Smart Health 2.0

Progetto Pon04a2_C

MIUR D.D. 626/Ric e 703/Ric

D*10.4a* – Subset di terminologia specialistica inerente malattie rare e croniche invalidanti

D10.4b – Thesaurus malattie rare e malattie croniche invalidanti

investiamo nel vostro futuro

Informazioni sul documento

Document Name: SH2.0_D*10.4a*_D104b_IIT CNR_ v.*0.2*

Revision: v.*0.2*

Revision Date: ***15/12/2014***

Author: *Lista dei contributori: <Elena,Cardillo> (<IIT CNR>), <Maria Teresa,Chiaravalloti> (< ICAR CNR>), <Erika,Pasceri> (< IIT CNR>),*

Security: *Pubblico/Confidenziale*

Storia del documento

La seguente tabella ha il fine di tracciare gli aggiornamenti del documento (sicuramente vanno tracciate le seguenti: versione per CGTS, versione con commenti CGTS, versione finale…,). Può essere rimossa nella versione finale

Versione	Data	Modifica
0.1	15/11/2014	Versione per CGTS
0.2	15/12/2014	Versione revisionata per CGTS
1.0	15/12/2014	Versione finale

Sommario

ABSTRACT

Obiettivo della presente attività è stato quello di costruire specifici thesauri multilingua e multiregistro, ossia, contenenti termini, rispettivamente, provenienti da due o più lingue ed appartenenti a diversi livelli di espressività linguistica. Poiché realizzare un thesaurus che copra tutto lo scibile medico è dispendioso in termini di tempo oltre che difficile da coordinare in maniera organica, si è deciso di applicare la metodologia definita dalla norma ISO 25964-1:2011 Information and documentation – Thesauri and interoperability with other vocabularies ai sottodomini delle malattie rare e delle malattie croniche invalidanti. La norma fornisce raccomandazioni circa lo sviluppo ed il mantenimento di thesauri per finalità di information retrieval e ben si inquadra, pertanto, nell'ottica generale dell'OR che mira ad organizzare le conoscenze in ambito clinico a fini di supporto decisionale. Per la costruzione delle suddette risorse terminologiche si è proceduto, quindi, all'individuazione di eventuali risorse terminologiche di dominio esistenti, native italiane e non, così da avere un subset di termini iniziale da arricchire con termini estratti da corpora documentali costituiti a partire da letteratura scientifica di dominio.

INTRODUZIONE

Scopo del documento

L'obiettivo principale dell'attività è stato quello di analizzare la rappresentazione terminologica delle malattie rare e delle malattie croniche invalidanti per costruire un modello di classificazione a supporto non solo dei professionisti del settore, ma anche e soprattutto degli utenti/assistiti ai fini di utilizzo quale strumento per la ricerca di informazioni cliniche utili ai percorsi di diagnosi e cura.

Struttura del documento

Il documento è così strutturato:

La prima parte descrive le risorse di partenza

La seconda parte la costruzione dei corpora documentali

La terza parte descrive lo strumento di classificazione creato

Acronimi e termini chiave

Patologie rare, patologie croniche invalidanti, sistemi di classificazione, thesaurus

1. IL CONTESTO DI RIFERIMENTO

1.1. Le Malattie rare

Una malattia rara, anche conosciuta come “malattia orfana”, colpisce ogni anno una piccola percentuale della popolazione mondiale. Il numero delle malattie rare ad oggi riconosciute e diagnosticate varia tra le 7.000 e le 8.000, ma è una cifra che tende ad aumentare, con l’avanzare della ricerca scientifica e i progressi in ambito della ricerca genetica. La maggior parte delle patologie rare sono di origine genetica e, molto spesso, non esistono cure o sono, nella maggior parte di casi, cure in fase sperimentale senza risultati certi sulla loro reale efficacia. In molti casi tali patologie, si manifestano nei primissimi anni di vita e circa il trenta per cento dei bambini affetti, non arriva a compiere il quinto anno di età.
La definizione di malattia rara non è univoca, per cui una malattia può essere considerata rara in alcune aree geografiche e “frequente” (o comune) in altre. La diffusione di una patologia e quindi la frequenza con cui questa si manifesta sulla popolazione può essere vincolata da una serie di fattori quali, ad esempio, ambientali, genetici ecc. l’alternarsi di queste condizioni fa si che la definizione di rara possa essere vera ed accertata in alcuni casi, ma non in altri. Ad esempio, per diversi anni l’AIDS è stata una malattia estremamente rara, ma ad oggi il numero degli affetti di questa patologia cresce sempre di più, specialmente in alcune popolazioni, perdendo quindi la definizione/status di “rara” diventando, piuttosto, una patologia abbastanza diffusa. Un altro esempio è quello della lebbra, una malattia rara in Italia, ma comune in Africa centrale. La talassemia, un’anemia di origine genetica, è rara nel Nord Europa, ma è frequente nelle regioni del Mediterraneo; la “malattia periodica” è rara in Italia, ma è molto diffusa in Armenia. Allo stesso modo, esistono molte malattie abbastanza “comuni” che possiedono delle varianti estremamente rare.

