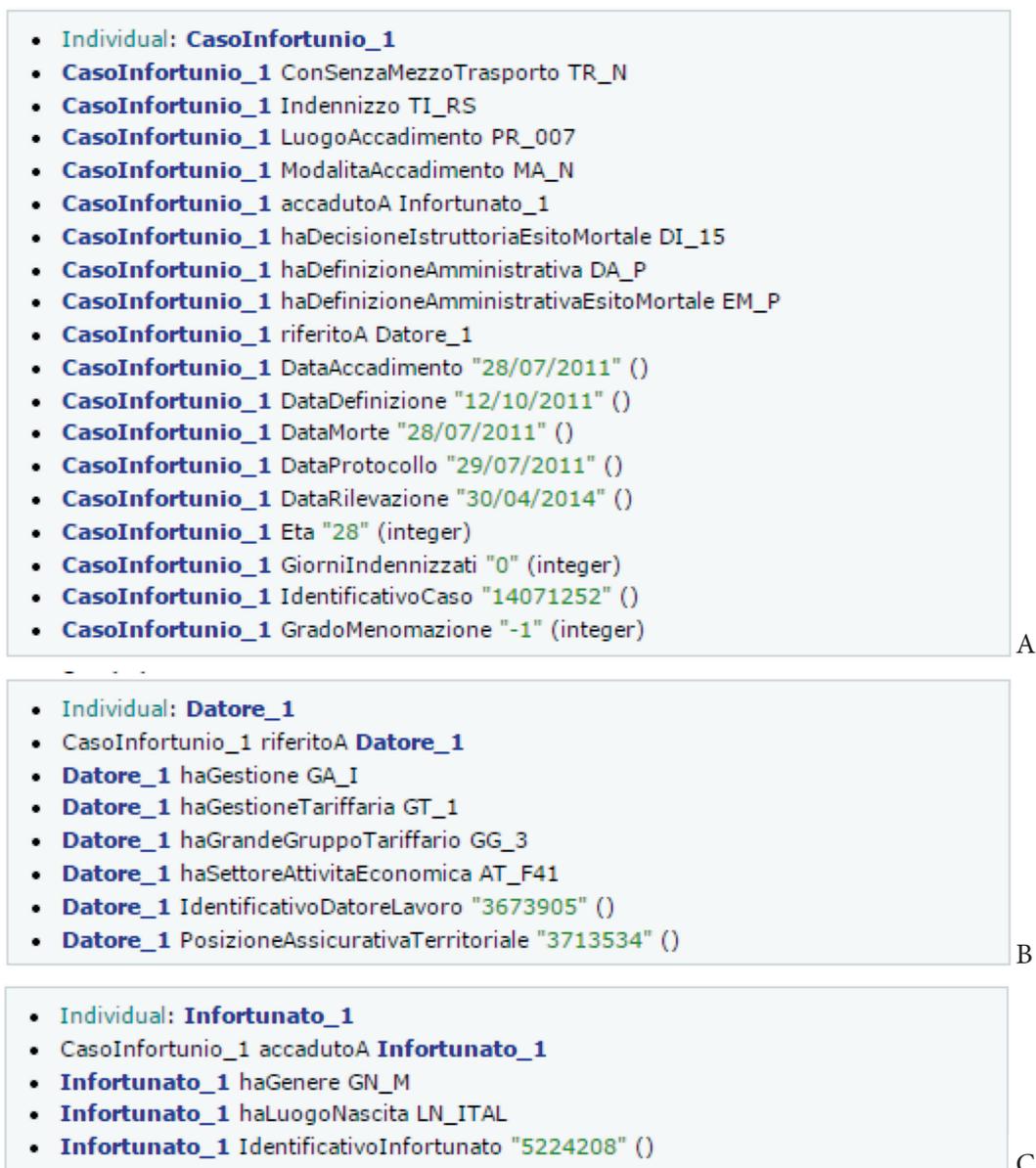


Figura 21 (A, B, C) - Record della Valle d'Aosta modellato attraverso gli individui di *ontoInfortuni*.

4.1 Modalità di navigazione ipertestuale: OWLDoc

Una modalità di visualizzazione fornita da Protégé è *OWLDoc*, una funzionalità che crea automaticamente una descrizione statica dell'ontologia in pagine HTML. La pagina è suddivisa in tre riquadri: quello in alto a sinistra contiene l'elenco completo delle ontologie utilizzate, in basso a sinistra le caratteristiche dell'ontologia selezionata nel riquadro precedente, a destra la pagina che descrive l'elemento selezionato nel secondo riquadro.

Questo tipo di visualizzazione permette di navigare nell'ontologia accedendo a tutte le sue caratteristiche mediante le modalità ipertestuali e può essere utile per comprendere il modello realizzato.

La figura 22 illustra la pagina iniziale, nel riquadro in alto a sinistra è stata selezionata la voce "tutte

le ontologie” (“*All Ontologies*”), l’elenco è riportato nel riquadro in basso a sinistra, ed in questo si è scelto *ontoInfortuni*, le cui caratteristiche generali sono visualizzate nella parte destra della figura, quali elenco e numerosità degli elementi che la compongono e le ontologie relative alle tipologiche che vengono importate per creare *ontoInfortuni*.

The screenshot displays the OWLDoc interface for the **ontoInfortuni** ontology. It features a sidebar on the left with two main sections: **Contents** and **All Ontologies (7)**. The **Contents** section lists: All Ontologies, Classes (34), Object Properties (32), Data Properties (58), Annotation Properties (9), Individuals (93), and Datatypes (5). The **All Ontologies (7)** section lists: *ateco* (c:2, op:2, dp:4, i:8), *common* (c:2, op:0, dp:6, i:8), *nazioni* (c:2, op:2, dp:4, i:9), *provincia* (c:4, op:7, dp:9, i:15), *servizio* (c:13, op:1, dp:7, i:17), *tipologicheinfortuni* (c:4, op:0, dp:12, i:33), and *ontoInfortuni* (c:20, op:20, dp:17, i:27). The main area on the right is titled **ontoInfortuni** and contains four sections: **Annotations (2)** (description "Ontologia degli infortuni dagli Open Data Inail" and label "Ontologia degli infortuni" (it)), **References** (Classes (20), Object Properties (20), Data Properties (17), Annotation Properties (5), Individuals (27), Datatypes (3)), and **Imports (6)** (listing URLs for *ateco.owl*, *common.owl*, *nazioni.owl*, *provincia.owl*, *servizio.owl*, and *tipologicheinfortuni.owl*).

Figura 22 - Pagina principale in modalità OWLDoc

In figura 23, si illustra un esempio in cui si è scelto di visualizzare le caratteristiche della classe *DI_DecisioneIstruttoriaEsitoMortale_Tipo*, che appartiene all’ontologia *tipologicheinfortuni*.

In particolare si può notare che alla voce *Equivalentents* c’è l’insieme di tutti i suoi individui, cioè la classe può contenere solo quegli elementi.

