



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

## Technical Report / Mar. 2015



ODINet un framework innovativo per l'accesso e la diffusione on-line di dati strutturati ed eterogenei

Codice documento: **D4.1.1**

Numero revisione: **2**

Data revisione: **23/03/2015**

S. Pieroni, M. Franchini, S. Molinaro, L. Caterino, M. Faraoni, M. Toigo, F. Pitto, A. Greco, F. Mariani

*Il Progetto ODINet è realizzato con il determinante contributo della Regione Toscana a valere sul Programma Operativo Regionale cofinanziato dal FESR per l'obiettivo "Competitività regionale e occupazione" anni 2007-2013*



Regione Toscana



REPUBBLICA ITALIANA



Unione Europea  
Fondo Europeo di Sviluppo Regionale

**Le ali alle tue idee**

## Sommario

Introduzione .....	3
Dominio semantico e risorse digitali utilizzate.....	5
Architettura tecnologica e motore di ricerca .....	9
Risultati e validazione del sistema.....	11
Discussione .....	24
Bibliografia.....	25

## Indice delle Figure

Figura 1 – Overview della piattaforma ODINet .....	4
Figura 2 - Struttura principale di EUROVOC .....	5
Figura 3 - Livello principale della ontologia salute .....	6
Figura 4 - Livello principale della ontologia sociale .....	7
Figura 5 – Settore Vita Economica dell’ontologia economica.....	8
Figura 6 - Il grafo di ricerca.....	10
Figura 7 - Testo della ricerca: Anziani.....	12
Figura 8 – Anziani: consultazione INDICE DI SUPPORTO AI GRANDI ANZIANI in forma cartografica.....	13
Figura 9 - Anziani: consultazione INDICE DI SUPPORTO AI GRANDI ANZIANI in forma georeferenziale .....	13
Figura 10 - Tumore della mammella.....	14
Figura 11 - Screening per tumore alla mammella, percentuale di estensione .....	14
Figura 12 - Concetti e infografica restituiti da ODINet per la ricerca del concetto ‘determinanti di salute’ ..	16
Figura 13 - Concetti e grafo ontologico restituiti da ODINet per la ricerca del concetto ‘determinanti di salute’ .....	17
Figura 14 - Concetti e grafo ontologico restituiti da ODINet per la ricerca del concetto ‘indicatori sugli stili di vita’ .....	17
Figura 15 - Dati restituiti da ODINet per la ricerca del concetto ‘lavoro’ all’interno della fonte ‘Regione Toscana’ .....	18
Figura 16 - Dati restituiti da ODINet per la ricerca del concetto ‘disabilità’ all’interno della fonte ‘Regione Toscana’ .....	19
Figura 17 - Dati restituiti da ODINet per la ricerca del concetto ‘immigrati’ all’interno delle fonti ‘Regione Toscana’ e ‘Rete Osservatori Sociali’ .....	20
Figura 18 - Dataset restituiti dalla ricerca ‘spesa pubblica’ filtrata per fonte SoldiPubblici.Gov.it.....	21
Figura 19 - Grafo ontologico per la query ‘pensioni’ .....	22
Figura 20 – Grafo ontologico per la query ‘PIL’ .....	23

## Introduzione

ODINet è un progetto di ricerca e sviluppo, approvato nell'ambito del Programma Operativo Regionale attraverso il Fondo europeo di sviluppo regionale 2007-2013. Il progetto prevede la realizzazione prototipale di un motore di ricerca semantico in grado di catalogare i dati in un grafo ontologico, di estrarre le informazioni più rilevanti per le richieste degli utenti e restituirle in un modo altamente fruibile.

Il dominio applicativo riguarda i settori sociale, economico e sanitario, in modo da coprire la maggior parte dei dati in possesso degli enti pubblici nel contesto nazionale. Inoltre, i tre ambiti sono strettamente connessi fra loro ed offrono l'opportunità di indagini conoscitive trasversali, attraverso l'individuazione di ontologie che descrivono i concetti interconnessi e i collegamenti tra i vari argomenti. Nell'ambito di ciascun dominio abbiamo progettato un'ontologia allo scopo di indicizzare il motore di ricerca e migliorare la potenza e coerenza dei risultati della ricerca semantica. Abbiamo inoltre sviluppato un componente di data harvesting in grado di estrarre i dati dal Web, interfacciandosi ai portali della pubblica amministrazione italiana.

Abbiamo indicizzato i dati insieme alle ontologie tematiche attraverso un grafo di ricerca che costituisce il supporto principale per il motore semantico. Questo componente, che è disponibile come servizio Web, si basa sugli algoritmi e le metodologie proprie della Social Network Analysis (SNA), quali la centralità dei nodi e i fattori di clustering, al fine di individuare quelle serie di dati che sono più pertinenti alle interrogazioni effettuate dall'utente. Il motore di ricerca esegue una ricerca semantica sul grafo: le relazioni semantiche delle ontologie sono arricchite da due procedure in grado di identificare i concetti presenti in ontologie distinte che sono semanticamente correlati tra loro. Infine i dataset individuati vengono restituiti dall'interfaccia web. L'organizzazione generale della piattaforma ODINet è mostrata nel diagramma di Figura 1.

Il focus di questo report è in primo luogo la descrizione delle componenti semantiche della piattaforma, sottolineando come le ontologie sono state utilizzate per costruire un indice in forma di grafo.

Presentiamo inoltre una descrizione del nostro motore di ricerca ed infine una analisi dei risultati ottenuti nella fase finale di sperimentazione del motore stesso avvenuta attraverso query trasversali ai tre domini semantici.'

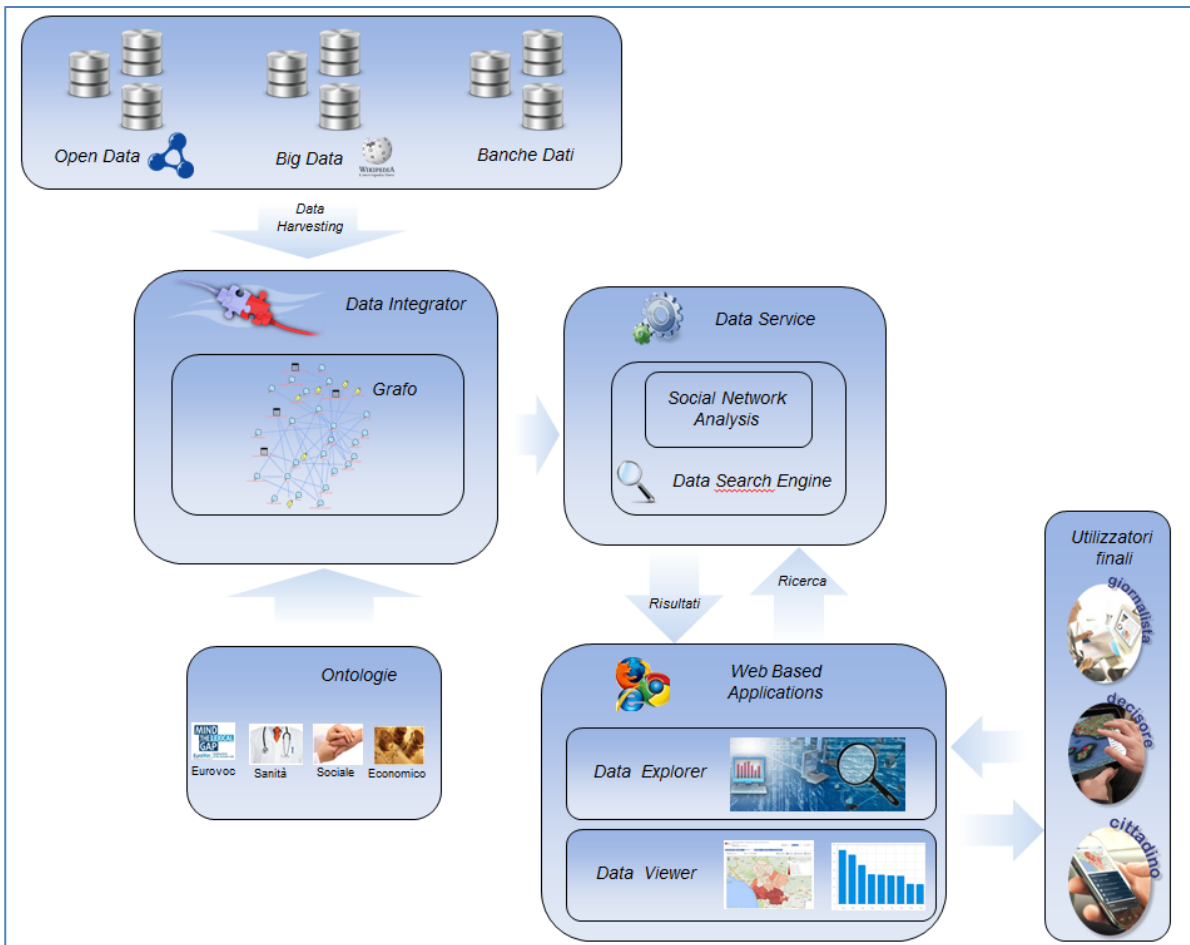


Figura 1 – Overview della piattaforma ODINET

## Dominio semantico e risorse digitali utilizzate

Costruire ontologie specializzate con il supporto di esperti del settore della conoscenza, richiede un grande sforzo di concettualizzazione e tempi molto lunghi[1], specie in settori così complessi e generali come quelli del dominio ODINet. Pertanto, la nostra scelta è stata di basarci su risorse standard a partire da EuroVoc [2], un thesaurus multilingue e multidisciplinare, gestito e mantenuto dall'ufficio pubblicazioni dell'Unione europea. EuroVoc è stato ampiamente utilizzato per software di classificazione e indicizzazione [3]. A partire dalla formalizzazione in SKOS [4] di EuroVoc e basandoci su altre risorse standard per il sociale, la sanità e il settore economico, abbiamo costruito le tre ontologie specializzate e creato le connessioni tra queste, a supporto del motore di ricerca semantico.

Poiché l'obiettivo principale del ODINet è di accedere e classificare una grande quantità di dati e di presentarli ad una vasta gamma di utenti, abbiamo scelto EuroVoc in quanto copre un insieme esaustivo di settori relativi alle attività delle istituzioni europee. Pertanto, nella fase iniziale del nostro lavoro, le ontologie sono state progettate per avere uno spettro orizzontale molto ampio, piuttosto che verticale, utilizzando un approccio top-down. In questo modo abbiamo sviluppato definizioni precise dei concetti di alto livello, che sono state successivamente approfondite in fase di validazione sui diversi casi d'uso. Dopo un'analisi approfondita dei vari settori e dei loro micro-thesauri, abbiamo scelto il dominio "Questioni Sociali" come risorsa principale per le ontologie sociale e sanitario. "Questioni Sociali" infatti affronta vari temi rilevanti per il progetto ODINet come la salute, la famiglia, la migrazione, la demografia, la struttura sociale e degli affari, della cultura e della religione fino alla protezione sociale. Per quanto riguarda l'ambito economico abbiamo scelto i settori "Vita Economica", "Scambi Economici e commerciali", "Finanze", "Impresa e Concorrenza", "Occupazione e Lavoro" ed "Industria" come risorse fondamentali.