Una patologia è considerata rara quando ha una prevalenza nella popolazione generale inferiore ad una data soglia, codificata dalla legislazione di ogni singolo paese: l’Unione europea definisce tale soglia allo 0,05% della popolazione, ossia 5 caso su 10.000 abitanti. Molte patologie, estremamente rare, arrivano ad avere una frequenza di un caso su 100.000 persone (0,001%).

Negli Stati Uniti, la legge sulle Malattie Rare del 2002 definisce “rara” una malattia o condizione che colpisca meno di 200.000 persone negli Stati Uniti", o circa 1 persona su 1.500. La definizione deriva dall’Orphan Drug Act del 1983, legge federale emanata per incoraggiare la ricerca sulle malattie rare e le possibili cure. Dieci anni dopo, nel 1993, è stato istituito anche l’NIH Office of Rare Diseases Research (ORDR), all'interno del National Institute of Health (NIH), l'agenzia federale responsabile della ricerca biomedica negli Stati Uniti e non solo.

Lo scopo dell’ORDR è quello di coordinare e sostenere la ricerca sulle malattie rare, cercando di fornire quante più possibili informazioni, ma soprattutto di promuovere la collaborazione e l'interoperabilità a livello internazionale.

Esistono diverse associazioni anche di stampo totalmente volontario che nascono dalla necessità di voler fare qualcosa di realmente concreto per chi è affetto da tali patologie, come ad esempio la National Organization for Rare Disorders. I leader delle organizzazioni di pazienti affetti da malattie rare avevano iniziato a rendersi conto che vi sono problemi comuni a tutte le persone affette da malattie rare e hanno, quindi, unito pe proprie forze per chiedere l’approvazione di una specifica normativa nazionale concretizzatasi nell’Orphan Drug Act [1].

Oggi, la NORD fornisce informazioni, sensibilizzazione, ricerca e servizi ai pazienti per aiutare tutti i pazienti e le famiglie colpite da malattie rare. Sono diverse le associazioni che si occupano di malattie rare soprattutto come punto di riferimento per i pazienti, che vogliano ottenere informazioni, mettersi in contatto con altre associazioni, o con specialisti del settore; tali associazioni, infatti, hanno sviluppato dei sistemi classificatori, più o meno dettagliati, o, in alcuni casi, semplici liste terminologiche con annesse descrizioni sulle malattie e relativi riferimenti verso centri specializzati di cure. Tali associazioni sono sia governative, ma anche private o semplicemente gestite da comunità spontanee formatesi dalle famiglie dei pazienti affetti da tali patologie.

Il comitato di esperti dell'Unione europea sulle malattie rare è stato istituito formalmente con la decisione della Commissione europea del 30 novembre 2009 (2009/872/EC). La missione principale del comitato è quella di aiutare la Commissione Europea per la preparazione e l'attuazione delle azioni comunitarie nel settore delle malattie rare, in collaborazione e con delle consultazioni con gli organismi specializzati negli Stati membri, le autorità europee competenti in materia di ricerca e di azione per la salute pubblica e altre parti interessate che operano nel settore.

Tra le principali associazioni di settore, esiste Orphanet, gestito da un consorzio europeo di cui fanno parte una quarantina di paesi, il cui coordinamento ha sede in Francia. I team nazionali hanno il compito di raccogliere informazioni sulle consulenze specialistiche, sui laboratori di diagnosi, sulle attività di ricerca in corso e sulle associazioni di pazienti nei rispettivi paesi. Tutti i team aderiscono alle norme di qualità di Orphanet. Il team coordinatore francese è responsabile della gestione del database e del sito web, del controllo di qualità, dell'elenco delle malattie rare, delle classificazioni e dell'edizione dell'enciclopedia di Orphanet.