Class: DI_DecisioneIstruttoriaEsitoMortale_Tipo

<http://10.11.5.245/RicercaInail.ontologie/OpenData/Tipologiche/tipologicheinfo>
Annotations (2)

- comment "Decodifica della decisione istruttoria relativa alla trattazione esiti mortali" (it)
- label "Decisione Istruttoria Esito Mortale" ()

Equivalentents (1)

- {DI_ND, DI_244, DI_22, DI_21, DI_300, DI_15, DI_28, DI_27, DI_17, DI_16, DI_24, DI_18}

Superclasses (2)

- DecisioneIstruttoriaEsitoMortale some string
- DescrDecisioneIstruttoriaEsitoMortale some string

Disjoints (12)

DI_15, DI_16, DI_17, DI_18, DI_21, DI_22, DI_24, DI_244, DI_27, DI_28, DI_300, DI_ND

Usage (14)

- Class: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_15: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_16: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_17: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_18: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_21: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_22: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_24: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_244: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_27: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_28: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_300: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- DI_ND: **DI_DecisioneIstruttoriaEsitoMortale_Tipo**
- haDecisioneIstruttoriaEsitoMortale Range **DI_DecisioneIstruttoriaEsitoMortale_Tipo**

Figura 23 - La classe Decisione Istruttoria Esito Mortale nell'ontologia *tipologicheinfortuni*.

Selezionando DI_15 dall'elenco delle istanze (*Usage*) si può accedere alla descrizione, illustrata nella figura 24, dell'elemento che nella tipologica relativa alla decisione istruttoria esito mortale ha codice 15, con la descrizione "Esiti mortali nei termini con superstiti", visibili alla voce *Annotations*. Si noti, tra le restrizioni di proprietà in cui tale individuo è utilizzato nel riquadro *Usage*, quella che determina l'equivalenza con la classe *InfortunioEsitoMortalePositivo* (la terza dall'alto); ciò sta ad indicare che se un individuo ha la decisione istruttoria esito mortale di tipo 15, allora rientra nella classe di infortunio mortale con esito positivo. Questa relazione è utilizzata dal *reasoner* per classificare automaticamente tutti gli individui che hanno questa proprietà.

Individual: DI_15

<http://10.11.5.245/RicercaInail.ontologie/OpenData/Tipologiche/tipologicheinfortuni>
Annotations (3)

- description "Esiti Mortali Nei Termini Con Superstiti" (it)
- label "Con Superstiti" ()
- notation "15" ()

Types (1)

- DI_DecisioneIstruttoriaEsitoMortale_Tipo

Usage (6)

- Individual: **DI_15**
- DI_DecisioneIstruttoriaEsitoMortale_Tipo \equiv {DI_ND, DI_244, DI_22, DI_21, DI_300, **DI_15**, DI_28, DI_27, DI_17, DI_16, DI_24, DI_18}
- InfortunioEsitoMortalePositivo \equiv haDecisioneIstruttoriaEsitoMortale some {**DI_15**, DI_17, DI_16, DI_18}
- CasoInfortunio_1 haDecisioneIstruttoriaEsitoMortale **DI_15**
- **DI_15** DecisioneIstruttoriaEsitoMortale "15" ()
- **DI_15** DescrDecisioneIstruttoriaEsitoMortale "Esiti Mortali Nei Termini Con Superstiti" ()

Figura 24 - Selezione di un individuo della classe decisione istruttoria esito mortale.

Si sono inoltre riportate tutte le definizioni contenute nel vocabolario, inserendole come descrizione degli elementi dell'ontologia.

Object Property: haGestione

<http://10.11.5.245/RicercaInail.ontologie/OpenData/ontoInfortuni#haGestione>
Annotations (3)

- description "Caratterizza le modalita' di esercizio dell'assicuraz...ione contro gli infortuni sul lavoro. Si hanno 5 gestioni: "industria e servizi", "agricoltura", "per conto dello Stato", "medici radiologi", "infortuni in ambito domestico". Il flusso prende in considerazione solo le gestioni "industria e servizi", "agricoltura", "per conto dello Stato". " (it)
- label "haGestione" ()
- seeAlso "DPR 1124/1965, titolo I, titolo II, art. 127, L 93/1958, DPR 1055/1960, L 493/1999, DM 15.09.2000, DM 31.01.2006, DLgs 38/2000, art. 1. " (it)

Domains (1)

- DatoreLavoro

Ranges (1)

- GA_GestioneAggregata_Tipo

Figura 25 - Visualizzazione della relazione *haGestione*.

In figura 25 è rappresentata, ad esempio, la relazione di tipo *ObjectProperty*, *haGestione*, la cui descrizione è contenuta alla voce *gestione* del vocabolario [pagina 18]. L'annotazione *seeAlso*, contenuta in RDFS, è stata utilizzata per riportare i riferimenti indicati nel vocabolario. Al momento sono semplici testi, ma potrebbero diventare *link* per accedere alle pagine riferite.

Questo procedimento, oltre alla navigazione nell'ontologia, può rappresentare un'altra modalità di "lettura" e consultazione del vocabolario.

4.2 Modalità di interrogazione: SPARQLinfortuni

SPARQL è il linguaggio di interrogazione e di gestione dei dati in formato RDF ed è standardizzato da W3C. I servizi *web* per l'interrogazione dei dati secondo questo protocollo sono detti SPARQL Endpoint.

Le Pubbliche Amministrazioni, oltre a rendere disponibili i dati nei diversi formati standard, spesso forniscono l'accesso libero a specifici SPARQL Endpoint. L'utente può perciò interrogare i *dataset* disponibili formulando e componendo direttamente le *query*. Se da una parte questa modalità garantisce completa libertà di interrogazione dei dati, dall'altra richiede una conoscenza piuttosto approfondita sia del modello in cui i dati sono organizzati sia della sintassi del linguaggio SPARQL. Non è obiettivo di questo *Quaderno*, e neppure competenza delle autrici, sviluppare uno SPARQL Endpoint relativo agli *Open Data* Inail sugli infortuni sul lavoro, tuttavia è stato sviluppato un prototipo applicativo che può interrogare gli OD Inail attraverso l'ontologia *ontoInfortuni*. Questo prototipo ha solo lo scopo di dimostrare la fattibilità di alcune modalità di interrogazione che sono illustrate in seguito. Si è perciò dato più risalto all'interrogazione dei concetti e delle loro relazioni, al fine di verificare che la struttura creata fosse coerente con gli OD, piuttosto che verificare la numerosità dei casi contenuti in essi. Per tale motivo si sono inseriti nel modello ontologico solo alcuni casi (*record*) estratti dagli OD.

Il prototipo, sviluppato in linguaggio Java, utilizza la libreria Jena [Jena] per la creazione e gestione dell'ontologia a partire dalla versione OWL di *ontoInfortuni*.

Un obiettivo è stato quello di poter verificare attraverso lo SPARQL Endpoint la coerenza e la completezza semantica del modello realizzato, rispetto alle tabelle rappresentate nell'Allegato 1 del *Quaderno1*, intendendolo perciò come strumento per verificare che fossero presenti tutte le relazioni, esplicite o implicite, contenute nel vocabolario.

A tale scopo si sono ripercorse le tabelle contenute formulando in linguaggio SPARQL, per ciascuna tabella, la *query* corrispondente per interrogare *ontoInfortuni*. L'insieme di *query* implementate, per l'obiettivo appena descritto, non è esaustivo ma rappresenta un campione significativo.

Tale insieme è stato registrato a priori e richiamato automaticamente dal sistema in base alla scelta effettuata dall'utente. Lo spazio nel secondo riquadro a sinistra, dove compare la *query* in linguaggio SPARQL, è uno spazio editabile dove l'utente, esperto di tale linguaggio, può comporre direttamente la domanda di proprio interesse.

L'esempio di figura 26 illustra il caso in cui è stata selezionata la tabella B1, che corrisponde all'interrogazione "*Denunce d'infortunio per modalità di accadimento e data di accadimento*" (evidenziata in rosso). La *query* in linguaggio SPARQL (secondo riquadro a sinistra) è stata automaticamente inserita dal sistema. Nel riquadro in basso si ottiene il risultato, ovviamente non aggregato, sui pochi dati inseriti manualmente, motivo per cui non si è implementato il contatore.