Per l'editing delle ontologie abbiamo adottato [5] Protégé, che fornisce un ambiente di sviluppo concettuale e uno strumento grafico interattivo per la progettazione e realizzazione di ontologie. La versione italiana dei settori EuroVoc di cui sopra, è stata trasferita in Protégé attraverso la creazione di una classe per ogni settore e per ogni micro-thesaurus. Al fine di rispettare la gerarchia originale ogni micro-thesaurus è stato collegato ai suoi termini più ampi e ristretti attraverso la property subClassOf di Protégé. In coerenza con le specifiche EuroVoc, le relazioni hasBroader, hasNarrower ed isRelated sono state implementate all'interno di Protégé. Il processo di annotazione è stato basato principalmente sulle definizioni Skos di preferredLabel, hiddenLabel, seeAlso, isDefinedBy.

Abbiamo scelto la versione italiana di EuroVoc poiché la validazione viene condotta su dati forniti dalle istituzioni pubbliche italiane, pertanto le ontologie sono scritte in lingua italiana.

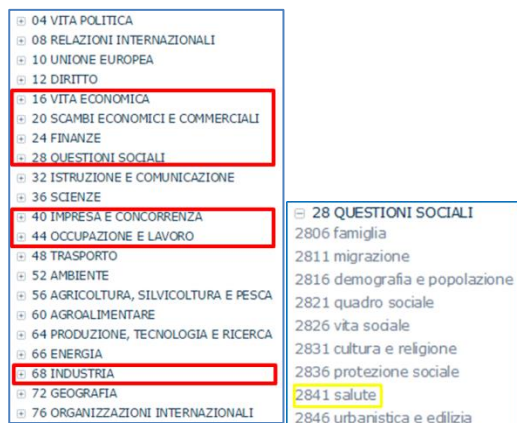


Figura 2 - Struttura principale di EUROVOC

### L'ontologia di dominio salute

L'obiettivo principale della ontologia di dominio salute è quello di fornire informazioni sulle malattie croniche, con particolare attenzione alle malattie cardiovascolari cercando di rispondere a 1) questioni generali provenienti dai cittadini e 2) domande specifiche provenienti dagli attori del sistema sanitario. Oltre ad EuroVoc, abbiamo utilizzato altre risorse specializzate, in particolare lo Unified Medical Language System [6] meta-thesaurus, un repository di concetti biomedici ed il Medical Subject Headings [7] thesaurus.

Quindi abbiamo migrato il micro-thesaurus '2841 salute' di Eurovoc (costituito delle gerarchie politica sanitaria, professione sanitaria, malattia, scienza medica, nutrizione ed industria farmaceutica) e una parte di MESH ovvero l'albero 'C14' delle Malattie Cardiovascolari in linguaggio OWL [8], divenuto uno standard del World Wide Web Consortium.

Questo passo della migrazione è stato fondamentale per eseguire l'editing di ontologie attraverso Protégé e la creazione di un file iniziale che fosse interoperabile. Successivamente, abbiamo identificato diverse estensioni, le principali riguardano i farmaci e le diagnosi di malattia: l'estensione relativa a farmaci ha comportato la migrazione dei primi due livelli dell'Anatomical Therapeutic Chemical System (ATC), sistema di classificazione adottato a livello

internazionale. L'estensione relativa alle diagnosi di malattia ha comportato la migrazione della nona revisione del sistema di classificazione internazionale delle malattie, (ICD-9-CM).

Il processo di annotazione è stato realizzato con il supporto del thesaurus MESH. Un'estensione finale è stata effettuata per mappare gli indicatori di salute e qualità delle cure in Toscana, sviluppato da Agenzia Regionale della Sanità (ARS Toscana). Abbiamo infine stabilito le relazioni tra i concetti delle diverse gerarchie, ad esempio, data una diagnosi come infarto miocardico acuto, l'ontologia fornisce i collegamenti con i le categorie di farmaco indicate per il trattamento e con i sintomi correlati. La Figura 3 fornisce una overview della ontologia di dominio salute.

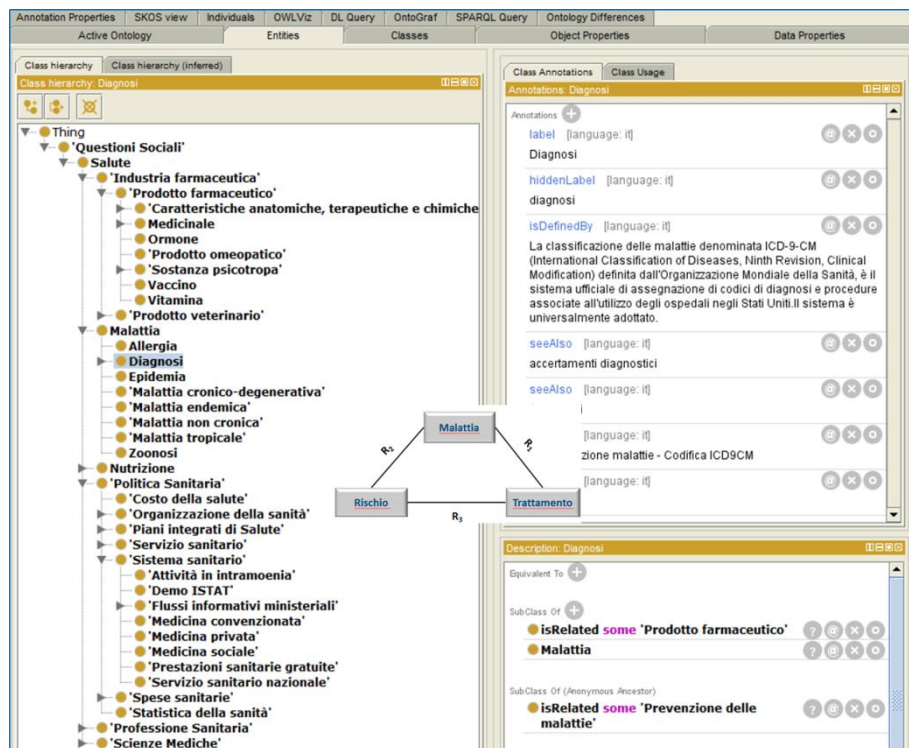


Figura 3 - Livello principale della ontologia salute

### L'ontologia di dominio sociale

La costruzione dell'ontologia per l'ambito sociale è partita dalla necessità preliminare di definire il termine "sociale". A questo proposito, sono stati valutati diversi approcci conoscitivi. Inizialmente, siamo partiti da un'impostazione di tipo epistemologico che ci ha condotto a identificare l'ambito sociale come quello delle "scienze sociali", ovvero tutte le scienze che studiano il comportamento dell'uomo come individuo e nell'ambito delle forme organizzate che storicamente si è dato. Il campo delle scienze sociali, anche limitando il campo di interesse ad alcune scienze sociali (sociologia, demografia), è però molto vasto e si è posta quindi la necessità di delimitare in modo più specifico e pragmatico il dominio conoscitivo oggetto del nostro lavoro. Dal momento che l'obiettivo pratico di ODINet è costruire un'applicazione (motore di ricerca) che consenta di trovare (principalmente) dati statistici pertinenti alle richieste dell'utente, abbiamo quindi scelto di fondare la costruzione dell'ontologia "sociale" sui concetti e le classificazioni già definite dal Sistema della statistica ufficiale internazionale e europea ed italiana in particolare.

In questo ambito è in corso da almeno 20 anni un lavoro di standardizzazione a livello internazionale che ha portato alla definizione di glossari, classificazioni, metodologie e concettualizzazioni condivise dai principali sistemi di produzione statistica internazionali. I sistemi che abbiamo preso come riferimento sono il lavoro di UNECE (United Nations Economic Commission for Europe), che ha elaborato il DISA (Database of International Statistical Activities in the UNECE Region) e un thesaurus delle attività di produzione statistica (Classification of statistical activities - CSA 2009) che ha portato all'identificazione di 5 domini principali:

- Domain 1: Demographic and social statistics
- Domain 2: Economic statistics
- Domain 3: Environment and multi-domain statistics
- Domain 4: Methodology of data collection, processing, dissemination and analysis
- Domain 5: Strategic and managerial issues of official statistics

Questa classificazione è stata confrontata con le classificazioni utilizzate presso Eurostat e Istat, individuando le differenze. Si tratta di classificazioni e concetti per lo più omogenei, con piccole differenze terminologiche che sono state recuperate in sede di costruzione dell'ontologia attraverso le alternative labels. L'ambito di lavoro è stato quindi

delimitato all'interno del Dominio 1 e del Dominio 3 e, all'interno di questi, ai temi più direttamente connessi alle politiche sociali e socio-sanitarie. I temi scelti per la costruzione dell'ontologia sociale sono quindi i seguenti:

- Dominio 1: Demografia e statistiche sociali
  - Settore 1.1: Popolazione e migrazioni
  - Settore 1.6: Protezione sociale
  - Settore 1.10: Politica e altre attività di comunità
- Dominio 3: Ambiente e statistiche multi dominio
  - Sotto-settore 3.3.1: Condizioni di vita, povertà e temi trasversali

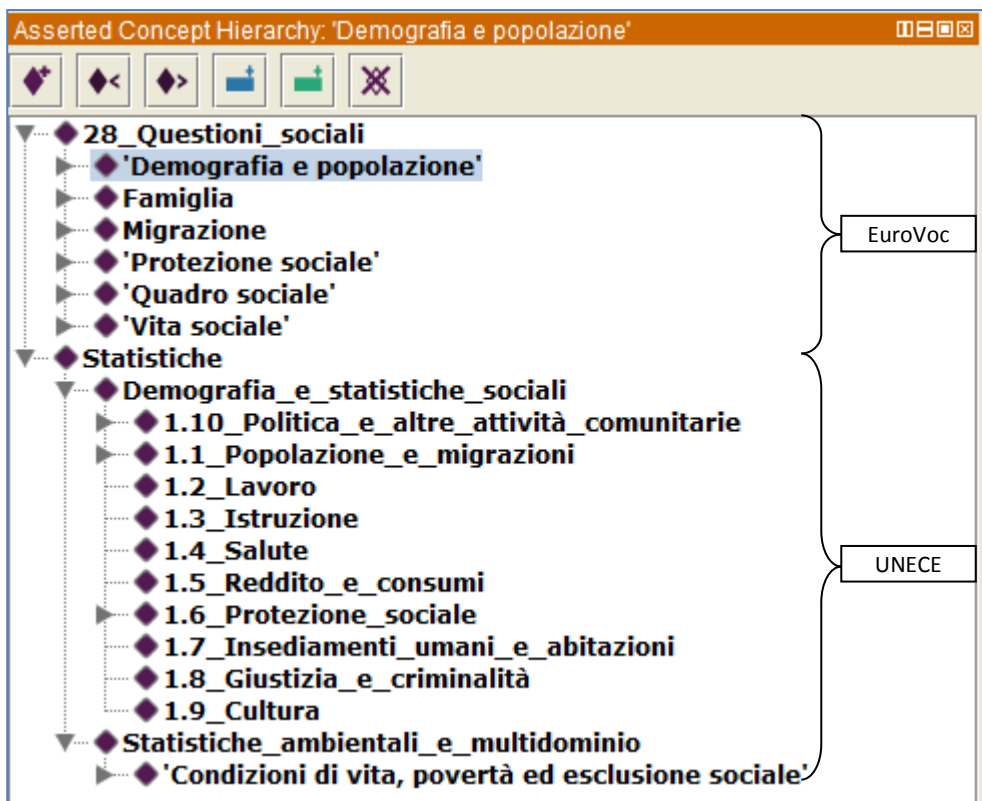


Figura 4 - Livello principale della ontologia sociale

### L'ontologia di dominio economico

La costruzione dell'ontologia di dominio economico è avvenuta all'interno di uno scenario in cui, tale ambito scientifico, presenta crescenti ambiti di sovrapposizione con gli altri due domini tematici oggetto della sperimentazione di ODINet, ovvero la sanità ed il sociale. L'ontologia costruita ha quindi perseguito il duplice obiettivo di 1) risultare completa e affidabile in riferimento all'ambito economico e 2) essere il più possibile trasversale rispetto a temi e concetti afferenti gli altri domini richiamati.