Orphanet è gestito, inoltre, da diversi comitati che si occupano in modo indipendente della supervisione del progetto, per assicurarne la coerenza, lo sviluppo tecnologico e la continuità. è impegnato sia a livello europeo che internazionale in diverse attività. In particolare, a livello europeo esiste un direttivo costituito dai rappresentanti delle agenzie che finanziano la gestione del database e del sito web e il coordinamento europeo di Orphanet; un comitato di coordinamento costituito dai coordinatori nazionali di Orphanet, con a capo il direttore dell'unità Inserm-Orphanet; un comitato editoriale composto da oltre 100 esperti internazionali. A livello nazionale, esiste un comitato direttivo e/o scientifico a seconda del paese (uno per ognuno dei 38 paesi partecipanti), costituito da esperti nazionali con competenze che coprono tutte le aree mediche.

La gestione del database e del sito web e il coordinamento sono cofinanziati dall'Inserm (Institut National de la Santé et de la Recherche Médicale), dal Direttorato Generale della Sanità francese e dalla Commissione Europea, mentre alcuni servizi specifici sono finanziati da altri partner.
Le attività di Orphanet a livello nazionale sono allo stesso modo finanziate dalle istituzioni dei rispettivi paesi e/o da fondi specifici.
Orphanet si impegna a sostenere, aggiornare e sviluppare un database on-line, multilingue, totalmente libero e gratuito, dedicato alle malattie rare e ai farmaci orfani. La raccolta dei dati e la diffusione delle informazioni si attengono alle disposizioni legali in vigore nei vari Paesi impegnati nel progetto: codice etico professionale, legge sull'elaborazione dati e libertà, sui diritti di proprietà intellettuale e qualsiasi altra legge o regolamento applicabile. Il database è posto sotto la responsabilità di un comitato scientifico e di un comitato editoriale, i cui membri sono nominati sulla base della loro esperienza e competenza nel campo delle malattie rare, su proposta delle società scientifiche, delle autorità della salute dei Paesi impegnati nel progetto, o di qualsiasi altra associazione del settore. Ogni qualvolta vengono inserite nuove informazioni, un

membro del Comitato Scientifico ne valida la pertinenza e la correttezza. Di seguito in figura 1, la schermata di rappresentazione di un termine all'interno del database Orphanet.

FIGURA 1 RAPPRESENTAZIONE DI UN TERMINE ALL'INTERNO DEL DATABASE ORPHANET

Come evidenziato nella figura 1, ogni termine inserito all'interno del database di Orphanet è corredato da:

- numero ORPHA: identificativo all'interno del database;
- sinonimo/i: tutti i sinonimi del termine in oggetto;
- prevalenza: stima della prevalenza della singola patologia;
- trasmissione: come la malattia si trasmette e la frequenza con la quale si trasmette;
- età di esordio: quando la malattia si manifesta;
- i codici corrispondenti del termine all'interno delle principali classificazioni internazionali (nell'esempio: ICD10, OMIM, UMLS, MeSH, MedDRA, SNOMED CT);
- riassunto: descrizione dettagliata della patologia in oggetto;revisore/i esperto/i: identificativo dell'esperto che ha revisionato la malattia;
- ultimo aggiornamento: data ultima di revisione della malattia: le informazioni vengono aggiornate ogni qual volta i progressi scientifici lo richiedano, o almeno una volta l'anno per tutti i dati presenti nel database.

L'Office of Rare Diseases Research situato presso il National Center for Advancing Translational Sciences (NCATS) sostiene e coordina la ricerca sulle malattie rare. L'ORDR, inoltre, fornisce il proprio supporto a pazienti che sono affetti da una delle migliaia tra le malattie rare conosciute oggi. Per poter svolgere al meglio le sue attività, ORDR ha attivato una rete di collaborazione con vari istituti con il compito di coordinare e promuovere i rapporti tra i vari soggetti così come altri Istituti e Centri di NIH.