Si è, invece, scelto di mettere in chiaro tutti i codici presenti nel risultato, attraverso le relazioni con le tipologiche, nell'esempio le voci relative alla modalità di accadimento e del mezzo di trasporto. Questo, anche per indicare una possibile chiave di lettura differente delle tabelle e per dimostrare la predisposizione dell'ontologia a navigare nei *dataset* completi attraverso le relazioni implementate.

Navigare e interrogare ontoInfortuni:
FAQ / Tabelle:

```

PREFIX tiposteco: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipoAteco#>
PREFIX tipoprovincia: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipoProvincia#>
PREFIX tipologiche: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipologiche#>
PREFIX tipologicheservizio: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipologicheServizio#>
PREFIX tiponazionicontinenti: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipoNazioniContinenti#>
PREFIX infortuni: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoInfortuni#>

SELECT DISTINCT ?ModalitaAccadimento ?ConSenzaMezzoTrasporto ?dataAccadimento
WHERE {
  ?caso infortuni:DataAccadimento ?dataAccadimento.
  ?caso infortuni:ModalitaAccadimento ?modaccad.
  ?modaccad tipologiche:servizio:haDescrizione ?ModalitaAccadimento.
  ?caso infortuni:ConSenzaMezzoTrasporto ?consenza.
  ?consenza tipologicheservizio:haDescrizione ?ConSenzaMezzoTrasporto.
}
ORDER BY (?ModalitaAccadimento)

```

Denunce d'infortunio (disponibili) e casi d'infortunio sul lavoro
Denunce d'infortunio per provincia, età, genere e settore economico
Infortuni per datore di lavoro, infortunato, modalità di accadimento, data di accadimento?
Infortuni per datore di lavoro, settore economico, infortunato, modalità di accadimento, data di accadimento?
Grado e classi di menomazione per settore economico e provincia?
Grado e classi di menomazione per settore economico e genere?
Quali decisioni per denunce di istruttoria esito mortale negative?

B1	B2	B3	B4	B5
B1.1	B1.2	B1.3	B1.4	B1.5

Esegui Reset

ModalitaAccadimento	ConSenzaMezzoTrasporto	dataAccadimento
"in itinere"	"con mezzo di trasporto"	"08/05/2009"
"in itinere"	"con mezzo di trasporto"	"25/09/2009"
"in occasione di lavoro"	"senza mezzo di trasporto"	"05/01/2009"
"in occasione di lavoro"	"senza mezzo di trasporto"	"06/06/2011"
"in occasione di lavoro"	"senza mezzo di trasporto"	"07/09/2010"
"in occasione di lavoro"	"senza mezzo di trasporto"	"07/10/2001"
"in occasione di lavoro"	"senza mezzo di trasporto"	"12/12/2010"
"in occasione di lavoro"	"senza mezzo di trasporto"	"17/07/2012"
"in occasione di lavoro"	"senza mezzo di trasporto"	"23/06/2011"
"in occasione di lavoro"	"senza mezzo di trasporto"	"27/08/2009"
"in occasione di lavoro"	"senza mezzo di trasporto"	"29/12/2008"

Figura 26 - Esempio di interrogazione guidata mediante tabelle predefinite

Inoltre, per rendere più semplice l'interrogazione, si sono formulate, in linguaggio naturale (in italiano), alcune domande (*FAQ - Frequently Asked Questions*) che potrebbero essere di interesse per un utente che voglia consultare gli OD senza però essere costretto ad imparare la sintassi SPARQL e conoscere approfonditamente *ontoInfortuni*.

Anche in questo caso, come in precedenza per le tabelle, ciascuna domanda è stata associata alla formulazione in linguaggio SPARQL che viene automaticamente resa disponibile.

La tabella 7 elenca le domande sviluppate e disponibili nel prototipo software.

Tabella 7 - Elenco delle domande *FAQ* disponibili.

Denunce d'infortunio (disponibili) e casi d'infortunio sul lavoro?
Denunce d'infortunio per provincia, età, genere e settore economico?
Infortuni per datore di lavoro, infortunato, modalità di accadimento, data di accadimento?
Infortuni per datore di lavoro, settore economico, infortunato, modalità di accadimento, data di accadimento?
Grado e classi di menomazione per settore economico e provincia?
Grado e classi di menomazione per settore economico e genere?
Quali decisioni per denunce di istruttoria esito mortale negative?

Nella figura 27 è presentato l'esito di una domanda selezionata dall'elenco proposto in tabella 7, nel caso riportato la seconda.

Navigare e interrogare ontoInfortuni:
FAQ / Tabelle:

PREFIX tipoateco: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipoAteco#>
 PREFIX tipoprovincia: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipoProvincia#>
 PREFIX tipologicheservizio: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipologicheServizio#>
 PREFIX tiponazionicontinenti: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoTipoNazioniContinenti#>
 PREFIX infortuni: <http://ontologie.ricerca.inail.it/opensdata/elements/2014/ontoInfortuni#>

```
SELECT DISTINCT ?Provincia ?Età ?Genere ?CodiceAteco ?DescrAteco
WHERE {
  ?Denunce rdf:type infortuni:CasoDiInfortunio.
  ?Denunce infortuni:LuogoAccadimento ?luogo.
  ?luogo tipoprovincia:DescrProvincia ?Provincia.
  ?Denunce infortuni:etaAccadimento ?Età.
  ?Denunce infortuni:accadutoA ?infortunato.
  ?infortunato infortuni:Genere ?genere.
  ?genere tipologicheservizio:haDescrizione ?DescrAteco.
```

Denunce d'infortunio (disponibili) e casi d'infortunio sul lavoro?
Denunce d'infortunio per provincia, età, genere e settore economico?
 Infortuni per datore di lavoro, infortunato, modalità di accadimento, data di ac
 Infortuni per datore di lavoro, settore economico, infortunato, modalità di acc
 Grado e classi di menomazione per settore economico e provincia?
 Grado e classi di menomazione per settore economico e genere?

B1	B2	B3	B4	B5	B6	B7
B1.1	B1.2					
B1.1.1	B1.3					
B1.1.2	B1.4					
B1.1.3	B1.5					

Esegui Reset

Provincia	Età	Genere	CodiceAteco	DescrAteco
"Aosta"	35	"maschio"	"C 25"	"Fabbricazione di prodotti in metallo (esclusi macchinari e attrezzature)"
"Aosta"	49	"maschio"	"C 25"	"Fabbricazione di prodotti in metallo (esclusi macchinari e attrezzature)"
"Aosta"	39	"maschio"	"C 25"	"Fabbricazione di prodotti in metallo (esclusi macchinari e attrezzature)"
"Aosta"	38	"maschio"	"C 25"	"Fabbricazione di prodotti in metallo (esclusi macchinari e attrezzature)"
"Aosta"	50	"maschio"	"F 41"	"Costruzione di edifici"
"Chieti"	47	"maschio"	"H 49"	"Trasporto terrestre e trasporto mediante condotte"
"Teramo"	43	"maschio"	"H 49"	"Trasporto terrestre e trasporto mediante condotte"
"Aosta"	65	"femmina"	"H 49"	"Trasporto terrestre e trasporto mediante condotte"
"Teramo"	54	"femmina"	"ND"	"Non Disponibile"

Figura 27 - Esito query selezionata dalle FAQ implementate

4.3 Modalità di rappresentazione grafica di ontoInfortuni

L'editore Protégé dispone di *plug-in* che permettono di illustrare l'ontologia, come ad esempio *OntoGraph*, utilizzato in questa trattazione per rappresentare graficamente le soluzioni adottate nel capitolo 3.