All'interno di EuroVoc, base comune alle tre ontologie tematiche, sono stati selezionati 43 microthesauri afferenti 6 settori EuroVoc legati all'area economica:

- Settore 16 - Vita economica
- Settore 20 - Scambi economici e commerciali
- Settore 24 - Finanze
- Settore 40 - Impresa e concorrenza
- Settore 44 - Occupazione e lavoro
- Settore 68 - Industria

Successivamente, l'ontologia economica è stata integrata attraverso il Glossario economico statistico multilingua dell'Istat, una risorsa fondamentale poiché contiene i metadati relativi all'intero repertorio delle statistiche ufficiali prodotte dall'Istat (principale *provider* di open data in Italia) e da altri enti del Sistan. Una volta individuato lo stakeholder di dominio coinvolto nel processo di validazione di ODINet - l'Istituto Regionale per la Programmazione Economica della Toscana (IRPET) - è acquisita la relativa base dati relativa ai conti economici della Toscana e alle unità



di lavoro presenti in regione nel periodo 2001-2012, si è quindi proceduto ad un ulteriore affinamento dell'ontologia, in modo che potesse 'recepire' le informazioni provenienti dai database Irpet.

La fase di editing dell'ontologia è stata sviluppata attraverso l'utilizzo di Protégé, il quale ha consentito l'arricchimento della risorsa attraverso l'inserimento delle annotazioni: in tal modo, i concetti SKOS presenti all'interno dell'ontologia sono stati arricchiti da proprietà e relazioni. In Figura 5 si mostra una overview dell'ontologia economica e del relativo processo di annotazione condotto.

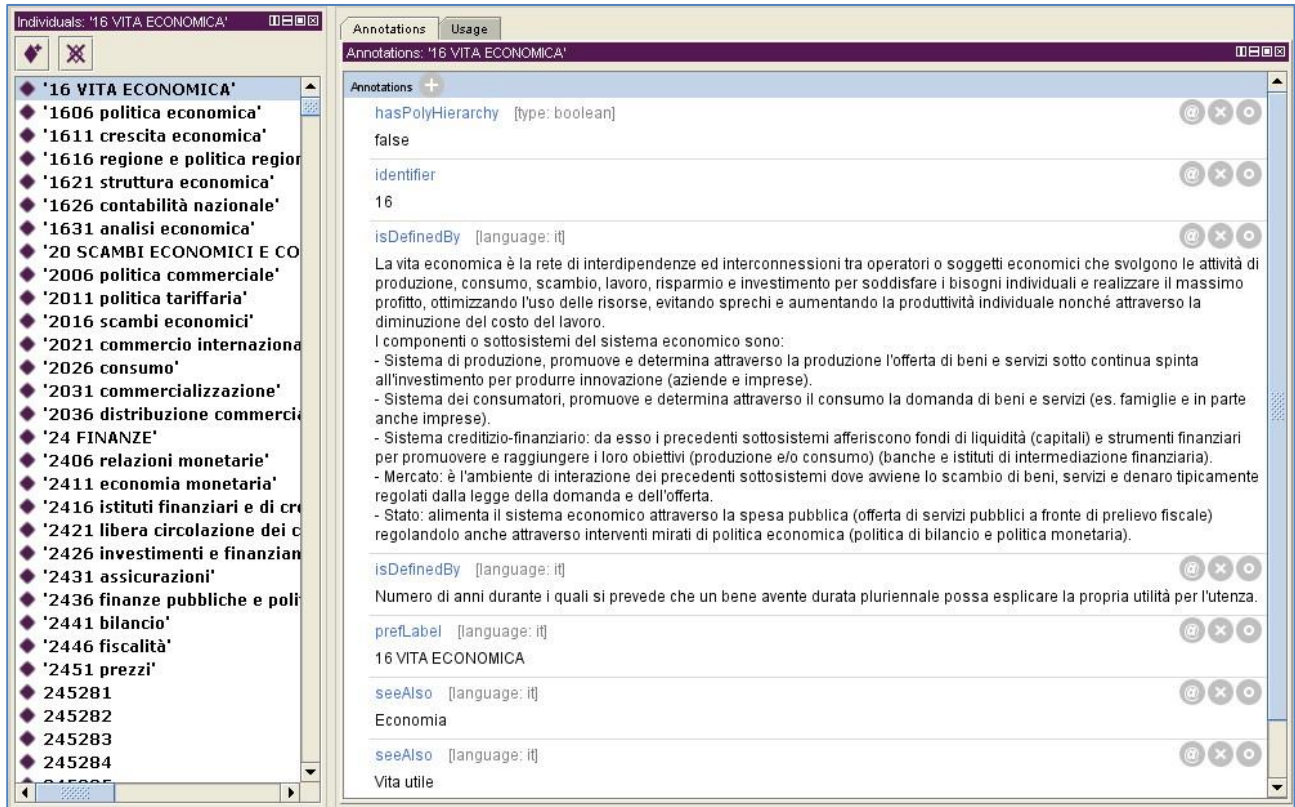


Figura 5 – Settore Vita Economica dell'ontologia economica



# Architettura tecnologica e motore di ricerca

## Il modulo di data harvesting

Abbiamo progettato e sviluppato un modulo in grado di interfacciarsi con i portali open data esistenti per l'individuazione e l'importazione di dataset disponibili sul web. In questo modulo è stato progettato uno strato di connettori universali che prendono in ingresso i formati e protocolli più diffusi con cui sono distribuiti attualmente i dati online (es. CKAN API, SDMX, OData, CSV, ShapeFile) ed espongono in uscita con un unico protocollo di output sia il contenuto del dato sia una sua descrizione (metadata).

I dataset vengono forniti principalmente da I.Stat (una banca dati ISTAT esposta sul web con il protocollo SDMX) e da dati.toscana.it (una piattaforma open-data sviluppata dalla Regione Toscana che espone dati con le CKAN API). Inoltre, alcuni stakeholder hanno aderito al progetto, fornendo i dati per lo scenario di validazione, vale a dire ARS Toscana (l'Agenzia Regionale di Sanità), IRPET (l'Istituto Regionale di Programmazione Economica della Toscana) e ROS (la Rete degli Osservatori Sociali).

## Il modulo di data indexing

Abbiamo costruito un grafo ontologico con ruolo di supporto principale al motore di ricerca. Il grafo permette di relazionare i concetti individuati dalle ontologie con i dataset importati con le procedure di web harvesting. La parte "concettuale" del grafo è stata costruita aggiungendo le relazioni tra concetti secondo i predicati definiti dalle ontologie. Poiché il grafo concettuale risultante era piuttosto sparso, abbiamo inoltre progettato e realizzato due procedure per potenziare le relazioni, aumentando in questo modo le informazioni semantiche.

Le procedure sono denominate WikiSimilarityDistance e WikiConceptConnection.

### a) WikiSimilarityDistance

Essa produce una stima della distanza semantica tra due concetti tramite un'approssimazione del Google Similarity Distance, misura che calcola la similarità semantica tra due parole (o due frasi) sulla base del numero di pagine indicizzate da Google in cui le due parole (frasi) appaiono insieme, rapportato al numero di pagine in cui appaiono singolarmente.

### b) WikiConceptConnection

Attraverso questa procedura siamo riusciti a trovare altre associazioni semantiche tra concetti sfruttando gli elementi nella sezione "Voci correlate" delle pagine di Wikipedia. Attraverso tale procedura siamo riusciti ad aggiungere circa 3.5K ulteriori relazioni per la ricerca sul grafo.

Per ogni dataset trovato importiamo quindi il titolo, le sue parole chiave, la sua descrizione, in modo da ottenere più informazioni possibili sul suo contenuto, oltre agli indicatori, ai nomi significativi delle colonne, ai valori più frequenti delle colonne ed al tema.

Ogni dataset viene quindi collegato nel grafo ai concetti espressi in esso. Abbiamo inoltre assegnato pesi diversi alle relazioni a seconda di dove e quante volte il concetto è stato trovato nel dataset.

La rappresentazione visiva del grafo di ricerca è mostrata in Figura 6. L'icona nuvola identifica i concetti, la tabella rappresenta i dataset, l'icona simile ad una etichetta rappresenta le label.

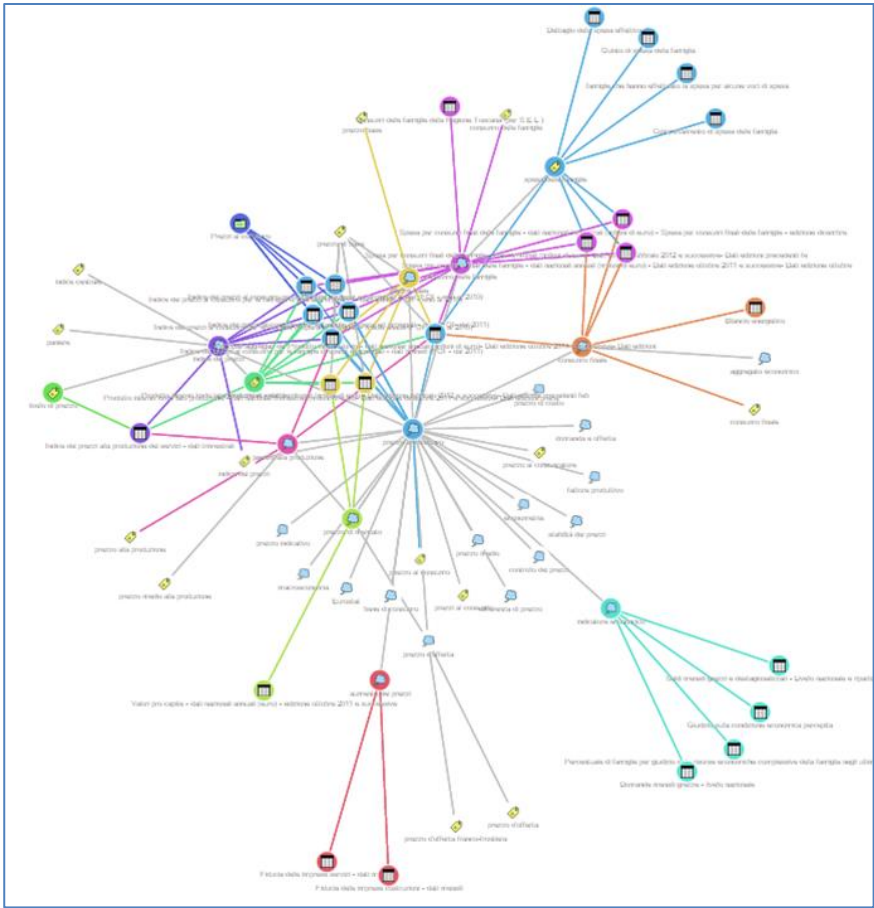


Figura 6 - Il grafo di ricerca

**Il modulo di data search**

Il risultato di una ricerca si ottiene a partire da un elenco di concetti e dataset ottenuti tramite una query full-text. Viene quindi effettuata una propagazione semantica della lista dei soli concetti ottenuti attraverso tecniche di Social Network Analysis, in particolare utilizzando l’algoritmo Page Rank, con lo scopo di trovare ulteriori concetti semanticamente connessi con quelli individuati nella fase precedente. I dataset direttamente connessi ai concetti trovati al termine di tale propagazione, insieme a quelli individuati direttamente tramite la ricerca full-text, vengono restituiti come risultato della ricerca.