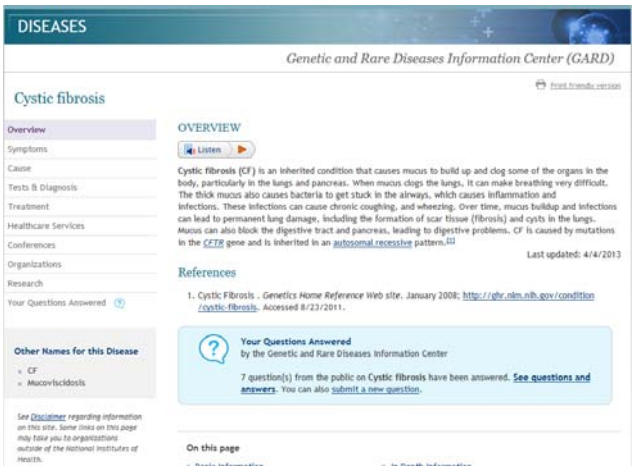

FIGURA 2 RAPPRESENTAZIONE DI UN TERMINE ALL'INTERNO DELL'ORDR

All'interno del sito dell'ORDR è possibile visualizzare, in ordine alfabetico e per ogni termine una descrizione dettagliata e i riferimenti bibliografici per ulteriori ricerche attraverso la rete della National Library of Medicine.

L'Organizzazione Nazionale per le Malattie Rare (NORD) è una federazione di organizzazioni sanitarie di volontariato dedicate ad aiutare le persone con rare malattie "orfane" e di assistenza alle organizzazioni che li servono. La NORD è impegnata nella identificazione, il trattamento e la cura delle malattie rare attraverso programmi di educazione, sensibilizzazione, ricerca e servizio.

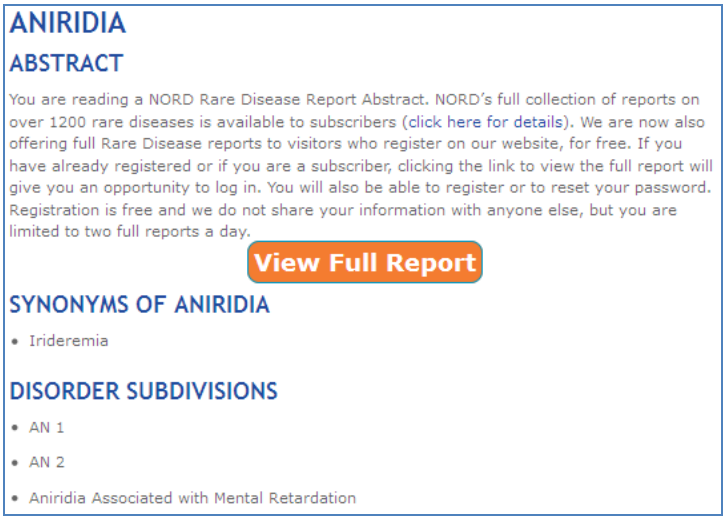

FIGURA 3 RAPPRESENTAZIONE DI UN TERMINE ALL'INTERNO DEL DATABASE DELLA NORD

Come mostra la figura 3, nel database NORD è possibile visualizzare, per ogni termine, una lista di sinonimi, laddove presenti, ed un abstract con i dettagli sulla patologia (funzione disponibile per gli utenti registrati, che possono farlo attraverso il sito senza dover pagare nulla.

L'analisi della situazione esistente nel panorama italiano si è rivelata una fase essenziale, sia per la corretta definizione dell'approccio metodologico da adottare, privilegiando la costruzione di un thesaurus che vuole essere un completamento delle attuali classificazioni esistenti in lingua italiana, partito dalla necessità di intervenire per un'unificazione a livello terminologico, sia per un'esigenza degli utenti finali della risorsa, i pazienti affetti da patologie rare, ma anche per un'esigenza emersa tra gli esperti di dominio coinvolti nelle diverse fasi dell'attività di ricerca. Nel già citato sistema di Orphanet esiste la versione italiana, anche se non per la totalità dei termini. Il coordinamento centrale in Italia avviene grazie al Centro Nazionale delle Malattie rare (CNMR) dell'Istituto Superiore di Sanità (ISS). Il centro opera anche a livello nazionale svolgendo attività di censimento, aggiornamento e ricerca nel campo delle malattie rare.