OntoGraph è un'applicazione che consente di navigare interattivamente tra le relazioni presenti nell'ontologia, sottoclassi, individui o istanze, proprietà di dominio o *range*, restrizioni, ma poiché è predisposto principalmente per lo sviluppatore, non pone in chiaro, nel grafo che produce, tutte le proprietà.

Aggiungiamo, infine, una modalità di lettura dell'ontologia che la renda più accessibile senza entrare nei dettagli implementativi, ma che renda leggibile il grafo selezionato di volta in volta su quegli elementi che rappresentano un punto importante o di interesse nello sviluppo e che non sarebbero visualizzabili con gli strumenti propri di Protégé.

La rappresentazione grafica è stata costruita *ad hoc* in modo da mettere in evidenza, rispetto ad una classe di interesse, le altre classi e le relazioni tra esse, tutte complete del loro nome, e caratterizzate da forme e colori significativi. L'applicazione, sviluppata in linguaggio PHP, utilizza la libreria Jena per la lettura del modello in OWL e le *query* SPARQL su di esso, e la libreria grafica *[Arbor].js* per la visualizzazione del grafo risultante dall'interrogazione, rispettando i ruoli degli elementi coinvolti in termini di classi, relazioni e proprietà.

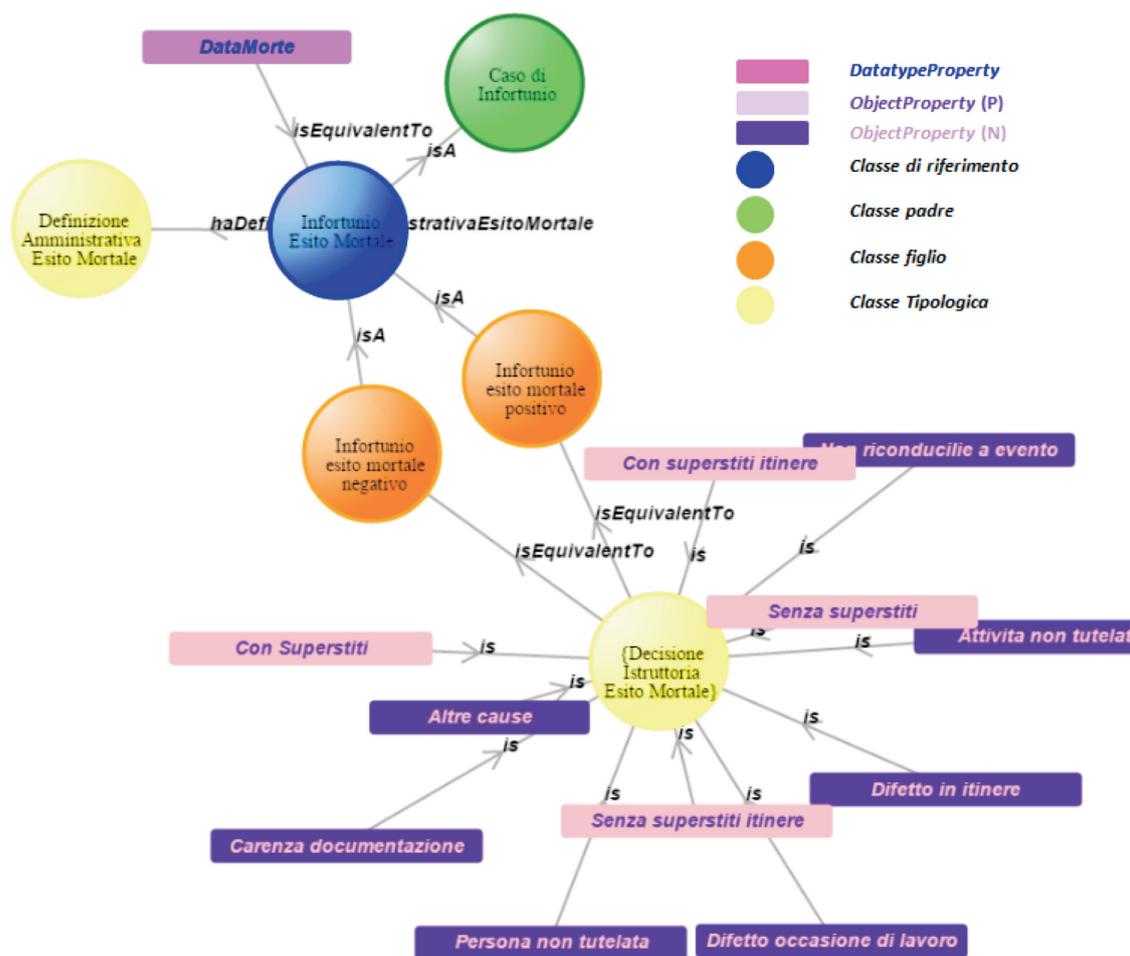


Figura 28 - Grafo relativo alla Decisione istruttoria esito mortale

Nella figura 28 viene rappresentato il grafo centrato sulla classe *InfortunioEsitoMortale* (in blu), completo di tutte le sue relazioni. La parte relativa alla Decisione istruttoria esito mortale (in giallo) rappresenta la visualizzazione grafica della tipologica già mostrata in modalità OWLDoc (paragrafo 4.1) nella figura 23. La legenda, presente nella figura 28, permette di individuare i ruoli e le relazioni e di trovare, ad esempio, l'esito positivo "Con superstiti" tra i rettangoli con fondo più chiaro, che corrisponde alla selezione dell'individuo *D_15* presentata in figura 24.

5 *Open Data* e ontologie nelle altre Pubbliche Amministrazioni

Avendo introdotto all'inizio i decreti vigenti per l'Amministrazione Digitale e più volte citato dati disponibili presso le altre PA, vogliamo riportare brevi cenni di quanto ad oggi presente sui portali di quelle amministrazioni con cui possiamo relazionarci, interessanti dal punto di vista del confronto con quanto realizzato in Inail.

Il punto di partenza è il portale dell'Open Data italiano a cura del Ministero per la Pubblica Amministrazione e la Semplificazione [ODPA-015].

Introduciamo per primo un portale [SPCData], di particolare interesse, quello del Sistema Pubblico di Connettività e Cooperazione - SPCData, ovvero il portale dei (*Linked*) *Open Data* in cui vengono raccolti i dati della Pubblica Amministrazione [CoSPC-012].

Nato all'interno dell'Agenzia per l'Italia Digitale, è attualmente in fase di integrazione con il portale del Ministero.

5.1 Portale dei *Linked Open Data* (SPCData)

Con il termine SPCData indichiamo il portale dei *Linked Open Data* (LOD) del Sistema Pubblico di Connettività e Cooperazione della PA da cui si possono scaricare i *dataset* SPC disponibili, interrogarli tramite uno SPARQL Endpoint, e utilizzare alcuni servizi dimostrativi che si basano sui *dataset* LOD. Alcuni di questi servizi sono rivolti alla ricerca di dati "elementari" (numeri di telefono, indirizzi, e-mail), oppure ad informazioni generali (quali PA si occupano di un argomento), altri riportano tabelle strutturate ed arricchite con rappresentazioni grafiche (come illustrato nell'esempio in figura 29).