**Il modulo di navigazione dei risultati**

Il risultato finale della ricerca viene restituito al termine della fase di propagazione semantica. La navigazione dei dataset è particolarmente fruibile considerando che i dataset sono consultabili sia in forma cartografica, sia in forma di grafico, sia in forma geo-referenziale. Inoltre la visualizzazione può essere guidata attraverso filtri impostati dinamicamente dagli utenti.

## Risultati e validazione del sistema

Un primo risultato di ODINet sono le tre ontologie che, oltre al ruolo primario che rivestono nel supporto semantico al motore di ricerca, possono funzionare indipendentemente e in modo integrato come base di conoscenza. Le tre ontologie nel loro insieme forniscono circa 4000 concetti distinti, con le relative relazioni ed etichette per la definizione dei sinonimi.

Il risultato più concreto ed importante del progetto è il motore di ricerca e la sua fruibilità per l'utente finale

Per effettuare la validazione del sistema nel suo complesso, abbiamo definito alcuni contesti e casi d'uso in base a problemi reali identificati con gli stakeholders.

L'attività di validazione, oltre a valutare la capacità di individuazione e di classificazione tematica dei dati, intende misurare la coerenza dei risultati forniti dal motore rispetto alle query impostate.

A fronte delle ricerche effettuate abbiamo quindi misurato la percentuale dei risultati restituiti che risulta "coerente" con le richieste.

### Validazione per il dominio salute

Nell'ambito del dominio salute, la definizione dello scenario di validazione delle tecnologie messe a punto da ODINet parte dalla analisi delle complessità di gestione di alcune patologie croniche, complessità in costante aumento anche per effetto dell'invecchiamento della popolazione.

Sia i dati di tipo clinico, normalmente storicizzati in cartella clinica/fascicolo sanitario elettronico, sia i dati di tipo amministrativo, disponibili attraverso flussi informativi ministeriali, presentano requisiti di privacy tali da non essere fruibili in modalità open.

Per questo motivo i dataset disponibili attraverso il motore di ricerca ODINet per il dominio salute sono esclusivamente in forma aggregata. In particolare attraverso ODINet non è possibile avere accesso a dati clinici, nè a dati per paziente. Sono invece disponibili dati in forma aggregata forniti principalmente dalla Agenzia Regionale di Sanità (ARS Toscana) e da altre fonti istituzionali quali l'ISTAT, l'INPS, la FONDAZIONE ZANCAN.

L'Ars ha messo a disposizione di ODINet una serie di indicatori di supporto alla predisposizione dei piani integrati di salute (Indicatori PIS) e la banca dati delle malattie croniche degli assistibili Toscani (Macro).

Considerata la tipologia di questi dataset abbiamo identificato due scenari:

- 1) il primo che fosse di supporto alle attività di governo della salute degli anziani
- 2) il secondo che fosse più vicino alle tematiche di medicina traslazionale nell'ambito delle patologie tumorali.

Mostriamo di seguito alcuni esempi che spiegano il tipo di validazione effettuata.

### Scenario 1: supporto alle attività di governo della salute degli anziani

Nell'ambito del primo scenario abbiamo chiesto al motore di ricerca la restituzione dei dati a fronte della query 'anziani'.

I concetti che il motore "conosce" in merito all'argomento richiesto, sono visibili in Figura 7.



Figura 7 - Testo della ricerca: Anziani

Come si può vedere la totalità dei concetti restituiti dalla ricerca risulta coerente con la query effettuata. Va sottolineato che alcuni dei concetti restituiti non sono affatto ovvi rispetto alla domanda, come per il caso di DIPENDENZA DEMOGRAFICA che consente l'accesso agli indicatori di invecchiamento. La coerenza e la completezza dei risultati si valuta ancora meglio esplosando i concetti restituiti dal motore. Ad esempio selezionando ASSISTENZA SOCIO-SANITARIA: ANZIANI si ottiene un insieme di 11 dataset di pertinenza:

1. Tasso di utenza zonale per assistenza domiciliare anziani
2. Spesa procapite per servizi residenziali per anziani
3. Spesa procapite per interventi domiciliari per anziani
4. Indice di copertura zonale delle strutture semiresidenziali in convenzione per anziani
5. Indice di copertura zonale delle strutture residenziali autorizzate per anziani
6. Indice di copertura zonale delle strutture residenziali in convenzione per anziani
7. Indice di copertura zonale delle strutture semiresidenziali autorizzate per anziani
8. Spesa procapite per servizi semiresidenziali per anziani
9. Indice di supporto ai grandi anziani (x 100)
10. Speranza di vita a 65 anni. Femmine
11. Speranza di vita a 65 anni. Maschi

Selezionando invece INVALIDITÀ CIVILE si ottiene un insieme di 6 dataset di cui il motore pre seleziona quelli evidenziati in grassetto, mostrando la sua capacità di effettuare un ranking di pertinenza dei risultati trovati.

- 1. Tasso di pensioni di invalidità civile (x 100)**
2. Tasso di pensioni di invalidità civile 20-64 anni (x 100)
- 3. Tasso di pensioni di invalidità civile 80+ anni (x 100)**
- 4. Tasso di pensioni di invalidità civile 65-79 anni (x 100)**
5. Tasso di pensioni di invalidità civile 0-19 anni (x 100)
6. Importo medio mensile pensioni invalidi civili erogate dall'INPS

Ognuno dei dataset è consultabile sia in forma cartografica (Figura 8), sia in forma geo-referenziale (Figura 9) aumentando in questo modo la portata informativa

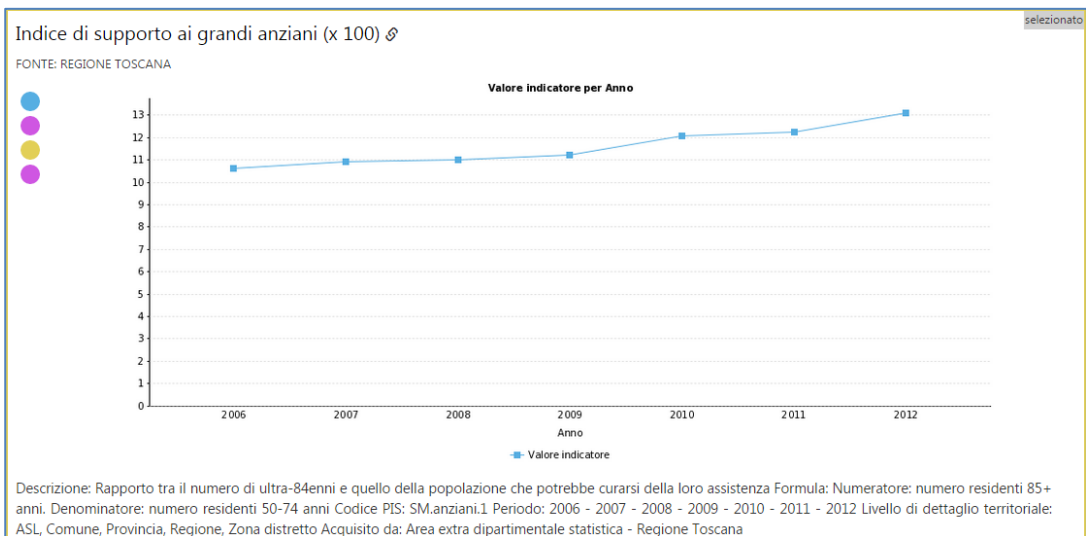


Figura 8 – Anziani: consultazione INDICE DI SUPPORTO AI GRANDI ANZIANI in forma cartografica

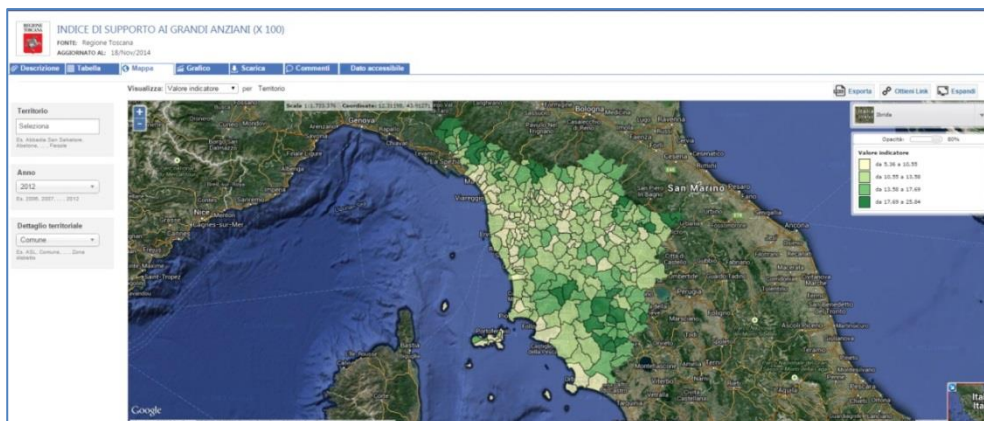


Figura 9 – Anziani: consultazione INDICE DI SUPPORTO AI GRANDI ANZIANI in forma georeferenziale

**Scenario 2: tematiche di medicina traslazionale nell’ambito delle patologie tumorali**

La medicina traslazionale, contribuisce all'individuazione di nuove terapie e strumenti diagnostici per il miglioramento della cura del paziente e cerca di colmare il gap tra i bisogni di tipo clinico e le soluzioni reali a tali bisogni.

In questa ottica risulta fondamentale riuscire a comprendere la complessità del processo della malattia, l'interazione del paziente con il sistema sanitario, le limitazioni nella diagnosi clinica e le sue implicazioni, l'impatto sulla salute dovuto alla (mancanza di?) aderenza del paziente alle politiche di prevenzione e alla sua (dis)percezione del rischio.

All'interno di questi temi molto generali, abbiamo scelto come tematica di validazione quella legata alle problematiche del tumore della mammella femminile.

Questo tipo di tumore è la neoplasia più frequentemente diagnosticata, rappresentando circa il 29% di tutte le diagnosi oncologiche nelle donne. È inoltre la principale causa di mortalità oncologica .