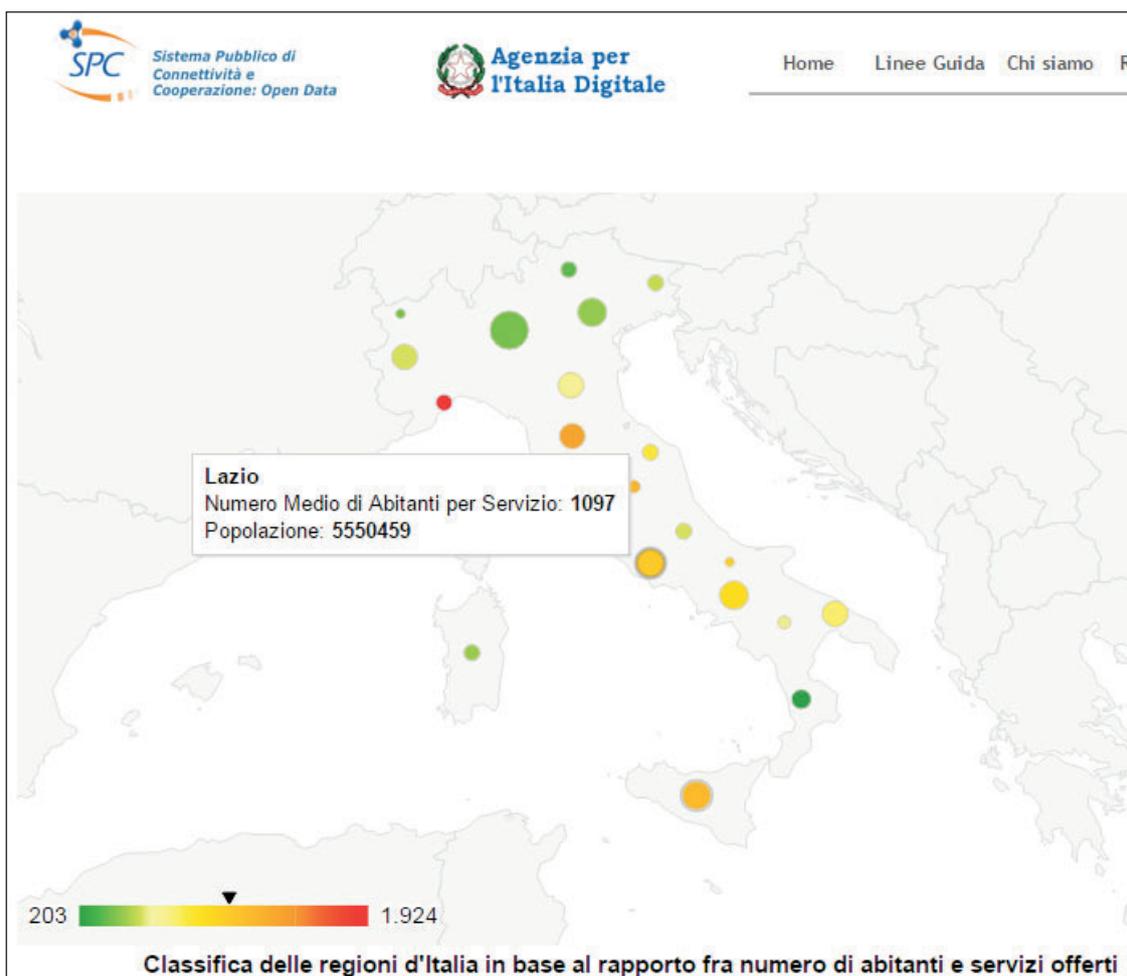


Figura 29 - Distribuzione degli OD come nella nota della figura

Da questo sito è possibile anche accedere e consultare le ontologie “ufficiali” dell’Agenzia Digitale per l’Italia, tra queste quella relativa alla rappresentazione geografica della Nazione Italia, con le sue regioni, province, comuni, ma anche quella relativa al codice Ateco 2007 per i settori di attività economica, o più semplicemente quella contenente l’indice delle PA italiane.

In particolare, le prime due potrebbero essere utilizzate anche in *ontoInfortuni*, in modo da potersi svincolare da specifiche tipologie create appositamente per la rappresentazione degli OD Inail. In figura 30 è illustrata una rappresentazione del codice Ateco 2007 (nell’esempio la schermata illustra le informazioni e i dati relativi al codice “20”). In questo caso, dal momento che il codice Ateco ha una rappresentazione gerarchica, è stato utilizzato lo standard XKOS, specifico per rappresentare classificazioni statistiche; con la relazione *ObjectProperty generalizes*, si mette in relazione l’elemento con il suo predecessore e con *specializes* lo stesso con i suoi successori.

at SPCData
<http://spcdata.digitpa.gov.it/browse/LinkedAteco2007/20>



Questa divisione include la trasformazione di materiale organico e inorganico grezzo tramite processi chimici e la formazione di prodotti specifici. Si distingue la produzione di elementi chimici di base, che costituiscono il gruppo industriale primario, dalla produzione di prodotti intermedi e finali ottenuti tramite un'ulteriore lavorazione degli elementi chimici di base.

Property	Value
?comment	<ul style="list-style-type: none"> Questa divisione include la trasformazione di materiale organico e inorganico grezzo tramite processi chimici e la formazione di prodotti specifici. Si ... »more»
is ?generalizes of	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007/C>
?generalizes	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007/201> <http://spcdata.digitpa.gov.it/LinkedAteco2007/202> <http://spcdata.digitpa.gov.it/LinkedAteco2007/203> <http://spcdata.digitpa.gov.it/LinkedAteco2007/204> <http://spcdata.digitpa.gov.it/LinkedAteco2007/205> <http://spcdata.digitpa.gov.it/LinkedAteco2007/206>
?inScheme	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007>
is ?includedItem of	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007>
is ?member of	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007/Divisioni>
?notation	<ul style="list-style-type: none"> 20
?prefLabel	<ul style="list-style-type: none"> FABBRICAZIONE DI PRODOTTI CHIMICI
is ?specializes of	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007/201> <http://spcdata.digitpa.gov.it/LinkedAteco2007/202> <http://spcdata.digitpa.gov.it/LinkedAteco2007/203> <http://spcdata.digitpa.gov.it/LinkedAteco2007/204> <http://spcdata.digitpa.gov.it/LinkedAteco2007/205> <http://spcdata.digitpa.gov.it/LinkedAteco2007/206>
?specializes	<ul style="list-style-type: none"> <http://spcdata.digitpa.gov.it/LinkedAteco2007/C>
?type	<ul style="list-style-type: none"> <http://www.w3.org/2004/02/skos/core#Concept>

Figura 30 - Ontologia per Ateco 2007 relativa al codice “20”

5.2 Open Data di INPS

Inps ha messo a disposizione una collezione di *open data* (643 al momento in cui è stato redatto questo *Quaderno*), divisi per argomento, fonte, periodo. È importante notare che i *dataset* Inps sono relativi a dati già elaborati e presentati in forma aggregata tabellare, sono pertanto di una granularità più grossolana rispetto a quelli dell’Inail che invece riportano i singoli casi con la granularità più fine che sia possibile ovvero rapporto uno a uno tra il dato e il singolo caso d’infornuto. I formati disponibili sono quelli consueti CSV, XLM, XLS, con l’aggiunta del formato OWL, che indica la presenza di un’ontologia di cui però non tutti gli *open data* sono forniti.

È stata anche pubblicata l’ontologia dei dati Inps, in cui vengono modellati i concetti fondamentali trattati da Inps negli *open data* compresi i tipi di *dataset* disponibili.

L’ontologia (*dominioInpsLinkedV1.0.owl*) definita dall’Inps, scaricabile dal sito, fornisce “*lo schema astratto di riferimento ufficiale per la rappresentazione semantica delle informazioni sulla sicurezza sociale e la mobilità, raccolte dall’Istituto Nazionale della Previdenza Sociale*”, inoltre “*garantisce l’accesso alle risorse descritte sia attraverso vocabolari semantici standard (Dublin Core, FOAF, Data Cube, SKOS, SDMX) sia attraverso vocabolari dedicati alle specifiche legali di utilizzo delle informazioni (LiMo) e alla definizione delle entità specifiche del dominio INPS*”⁷.

L’approccio adottato in questo caso è quello di modellare, attraverso l’ontologia, i concetti che sono presentati come *dataset* in forma tabellare.

⁷ La citazione è tratta dalla presentazione dei LOD Inps che si trova nel portale Inps nella pagina sugli Open Data.

Alcune di queste tabelle (come ad esempio quella mostrata in figura 33) forniscono dati aggregati derivati dal contributo dei diversi Enti che li producono e poi li forniscono ad Inps.

Nell'esempio illustrato in figura 31 si evidenzia che l'istanza Inail è collegata ad un certo numero di concetti (quali: aziende ispezionate, aziende irregolari,...) attraverso l'individuo *ID_56* che nell'ontologia è il *dataset* di riferimento per quelle informazioni la cui fonte, tra gli altri Enti, è Inail⁸.

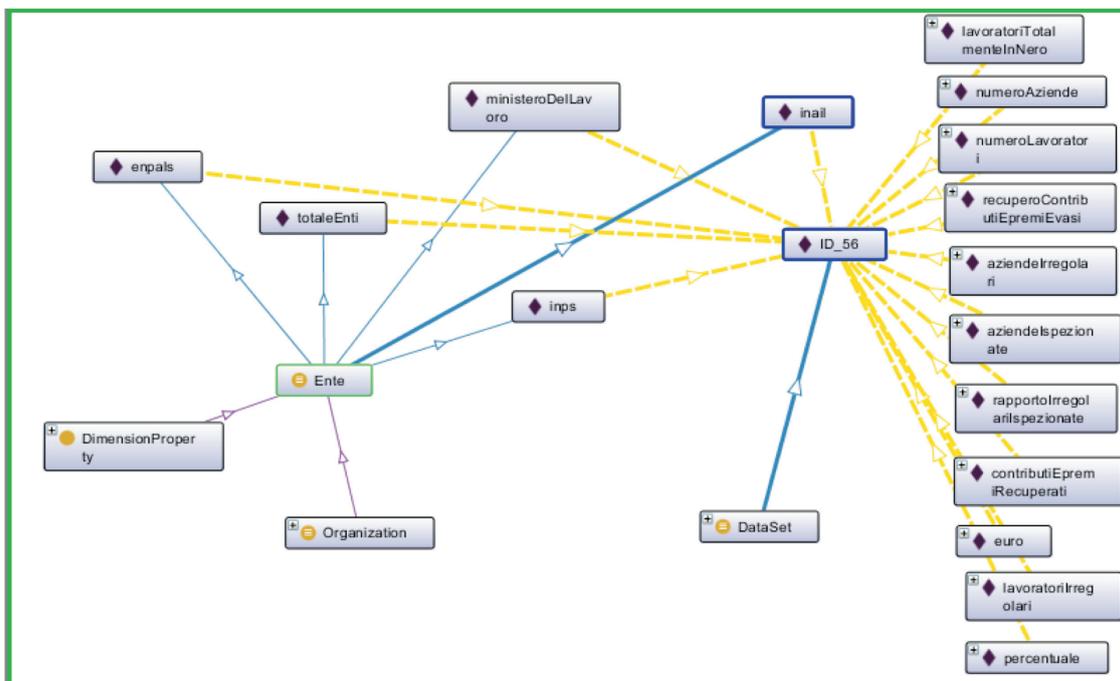


Figura 31 - Collegamento di alcuni concetti (dati) con Inail

Leggendo il file OWL attraverso Protégé, e aprendo il corrispondente file CSV con l'applicativo *excel*, dal confronto delle due figure seguenti, si nota come la cella della tabella, evidenziata in figura 33, viene rappresentata individualmente come una istanza (*observation_13*) nella figura 32 in cui si mostra sia il contenuto della cella (*measureValue*), sia la relazione con altri individui (per esempio *Aziende ispezionate*).

⁸ Le immagini dell'ontologia Inps sono state prodotte per questo *Quaderno* importando il file OWL di Inps nel sistema Protégé ed utilizzando le modalità di visualizzazione in esso disponibili.

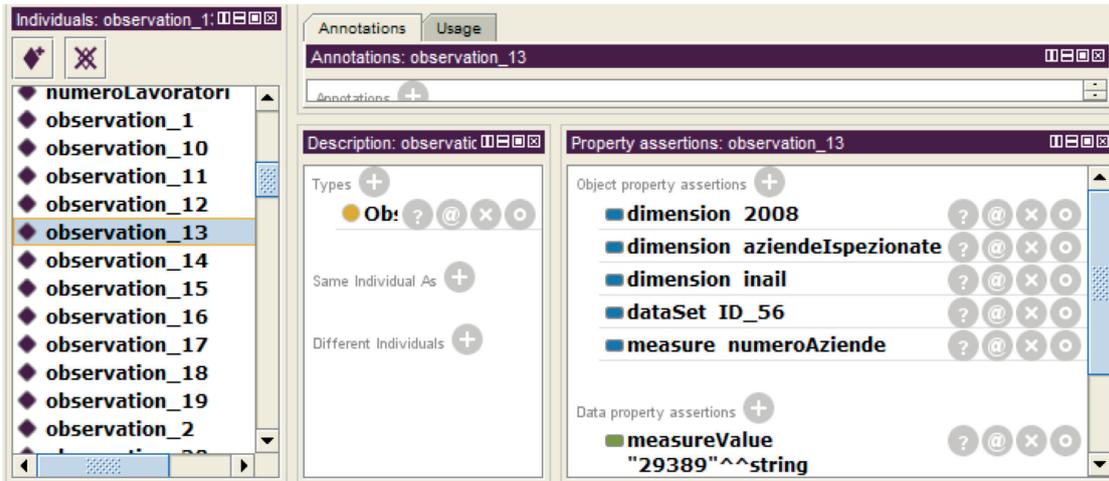


Figura 32 - Rappresentazione dell'individuo *observation_13*

Proseguendo nel confronto tra le figure 32 e 33, possiamo notare che, nel riquadro di destra della figura 32, l'*ObjectProperty* "dimension" è in relazione con gli individui (2008, *aziendeIspezionate*, *inail*) che corrispondono ai titoli delle colonne (*ispezionate*, la prima delle tre relative alle *Aziende*, 2008 l'anno di riferimento) e della riga di riferimento dell'Ente (nel nostro caso *INAIL*). La *DatatypeProperty* "measureValue" corrisponde al valore "29389" contenuto nella cella evidenziata in figura 33.