L’idea è dunque quella di cercare di acquisire un quadro conoscitivo della situazione per le donne residenti in Toscana.

Di seguito un esempio che spiega il tipo di validazione effettuata per questo scenario.

Abbiamo chiesto al motore di ricerca la restituzione dei dati a fronte della query ‘Tumore della mammella’.

I concetti che il motore "conosce" in merito all'argomento richiesto, sono visibili in Figura 10.



Figura 10 - Tumore della mammella

Questa query restituisce un quadro conoscitivo particolarmente esaustivo. I risultati forniti dal motore permettono infatti di valutare la complessità della malattia dando spunti sia per le limitazioni nella diagnosi clinica (dataset mortalità per tumore alla mammella) sia per l’impatto della malattia (dataset ricoveri per tumore alla mammella).

Inoltre il dataset ‘screening tumore mammella: percentuale estensione’ presente tra i ‘RISULTATI PER TUMORE DELLA MAMMELLA’ fornisce indicazioni sull’impatto dovuto alla (mancanza di?) aderenza del paziente alle politiche di prevenzione e alla sua (dis)percezione del rischio.

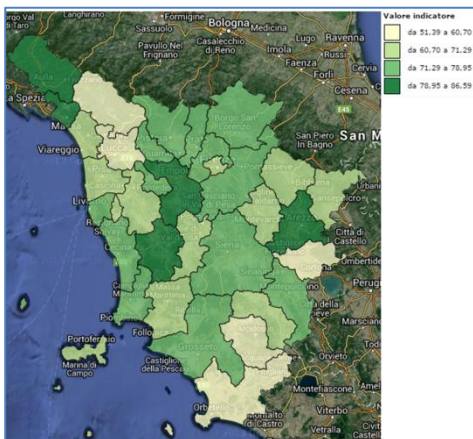


Figura 11 - Screening per tumore alla mammella, percentuale di estensione

Per sintetizzare i risultati dei test effettuati sul motore di ricerca ODINet relativi al dominio salute, si possono considerare in linea con quanto previsto. In particolare le singole query hanno mostrato un ottimo livello di coerenza tra chiave di ricerca e risultati restituiti. Laddove la coerenza non avesse raggiunto il 100% abbiamo identificato alcune lacune nei dataset sottostanti o nella definizione ontologica di supporto, che hanno determinato il completo raggiungimento dell'obiettivo.



## Validazione per il dominio sociale

Al fine di validare il prototipo ODINet, in relazione al dominio 'sociale' si è proceduto attraverso interrogazioni specifiche per testare il livello di coerenza e completezza dei risultati in relazione ai concetti e alle tematiche ricercati. Sulla base di tali obiettivi, e delle specifiche dei dataset caricati su ODINet in chiave di validazione, il test di qualità è stato dunque orientato in due diverse direzioni:

- I. verifica della **pertinenza dei risultati (concetti, relazioni tra concetti, dati) restituiti da ODINet in relazione alle chiavi di ricerca impostate**. Nello specifico si è cercato di valutare i concetti restituiti, le relazioni tra i concetti (capacità del motore di ricerca di estendere i risultati rispetto ai termini/concetti ricercati) e i dati/indicatori rintracciati.
- II. Verifica della effettiva **restituzione di alcuni dati/indicatori derivati da dataset campione in base alle ricerche impostate**. Su questo versante si è proceduto impostando soprattutto chiavi di ricerca pensate dal lato dei potenziali utenti di specifici dataset (figure preposte alla programmazione sociale e socio-sanitaria), in considerazione del fatto che una parte dell'ontologia del dominio sociale è stata creata proprio a partire da tali esigenze.

In merito al primo punto, si illustrano di seguito i risultati di n. 2 ricerche. Nella selezione dei temi e dei concetti da ricercare si è cercato di privilegiare quelli che in qualche modo potessero tagliare trasversalmente le tre ontologie. Questo sia per la natura delle tematiche sociali, che sono in qualche modo trasversali anche agli altri due domini sviluppati in ODINet, che per testare le capacità del motore di ricerca semantico di mettere in relazione concetti e dati appartenenti ai tre domini.

### Concetto ricercato: **Determinanti di salute**

Sono stati restituiti 6 concetti in relazione tra loro e 53 dati/indicatori puntuali che fanno riferimento a 5 diverse fonti (INPS, ISTAT, MeS-Scuola Sant'Anna, Regione Toscana e Rete degli Osservatori Sociali). Tra i concetti c'è già anche quello stesso di determinanti di salute, che racchiude al proprio interno 47 dati/indicatori che spaziano su temi assai differenti tra loro (reddito, situazione occupazionale, fragilità familiari, stili di vita, fattori ambientali). La risposta ricevuta, inoltre, mette in relazione altri dati sul reddito, sullo stato di salute, sui siti territoriali contaminati e l'intero Set di indicatori per il profilo di salute (che contiene, oltre agli indicatori sui determinanti di salute, indicatori di altre aree tematiche inerenti il concetto di salute quali stato di salute, livelli essenziali di assistenza sanitari e sociali, assistenza socio-sanitaria per gruppi target di popolazione). Sulla base di tali evidenze, si può dunque concludere che risulta un buon livello di coerenza e appropriatezza tra i concetti ricercati e i risultati restituiti, dato che sono state rintracciate relazioni tra concetti che hanno espanso la iniziale chiave di ricerca e che sono stati catturati, tra i dataset presenti in ODINet, tutti i dati in qualche modo riconducibili al tema ricercato.

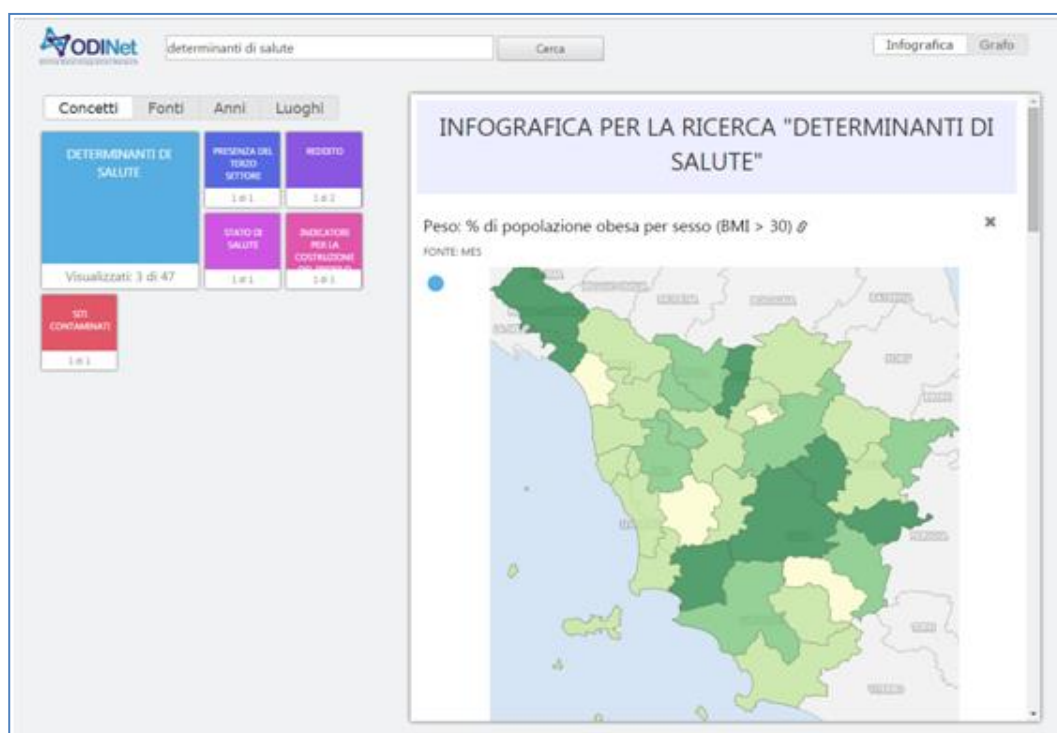


Figura 12 - Concetti e infografica restituiti da ODINet per la ricerca del concetto 'determinanti di salute'

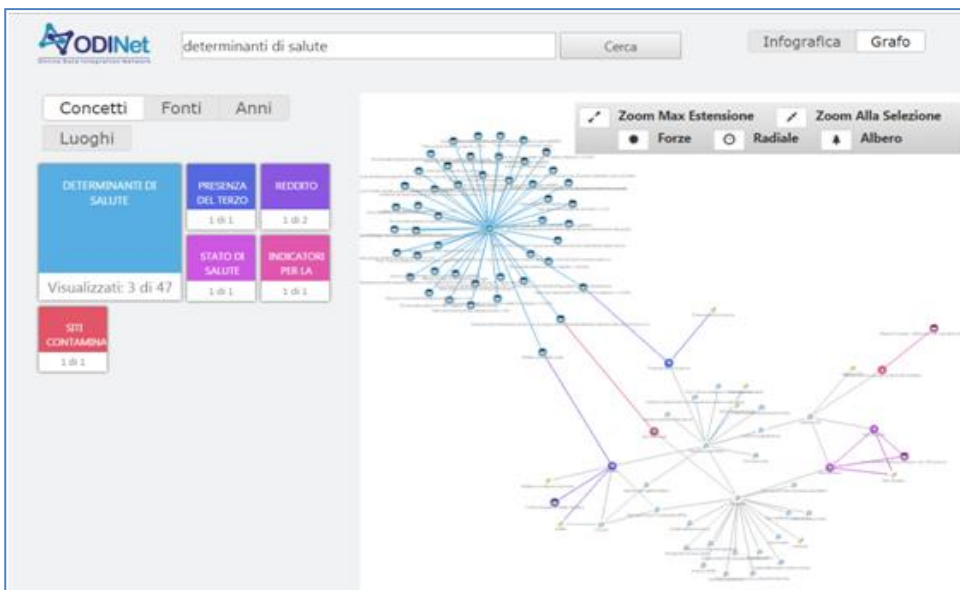


Figura 13 - Concetti e grafo ontologico restituiti da ODINet per la ricerca del concetto 'determinanti di salute'

**Concetto ricercato: Stili di vita**

Sono stati restituiti 7 concetti e 8 dati/indicatori puntuali che fanno riferimento a 5 diverse fonti (ISTAT, MeS-Scuola Sant'Anna, Ministero delle Finanze, Regione Toscana e Rete degli Osservatori Sociali). I risultati restituiti coprono quasi tutti i versanti sui quali vengono tradizionalmente indagati e misurati gli stili di vita (consumo di tabacco, abitudini alimentari, attività fisica, etc) ma occorre dire che non sono stati rintracciati dati/indicatori relativi al consumo di alcol, che dunque resta un elemento da migliorare. Un valore aggiunto che si riceve da ODINet su questa ricerca è quello delle connessioni tra i tre diversi domini perché sono messi in relazione i dati tipicamente riconducibili agli stili di vita con altri che hanno invece un collegamento indiretto (reddito – dominio economico, stato di salute – dominio sanità), cogliendo in questo senso la multidimensionalità di un concetto come quello degli stili di vita.