Tavola I.3.6.1 - Attività ispettiva di vigilanza per ente controllore, aziende ispezionate e lavoratori non regolari - Anno 2008-2009 e 1° semestre 2010 (valori assoluti e percentuali)

ENTE	Aziende			Lavoratori		Recupero contributi e premi evasi
	ispezionate	irregolari (a)	irregolari/ ispezionate	irregolari	totalmente in nero	
2008						
Ministero del Lavoro	188.655	92.885	49,2	173.289	49.510	1282.586.718,82
INPS	96.375	79.237	82,2	68.242,0	52.327	11548.010.000,00
INAIL	29.389	25.110	85,4	57.153	25.271	187.521.864,00
ENPALS	751	611	81,4	8.941	241	124.393.343,66
TOTALE	315.170	197.843	62,8	307.625	127.349	11.942.511.926,48

Figura 33 - Porzione dell'open data relativo alle attività ispettive di vigilanza sul lavoro

Il metodo utilizzato in Inps per produrre e presentare le tabelle dei suoi open data è quello comunemente seguito per rappresentare, attraverso le ontologie, tabelle di dati aggregati precedentemente elaborati in maniera da riprodurle ad ogni aggiornamento dei dati di origine.

5.3 Altri esempi di interesse

Si riportano due esempi di utilizzo dell'ontologia con l'obiettivo di rendere immediatamente fruibili le cosiddette informazioni sul "chi fa cosa?" che è spesso la forma che assume la domanda con cui iniziamo una ricerca di informazioni in un sito collegato all'argomento di nostro interesse. Spesso, se non si conosce la struttura e il contenuto del sito che può avere l'informazione cercata, la ricerca attraverso la navigazione nel sito stesso potrebbe non dare il risultato atteso.

In LOV, si intende vocabolario l'insieme delle definizioni di classi e proprietà utili a descrivere un certo dominio, inoltre si fornisce il legame con i dati, sempre attraverso l'indirizzo URI. Al momento LOV fornisce una scelta di centinaia di vocabolari, raggruppati sulla base di un certo numero di categorie, illustrate in figura 35.

I vocabolari devono rispondere a requisiti di qualità e stabilità degli stessi, e comunque la loro pubblicazione richiede la revisione e validazione da parte dei curatori del sito, per garantire che siano state adottate le migliori pratiche in questo settore, e deve superare i requisiti richiesti per la garanzia di qualità.

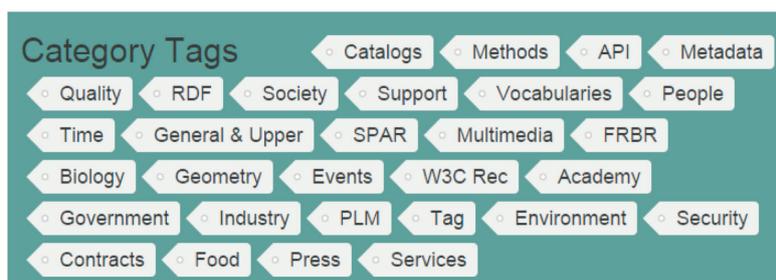


Figura 35 - Categorie disponibili in LOV

Il motivo per cui, a conclusione della panoramica di quanto esistente sugli open data, viene presentato questo sito è quello di stimolare un orientamento per il vocabolario proprio dell'Inail, ad esempio sui termini propri del dominio "assicurazione sul lavoro", perché venga realizzato secondo i criteri di riusabilità per renderlo poi disponibile attraverso LOV ad altri istituti omologhi. Analogamente scegliere, per alcuni settori di proprio interesse, dei vocabolari già esistenti con il vantaggio di doverli eventualmente solo personalizzare.

6 Considerazioni conclusive e possibili utilizzi

La proposta del presente *Quaderno*, rappresenta l'ontologia degli infortuni fortemente legata agli *open data* da cui è stata originata, sulla base delle scelte fatte dalle autrici per rendere più chiaro il modello concettuale che ne è derivato.

Rispetto alle altre amministrazioni che si sono servite delle ontologie per gli *open data* per ottenere una maggiore accessibilità agli *open data* stessi o alle informazioni di carattere organizzativo presenti nei siti, la proposta può essere utilizzata per condurre analisi sui dati originari sfruttando appieno la granularità dei dati e le potenzialità dell'ontologia di estrarre conoscenza, nel nostro caso, ad esempio, ai fini della prevenzione e riduzione degli infortuni.

L'ontologia *ontoInfortuni* non è la rappresentazione del formato dell'*Open Data* Inail per la loro visualizzazione nel senso che non riproduce le tabelle regionali (vedi Inps) ma è la struttura concettuale che pertanto risulta vuota ma è predisposta per essere agganciata ai dati attraverso i metadati modellati.

Il procedimento adottato è stato possibile grazie alla granularità fine dei dati, difatti se fossero state presentate come *open* le sole tabelle di dati aggregati presenti nell'allegato del *Quaderno1*, questo procedimento non avrebbe avuto senso.

Qualora venisse adottata anche per un uso sui dati completi degli infortuni sul lavoro, e non solo sugli *open data*, può essere migliorata in termini di essenzialità laddove la ridondanza è stata dettata da esigenze di legame con gli OD o addirittura ricostruita, in base alle nuove informazioni e alla nuova conoscenza, che possono venire fuori dall'analisi del *dataset* e degli ambiti in esso coinvolti.

Ciò significa che se si volesse modellare il dominio "infortuni sul lavoro" completo di tutte le sue componenti in termini di procedure amministrative, mediche ed economiche, l'ontologia che ne deriverebbe sarebbe completamente differente. Ciascuna delle componenti elencate diverrebbe a sua volta un dominio da modellare separatamente con l'ausilio di esperti.

Riferimenti bibliografici

[AAB-015] Agnello, P., Ansaldi, S., Bragatto, P., *Pooling Knowledge and Improving Safety for Contracted Works at a Large Industrial Park*, Work: a Journal of Prevention, Assessment and Rehabilitation; Vol. 51(3), 2015, pp. 537-548.

[AMAG-012] Ansaldi, S., Monti, M., Agnello, P., Giannini, F., *An Ontology for the Identification of the most Appropriate Risk Management Methodology*; P. Herrero et al. (Eds.): OTM 2012 Workshops, LNCS 7567, © Springer-Verlag Berlin Heidelberg, 2012, pp. 444-453.

[BCMGNP-003] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., *The Description Logic Handbook: Theory, Implementation, Applications*. Cambridge University Press, Cambridge, UK, 2003.

[BMSFC-009] Batres, R., Muramatsu, H., Shimada, Y., Fuchino, T., Chung, P.W., *Systematic Ontology Development of Accident Databases*, Computer Aided Chemical Engineering; 2009, 26, pp. 1171-1175.

[BCMM-014] Boselli, R., Cesarini, M., Mercorio, F., Mezzanzanica, M., (2014), *Are the Methodologies for Producing Linked Open Data Feasible for Public Administrations?*. In *Proceedings of Special Session in Knowledge Discovery meets Information Systems (KomIS) at DATA2014*, ISBN 978-989-758-035-2.

[BADs-013] Bragatto, P., Ansaldi, S., Delle Site, C., *A Pooled Knowledge Basis on Pressure Equipment failures to Improve Risk Management in Italy*, Chemical Engineering Transactions, 2013, 33, pp. 433-438.

[CGMBPA-008] Camossi, E., Giannini, F., Monti, M., Bragatto, P., Pittiglio, P., Ansaldi, S., *Ontology Driven Certification of Pressure Equipment*, Process Safety Progress, 2008, 27(4), pp. 313-322.

[CDFMV-013] Ciriello, G., De Felice, M., Mosca, R., Veltroni, M., *Infortuni sul lavoro. Un modello di lettura (della numerosità) su "open data" dell'Inail*, Inail, Quaderni di ricerca, 1- maggio 2013.