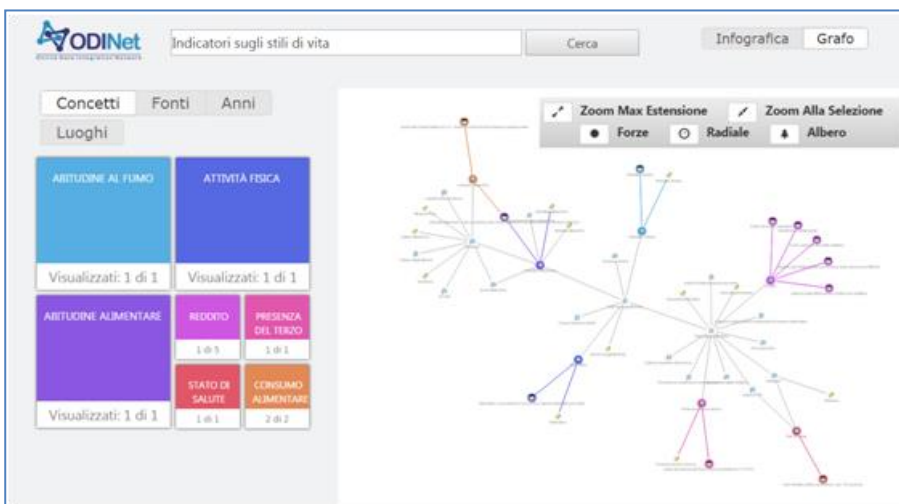


Figura 14 - Concetti e grafo ontologico restituiti da ODINet per la ricerca del concetto 'indicatori sugli stili di vita'

In relazione al secondo punto, si è scelto di concentrarsi sul dataset relativo al *Set di indicatori per il profilo di salute*. Si tratta di un patrimonio informativo fornito dall'Osservatorio Sociale Regionale della Regione Toscana, uno degli *stakeholder* del progetto e rappresenta uno degli strumenti che la Regione Toscana ha costruito per le AUSL/Zone distretto/Società della salute a supporto della programmazione sociosanitaria.<sup>1</sup> Il dataset contiene una vasta mole di indicatori di differenti aree tematiche (demografia, determinanti di salute, stato di salute, assistenza sanitaria, assistenza sociosanitaria a anziani, immigrati, disabili, minori e famiglie) e provenienti da fonti eterogenee (Istat, Inps,

<sup>1</sup> <http://servizi2.regione.toscana.it/osservatoriosociale/index.php?idDocumento=19157>

Inail, flussi doc sanitari, fonti amministrative ministeriali regionali e comunali, indagini epidemiologiche nazionali e regionali), per cui si presta ad interrogazioni provenienti da più versanti e può costituire una valida sollecitazione per il motore di ricerca.

**Concetto ricercato: Lavoro**

L'interrogazione restituisce tutte le misure legate con il concetto ricercato presenti all'interno del Set indicatori per il profilo di salute. Risulta particolarmente positivo che siano stati rintracciati indicatori afferenti a diverse aree tematiche presenti nel Set (determinanti di salute, immigrati, disabilità) e che sia stata dunque individuata la relazione con il concetto di 'lavoro'.

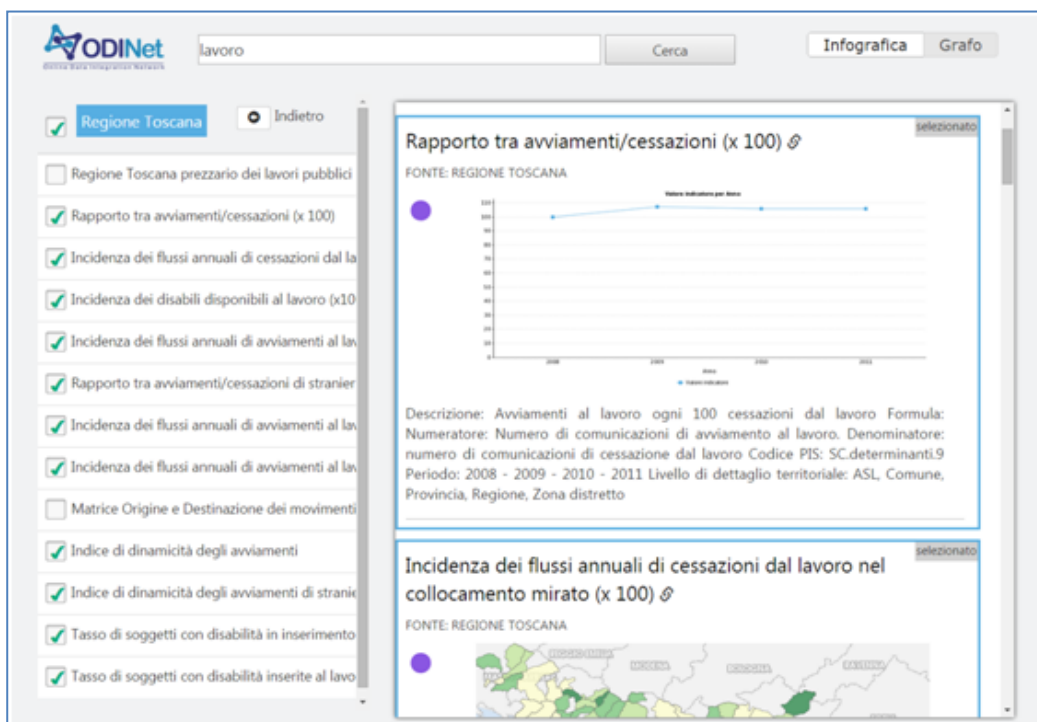


Figura 15 - Dati restituiti da ODINet per la ricerca del concetto 'lavoro' all'interno della fonte 'Regione Toscana'

**Concetto ricercato: Disabilità**

L'interrogazione restituisce tutte le misure legate con il concetto ricercato presenti all'interno del Set indicatori per il profilo di salute. Anche in questo caso, si riscontra la capacità di ODINet di mettere in relazione indicatori che all'interno del Set appartengono a differenti aree tematiche (assistenza socio-sanitaria disabili e determinanti di salute).

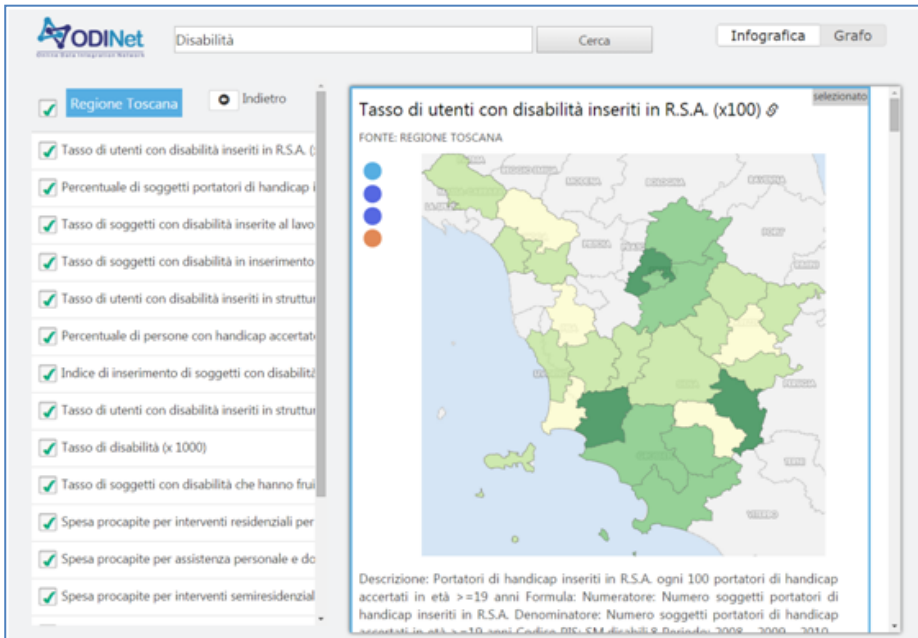
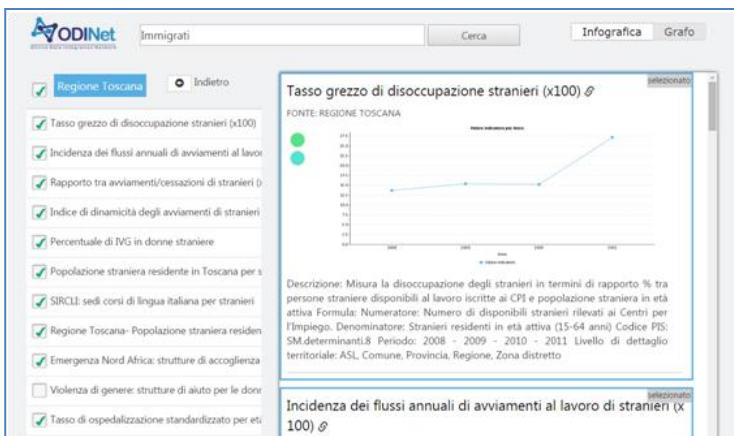


Figura 16 - Dati restituiti da ODINet per la ricerca del concetto 'disabilità' all'interno della fonte 'Regione Toscana'

**Concetto ricercato: Immigrati**

L'interrogazione restituisce tutte le misure legate con il concetto ricercato presenti all'interno del Set indicatori per il profilo di salute. Come per i due casi precedenti, si riscontra anche in questo frangente la capacità di ODINet di aggregare indicatori che all'interno del Set appartengono a differenti aree tematiche (determinanti di salute, assistenza socio-sanitaria immigrati e stato di salute).



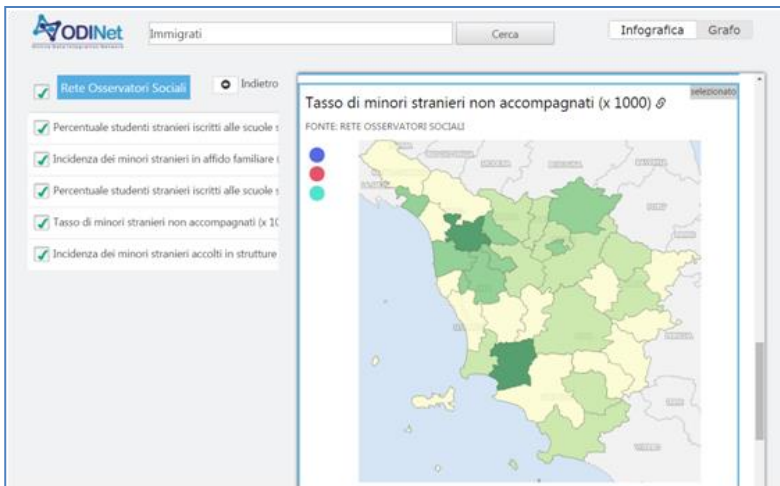


Figura 17 - Dati restituiti da ODINet per la ricerca del concetto 'immigrati' all'interno delle fonti 'Regione Toscana' e 'Rete Osservatori Sociali'

Per concludere, si segnala in sintesi che all'interno di entrambi gli approcci è stata riscontrata una soddisfacente capacità di ODINet di mettere semanticamente in relazione concetti appartenenti ai diversi domini e di restituire dati/indicatori coerenti con le chiavi di ricerca impostate, e dunque di fornire all'autore delle interrogazioni un quadro conoscitivo concettualmente più ampio di quello tradizionalmente desumibile da query effettuate attraverso motori di ricerca di uso comune.

## Validazione per il dominio economico

Per il dominio economico lo scenario di validazione del progetto ha preso in considerazione dei temi fortemente connessi con gli altri due domini coinvolti nella sperimentazione delle tecnologie di ODINet: il sociale e la sanità. I dati afferenti queste tre macro aree, seppur disponibili in misura quantitativamente rilevante, difettano tuttavia per l'estrema settorialità delle fonti, ognuna delle quali difficilmente si integra con le informazioni – pur concettualmente collegate – provenienti da altri *data provider*.

Per l'ambito economico la validazione del prototipo di ODINet è stata quindi condotta su scenari di riferimento trasversali ai tre domini. In particolare, i test condotti hanno inteso perseguire due principali obiettivi:

- 1) Verificare la coerenza dei *dataset* restituiti da specifiche chiavi di ricerca;
- 2) A partire dall'individuazione di alcuni *dataset* campione, verificare se specifiche chiavi di ricerca, concettualmente legate a quei *dataset*, restituiscono effettivamente i risultati attesi.

Per quanto riguarda il punto 1), ovvero la verifica di coerenza tra chiave di ricerca e *dataset* restituiti, la validazione ha utilizzato le seguenti *query*: Spesa pubblica; Pensioni; PIL.

**Spesa pubblica** - la chiave di ricerca restituisce 12 concetti legati alla *query*: finanziamento pubblico, spesa farmaceutica, spesa delle famiglie, spesa corrente, spesa di consumo, spesa per interventi residenziali, spesa per interventi semiresidenziali, spesa nazionale, spesa sociale, spesa sanitaria, spesa per assistenza personale e domestica a favore di persone disabili, servizio pubblico.

La pertinenza dei concetti restituiti dalla ricerca 'Spesa pubblica' consente anche che vi sia coerenza con i n.51 *dataset* mostrati dal motore di ricerca, che fanno riferimento alle principali fonti di dati sul tema: Istat (contabilità nazionale), SoldiPubblici.Gov.it (spese dei Comuni), Regione Toscana (spesa sociale), Fondazione Zancan (spesa sociale). L'immagine seguente mostra i *dataset* restituiti attraverso la *query* 'spesa pubblica' dal provider SoldiPubblici.Gov.it

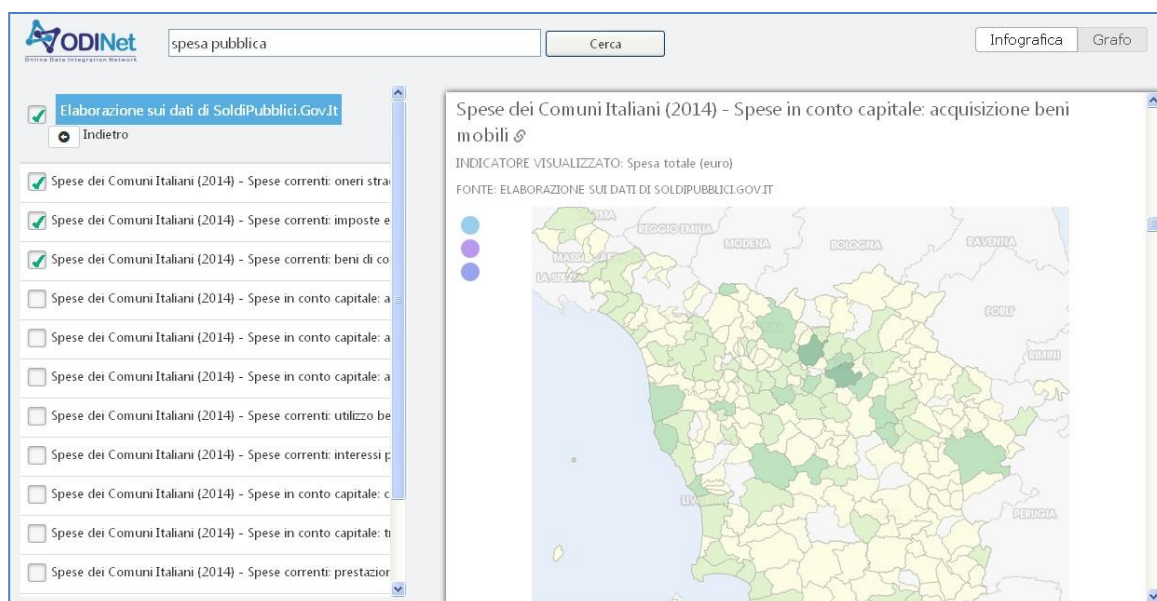


Figura 18 - Dataset restituiti dalla ricerca 'spesa pubblica' filtrata per fonte SoldiPubblici.Gov.it

**Pensioni** – la chiave di ricerca restituisce 12 concetti legati alla *query*: pensionato, anziano, previdenza sociale, pensione d'invalidità, pensione d'invalidità civile, pensioni sociali, pensioni ai superstiti, pensione di vecchiaia, pensione integrativa, lavoro sociale, disabile, prestazione sociale. Pare rilevante sottolineare come i risultati rimandino a concetti che non sono legati esclusivamente alla dimensione economica del concetto (la spesa pensionistica), ma che ne consentono di esplorare anche gli aspetti più specificamente legati agli ambiti socio-sanitari (pensioni sociali, civili e di invalidità) e demografici (indici di vecchiaia della popolazione): ciò è reso possibile dal *matching* ontologico effettuato per le 3 ontologie di dominio economico, sociale e sanitario sviluppate per il progetto. Come è possibile osservare dall'immagine seguente, la rappresentazione a grafo data dal motore di ricerca restituisce l'elevata densità e connessione tra i concetti e i *dataset* legati a 'pensioni', consentendo così – sul lato Infografica – di poter accedere a n.42 archivi dati in grado di esplorare il concetto in maniera multidimensionale.



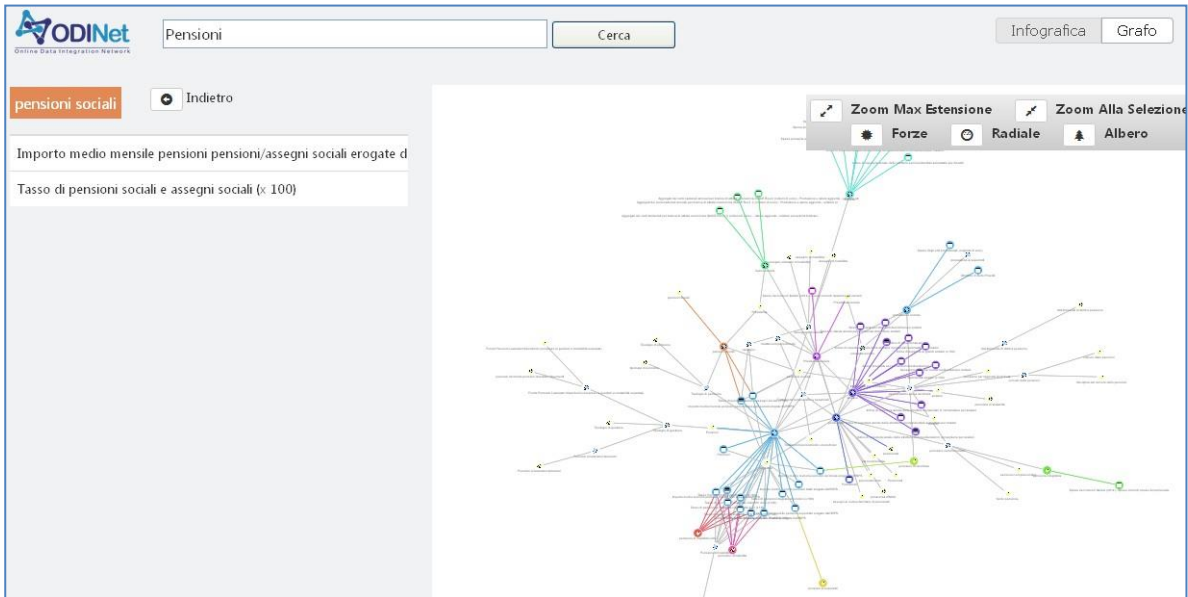


Figura 19 - Grafo ontologico per la query 'pensioni'

**PIL (o Prodotto Interno Lordo)** – la chiave di ricerca restituisce 15 concetti legati alla *query*: valore aggiunto, consumo finale, consumo, spesa nazionale, prodotto nazionale, economia, prezzo al consumo, consumo delle famiglie, consumo d'energia, prezzo alla produzione, beni e servizi, reddito nazionale, consumi finali delle famiglie, consumo interno, spesa di consumo. Anche in questo caso la chiave di ricerca individua correttamente i concetti legati al termine PIL, in particolare le principali voci ed aggregati che lo compongono<sup>2</sup> presenti all'interno dell'ontologia di dominio economico. I risultati della ricerca effettuata consentono quindi, rispetto al valore complessivo della ricchezza nazionale prodotta in un dato periodo, di discriminare anche il diverso contributo – e l'evoluzione nel tempo – portato dalle singole voci. Anche in questo caso, l'estrema ricchezza informativa restituita dalla *query* 'PIL' è resa possibile dalla densità del grafo ontologico, che presenta un elevato grado di connessione tra concetti e dataset, con la pressoché assenza di nodi isolati. Il risultato finale della ricerca è dato da n.47 dataset.

<sup>2</sup> Il PIL è pari alla somma dei beni e dei servizi finali prodotti da un Paese in un dato periodo di tempo. Si dice *interno* perché si riferisce a quello che viene prodotto nel territorio del Paese, sia da imprese nazionali sia da imprese estere. [...] Il PIL può essere misurato in tre diversi modi. Il primo consiste nel considerare la somma dei prodotti e servizi finali di un paese. [...] c'è un secondo metodo: sommare il valore aggiunto dell'economia in un dato periodo di tempo, dove per *valore aggiunto* si intende la differenza tra il valore della produzione di ogni impresa e il valore dei beni intermedi usati dalle stesse imprese nella produzione. Infine, poiché il valore della produzione è uguale a quanto è stato pagato dalle imprese per i fattori di produzione utilizzati (lavoro, capitale e servizi pubblici), il PIL può essere calcolato come somma dei redditi (salari e stipendi dei lavoratori, profitti delle imprese, imposte pagate allo stato) di tutta l'economia.