[ERS-010] Ebhrahimipour, V., Rezaie, K., Shokravi, S., *An Ontology Approach to Support FMEA Studies*, Expert Systems with Applications; 2010, 37(1), pp. 671-677.

[ECT-013] Elhdad, R., Chiamkurti, N., Torabi, T., *An Ontology-Based Framework for Process Monitoring and maintenance in Petroleum Plant*, Journal of Loss Prevention in Process Industries, 2013, 26, pp. 104-116.

[Gr-993] Gruber, T.R., *A Translation Approach to Portable Ontology Specifications*, Knowledge Acquisition, 1993, 5, pp. 199-220.

[Gr-995] Gruber, T.R., *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, International Journal of Human and Computer Studies, 43(5/6), 1995, pp. 907-928.

[Gu-995] Guarino, N., *Formal Ontology, Conceptual Analysis and Knowledge Representation*, International Journal of Human-Computer Studies, 43, Issue 5-6, Nov./Dec. 1995, 1995, pp. 625-640.

[Gu-998] Guarino, N., *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98)*, June 6-8, Trento, Italy, 1998*.

[Ho-011] Horridge, M., *A practical guide to building OWL ontologies using Protégé 4 and CO-ODE tools*, Edition 1.3, The University of Manchester, 2011.

[MWM-009] Morbach, J., Wiesner, A., Marquardt, W., *OntoCAPE – A (re)usable Ontology for Computer-Aided Process Engineering*, Computers and Chemical Engineering, 2009, 33, pp. 1546-1556.

[NMG-001] Noy, N.F., McGuinness, D.L., *Ontology development 101: a guide to creating your first ontology*, Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.

[Sm-999] Smith, B., *Ontology and Information Systems*, http://ontology.buffalo.edu/ontology_long.pdf, 1999.

[So-984] Sowa, J. F., *Conceptual Structures. Information Processing in Mind and Machine*, Reading, MA: Addison Wesley, 1984.

[SBFSC-006] Suzuki, M., Batres, R., Fuchino, T., Shimada, Y., Chung, P.W., *A Knowledge-based Approach for Accident Information Retrieval*, Computer Aided Chemical Engineering, 21(C), 2006, pp. 1057-1062.

[SRKR-997] Swartout, B., Ramesh, P., Knight, K., Russ, T., *Toward Distributed Use of Large-Scale Ontologies*, in Farquhar, A., Gruninger, M., Gómez-Pérez A, Uschold M., van der Vet P (eds) AAAI'97 Spring Symposium on Ontological Engineering. Stanford University, California, 1997, pp 138-148.

[TYDWTC-009] Tserng, H.P., Yin, S.Y.L., Dzung, R.J., Wou, B., Tsai, M.D., Chen, W.Y., *A Study of Ontology-Based Risk Management Framework of Construction Projects Through Project Life Cycle*, Automation in Construction; 2009, 18(7), pp. 994-1008.

[ZCZQC-009] Zhao, J., Cui, L., Zhao, L., Qiu, T., Chen, B., *Learning HAZOP expert system by case-based reasoning and ontology*, Computer and Chemical Engineering; 2009, 33(1), pp. 371-378.

Documenti disponibili su web:

[AGID-014a] <http://www.agid.gov.it/agenda-digitale>, aggiornamento 10 Dicembre 2014.

[AGID-014b] Agenzia per l'Italia Digitale, *Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico*, http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.7finale.pdf, 2014.

[Gr] Gruber, T. R., *What is an Ontology?*

<http://www-ksl.stanford.edu/kst/what-is-anontology.html>.

[CoSPC-012] Commissione di Coordinamento SPC, *Linee guida per l'interoperabilità semantica attraverso i linked open data*, Agenzia per l'Italia Digitale, <http://spcdata.digitpa.gov.it/index.html>, 12 novembre 2012.

[CSV-005] IETF, RFC 4180 – *Common Format and MIME Type for Common-Separated Values (CSV) Files*, <http://tools.ietf.org/html/rfc4180>, 2005.

[Inail-013] *Open data Inail*, <http://dati.inail.it/opendata/default/Infortuni/index.html>, 2013.

[ODPA-015] *I dati aperti della PA*, <http://www.dati.gov.it/>, 2015.

[OWL-004a] W3C, *OWL Web Ontology Language Guide*, <http://www.w3.org/TR/owl-guide/>, 2004.

[OWL-004b] W3C, *OWL Web Ontology Language Semantics and Abstract Syntax*, <http://www.w3.org/TR/owl-semantics/>, 2004.

[OWL-012] W3C, *OWL 2 Web Ontology Language Document Overview (Second Edition)*, <http://www.w3.org/TR/owl2-overview/>, 2012.

[RDF-014a] W3C, *RDF 1.1 Concepts and abstract Syntax*, RDF, 2014.

[RDF-014b] W3C, *RDF 1.1 Semantics*, RDF, 2014.

[RDF-014c] W3C, *RDF 1.1 Schema*, RDF, 2014.

[RDF-014d] W3C, *RDF Vocabulary Description Language 1.0: RDF Schema*, RDF, 2014.

[SPA-008a] W3C, *SPARQL Query Language for RDF*, <http://www.w3.org/TR/rdf-sparql-query/>, 2008.

[SPA-008b] W3C, *SPARQL Protocol for RDF*, <http://www.w3.org/TR/rdf-sparql-protocol/>, 2008.

[SPA-013] W3C, *SPARQL 1.1 Overview*, SPARQL, 2013.

Link

[Arbor] <http://arborjs.org/>

[ATECO-007] [<http://spcdata.digitpa.gov.it/classificazioni.html>]

[CNR-onto] <http://data.cnr.it/site/ontology>

[Jena] <https://jena.apache.org/documentation/>

[LOV] <http://lov.okfn.org/dataset/lov/>

[OntoGraf] <http://protegewiki.stanford.edu/wiki/OntoGraf>

[Protégé] <http://protege.stanford.edu/>

[RDF] http://www.w3.org/2011/rdf-wg/wiki/Main_Page

[SPARQL] http://www.w3.org/2009/sparql/wiki/Main_Page

[SPCData] <http://spcdata.digitpa.gov.it/index.html>

I Quaderni di ricerca

- 1 - G. Ciriello, M. De Felice, R. Mosca, M. Veltroni, *Infortuni sul lavoro. Un modello di lettura (della numerosità) su "open data" dell'Inail.*
- 2 - L. Cortis, F. M. Fabiani, C. Ratti, L. Rossi, D. G. Svampa, C. Vitale, *Analisi e valutazione del comportamento dell'assorbitore di energia nei sistemi di arresto caduta con differenti masse di prova.*
- 3 - P. Rossi, *Danno biologico: monitoraggio, criticità e prospettive d'interesse medico-legale a dieci anni dall'introduzione della nuova disciplina.*
- 4 - M. De Felice, A. Goggiamani, R. Mosca, M. Veltroni, *Malattie professionali. Un modello di lettura (della numerosità) su "open data" dell'Inail.*
- 5 - L. Cortis, F. M. Fabiani, L. Rossi, D. G. Svampa, *Nuovi strumenti per la valutazione dello scivolamento e ribaltamento delle scale portatili.*
- 6 - L. Cortis, F. M. Fabiani, L. Rossi, D. G. Svampa, *Determinazione della resistenza caratteristica dei dispositivi di collegamento montante-traverso nei ponteggi metallici fissi di facciata.*
- 7 - S. Carruba, R. Grimaldi, M. Sparro, S. Tomasini, *Vendor Rating: il modello di valutazione delle forniture IT dell'INAIL.*