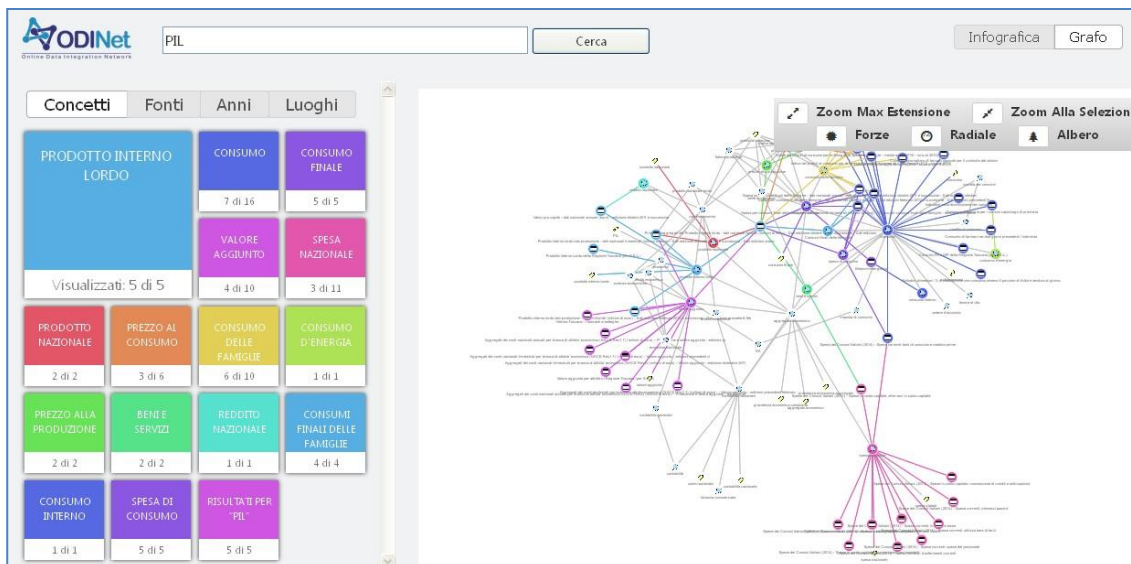


Figura 20 – Grafo ontologico per la query 'PIL'

Per quanto riguarda il punto 2), ovvero la verifica sulla effettiva restituzione di un determinato *dataset* a partire da chiavi di ricerca concettualmente legate ad esso, si riportano i risultati di n.3 ricerche effettuate:

**Valore aggiunto in Toscana per macro settore ATECO** – il *dataset* preso in considerazione è stato fornito da uno degli *stakeholder* di progetto coinvolti per la validazione: IRPET. L'archivio dati contiene le informazioni sul valore aggiunto prodotto nelle imprese e amministrazioni pubbliche della Toscana, nel periodo temporale 2001-2012, con il dettaglio del macro settore di attività economica ATECO e aggregazione territoriale per Sistema Economico Locale.

Quattro delle 5 ricerche effettuate (Agricoltura, PIL, Valore aggiunto, Irpet) hanno restituito il risultato atteso, mentre in un caso l'esito è stato negativo: la ricerca 'Casentino' fa riferimento alla tipologia di aggregazione territoriale (Sistema Economico Locale) utilizzata dal *dataset* prodotto da IRPET, ma non dagli enti SISTAN. Occorrerà, in tal senso, implementare l'ontologia attraverso l'inserimento dei 42 SEL toscani, che andranno così ad affiancarsi ad altre tipologie di aggregazioni territoriali già presenti (Regioni, Province, Asl, Zone distretto, ecc.).

**Importo medio mensile pensioni vecchiaia erogate dall'INPS** – il *dataset*, prodotto da INPS, offre informazioni relative all'importo medio mensile delle pensioni maturate attraverso la compresenza di età anagrafica, contribuzione minima e cessazione del rapporto di lavoro. Il periodo temporale coperto va dal 2007 al 2012, mentre il livello di dettaglio territoriale è per Asl, Comune, Provincia, Regione, Zona distretto.

Le ricerche effettuate hanno dato esito positivo in tutti e 4 i casi considerati, i quali hanno preso in considerazione gli aspetti strettamente connessi al contenuto informativo dell'archivio (pensioni), i beneficiari delle prestazioni (anziano), la macro area di riferimento (Previdenza sociale) e l'ente che eroga gli assegni mensili (INPS), che in questo caso corrisponde anche all'ente *provider* dei dati.

**Indicatori sulla distribuzione del reddito netto familiare** – il *dataset*, prodotto da Istat, riporta informazioni relative all'omogeneità nella distribuzione del reddito netto familiare (indice di Gini), relativamente agli anni dal 2003 al 2011 e con i seguenti livelli di dettaglio territoriale: Regione, ripartizione, nazione.

Dai risultati del test effettuato attraverso 5 diverse *query* di ricerca (ricchezza, povertà, reddito, distribuzione del reddito, disparità economica) emerge come in tutti i casi considerati la chiave di ricerca inserita abbia restituito il *dataset* relativo alla distribuzione del reddito.

In sintesi, i risultati dei test di qualità effettuati sul prototipo di progetto, e relativi al dominio di sperimentazione 'Economia', possono essere considerati allineati rispetto agli obiettivi posti. In particolare, così come riportato in precedenza, le singole casistiche analizzate hanno mostrato l'assoluta coerenza tra chiave di ricerca e *dataset* mostrati dal prototipo, grazie alla natura semantica del motore di ricerca che sfrutta in maniera ottimale le basi ontologiche sviluppate a supporto. I test hanno altresì consentito di individuare alcuni nodi scollegati all'interno dell'ontologia, fattispecie che può limitare le potenzialità del motore di ricerca semantico e che, quindi, richiederanno un ulteriore lavoro di sviluppo e affinamento dell'ontologia economica e relativo *matching* con le altre due ontologie di dominio sviluppate per la sperimentazione (sociale e sanità).

## Discussione

Uno dei punti di forza di ODINet riguarda sicuramente l'innovazione: è un motore di ricerca specializzato sui dataset che acquisisce dati dal web in modo automatizzato e usa al meglio gli algoritmi basati sulle tecnologie standard della SNA (Social Network Analysis) per la ricerca semantica. Dalla ricerca effettuata in letteratura possiamo affermare che non ci sono motori di ricerca semantica con caratteristiche simili.

In termini di validazione complessiva del sistema, i risultati dei test effettuati sul motore di ricerca ODINet si possono considerare soddisfacenti. Complessivamente le query effettuate sugli scenari individuati hanno mostrato un ottimo livello di coerenza tra chiave di ricerca e risultati restituiti. Laddove la coerenza, ovvero la forza delle connessioni tra concetti e dataset sottostanti, non ha raggiunto il livello massimo, la causa è principalmente da attribuirsi alla mancanza di dataset specifici e/o nel dettaglio delle meta-informazioni associate a tali dataset e/o nella definizione ontologica di supporto. Questi aspetti, che grazie all'esito della fase di validazione potranno essere migliorate in futuro.

Come ulteriore spunto di miglioramento, va segnalata la necessità di rafforzare la strutturazione delle relazioni tra concetti sia comuni che trasversali alle tre ontologie sviluppate. Questo potenzierebbe le capacità di risposta di ODINet in termini di interdisciplinarietà dei quadri conoscitivi restituiti, un valore aggiunto soprattutto per l'utenza composta da ricercatori, decisori politici e da professionalità preposte alla programmazione.

Così come sviluppato in fase prototipale, ODINet si rivolge ad una platea di utenti generica, composta anche di cittadini ed imprese. Sempre in ottica di miglioramento dello strumento potrebbe essere previsto un sistema di 'riconoscimento' dell'utente più esperto (il caso più tipico è quello del ricercatore), ad esempio attraverso login profilate, da utilizzare per mantenere in memoria i risultati restituiti dalle query già sottoposte al sistema e 'ripuliti' dall'utente sulla base dei propri specifici interessi di ricerca.

Evidenziamo infine come le ontologie realizzate, grazie al formalismo SKOS, sono già predisposte per eventuali traduzioni nelle altre lingue. Questo permetterebbe l'acquisizione di metadati e dataset derivanti dalla produzione scientifica di riferimento ai tre ambiti, potenziando la portata informativa e semantica del sistema nel suo complesso.

## Bibliografia

- [1] S. Gedzelman, M. Simonet, D. Bernhard, G. Diallo, and P. Palmer, "Building an Ontology of Cardio-Vascular Diseases for Concept-Based Information Retrieval", *Computers in Cardiology*, vol. 32, pp. 255–258, 2005.
- [2] Office for Official Publications of the European Communities, "Thesaurus EUROVOC" Annex to the index of the Official Journal of the EC, Luxembourg Publications of the European Communities, 1995.
- [3] R. Steinberger, M. Ebrahim, and M. Turchi, "JRC EuroVoc Indexer JEX - A freely available multi-label categorisation tool" *Proc. LREC'2012*, 21-27 May 2012 pp. 798-805
- [4] N. Ivezic, A. Farhad, K. Khosrow, and B. Kulvatunyou, "Ontological Conceptualization Based on the Simple Knowledge Organization System (SKOS)", *Journal of Computing and Information Science in Engineering*, doi:10.1115/1.4027582, May 2014.
- [5] M. Horridge, "A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools", The University Of Manchester Edition 1.3, March 2011
- [6] O. Bodenreider, "The Unified Medical Language System (UMLS):integrating biomedical terminology", *Nucleic Acids Res.* doi:10.1093/nar/gkh061, vol. 32, pp 267–270
- [7] W. D. J. Stuart, J. Nelson, and B.L. Humphreys, "Relationships in Medical Subject Headings (MeSH)." National Library of Medicine, Bethesda, MD, USA, 2002. Available from: <http://www.nlm.nih.gov/mesh/meshrels.html>
- [8] D. L. McGuinness , and F. van Harmelen, "OWL Web Ontology Language Overview". W3C Recommendation, February 2004. Available from <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- [9] World Health Organization on behalf of the European Observatory on Health Systems and Policies, "Health, health systems and economic crisis in Europe: impact and policy implications", 2013. Available from <http://www.euro.who.int>
- [10] Italian Public system of connectivity and cooperation, "Guidelines for semantic interoperability through linked open data", 2012 Agency for Digital Italy
- [11] M. Frustaci, "Glossary Economic-Statistical Multilingual". Italian National Institute for Statistics ISTAT, Document 8/2004, Available from <http://www.istat.it>
- [12] European Commission - Eurostat, "Eurostat: The Statistic Explained Glossary". Available from [http://epp.eurostat.ec.europa.eu/statistics\\_explained/index.php/Thematic\\_glossaries](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Thematic_glossaries)
- [13] European Commission - Eurostat, "RAMON Eurostat's Metadata Server". Available from <http://ec.europa.eu/eurostat/ramon/index.cfm>
- [14] European Commission - Eurostat, "Coded - The Eurostat concepts and definitions database". Available from [http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST\\_NOM\\_DTL\\_GLOSSARY&StrNom=CODED2&StrLanguageCode=EN](http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL_GLOSSARY&StrNom=CODED2&StrLanguageCode=EN)
- [15] United Nations Statistics Division, "United Nations Common Database – Methods and classifications". Available from <http://unstats.un.org/unsd/methods.htm>
- [16] The International Statistical Institute ISI, "The multilingual ISI glossary of statistical terms". Available from <http://isi.cbs.nl/glossary/>
- [17] M. Baglioni, S. Pieroni, F. Geraci, S. Molinaro, M. Pellegrini, E. Lastres "A New Framework for Distilling Higher Quality information from Health Data via Social Network Analysis". 13th International Conference on Data Mining (ICDMW.2013) IEEE , Dec 2013 DOI 10.1109/.142.
- [18] EuroVoc Multilingual Thesaurus of the European Union (<http://EuroVoc.europa.eu>)
- [19] EUROSTAT (2008), ESSPROS Manual 2008 edition. The European System of integrated Social PROtection Statistics (ESSPROS), Luxembourg
- [20] ISTAT-CNEL (2013), Rapporto Bes 2013: il benessere equo e sostenibile in Italia, Roma.
- [21] UNECE (2009), Classification of Statistical Activities (CSA rev. 2009), UNECE.