Tra i vari progetti coordinati dal CNMR, è importante citare EPIRARE (European Platform for Rare Registries), piattaforma europea per registri e database finalizzata alla costruzione di un set di dati che condivisibile tra i diversi Paesi. Il progetto EPIRARE, co-finanziato dall'Unione Europea e coordinato dal Centro Nazionale Malattie Rare dell'ISS. Oltre che con il Ministero della Salute e l'Istituto di Fisiologia Clinica del Consiglio Nazionale delle Ricerche, le attività sono svolte in associazione con altri istituti europei e extra-europei, tra i quali gli USA. Il Registro Nazionale delle Malattie Rare è stato istituito presso l'Istituto Superiore di Sanità con il Decreto Ministeriale 279/2001 (Art. 3) e ha avuto successive implementazioni mediante gli Accordi Stato-Regioni del 2002 e 2007. Il RNMR costituisce lo strumento principale di sorveglianza delle Malattie Rare su scala nazionale; l'obiettivo prioritario è la produzione di evidenze epidemiologiche a supporto sia della definizione e dell'aggiornamento dei livelli essenziali di assistenza (LEA), sia delle politiche e della programmazione nazionale.

Un importante passo in avanti in Italia è stato compiuto lo scorso 18 marzo 2014 quando la Camera dei Deputati ha approvato all'unanimità, in tema di malattie rare, una mozione con la quale il Governo si impegna a verificare lo status quo attuale del monitoraggio all'interno del Paese sulle malattie rare e come le stesse siano supportate dal Servizio Sanitario Nazionale. È stato, inoltre, richiesto il coordinamento a livello nazionale dei vari registri delle patologie di rilevante interesse sanitario in modo tale da avere un quadro chiaro sulla numerosità effettiva dei pazienti affetti da queste patologie. L'azione mira anche a introdurre misure a favore dei farmaci orfani provando a seguire il modello attualmente vigente negli USA: l'esenzione dei diritti da versare per l'immissione in commercio; una procedura di registrazione accelerata, un credito di imposta pari al 50% delle spese sostenute per la sperimentazione clinica.

1.2. Le Malattie croniche

A differenza di altri settori specialistici della medicina, quello relativo alle malattie croniche non è facilmente definibile in quanto rimangono molto labili i confini fra le patologie in esso incluse o da esso escluse. Sono, difatti, molteplici e diversi i fattori che vengono presi in considerazione per definire una patologia come "cronica". Diversi studi [2] hanno provato a circoscriverne i parametri di definizione, ma permane tuttora una parte di incertezza su ciò che possa essere considerato "cronico". Ciò a dispetto della loro crescente diffusione nel mondo, in pari misura nei Paesi "ricchi" e in quelli in via di sviluppo, e nella popolazione maschile come in quella femminile. L'Organizzazione Mondiale della Sanità (OMS) ha condotto nel corso degli ultimi anni diversi studi mirati ad analizzare il problema, individuarne le cause, tracciarne la diffusione, identificare soluzioni promuovendo campagne di prevenzione e comportamenti incentrati su uno stile di vita corretto.

FIGURA 4. CAUSE DELLE MALATTIE CRONICHE INDIVIDUATE DALL'ORGANIZZAZIONE MONDIALE DELLA SANITÀ

Principale motivo della crescente attenzione rivolta alle malattie croniche sono gli effetti invalidanti che esse causano e che possono manifestarsi per una parte considerevole della vita delle persone che ne vengono colpite. Questo causa una degenerazione progressiva delle condizioni di salute della popolazione ed impone l'urgenza di pensare soluzioni idonee a garantire la qualità della vita e dello stato di salute dei malati cronici. A differenza di quanto comunemente ritenuto, le malattie croniche non sono un problema degli anziani in quanto quasi la metà dei decessi ad esse dovuti avviene prima dei 70 anni ed un quarto entro i 60 anni. Sono, inoltre, sempre più frequenti i casi di patologie croniche diagnosticate in età infantile, primo fra tutti il diabete. A questo proposito, nell'ultimo rapporto sulle Non Communicable Diseases (NCDs) [3], l'OMS sottolinea che grande attenzione deve essere posta all'insorgenza precoce di queste malattie perché porterà nel tempo ad avere adulti che si sono ammalati prima e che potranno quindi avere manifestazioni attualmente non conosciute della malattia stessa, perché non ancora arrivata ad un corso di vita così lungo.
Considerate queste premesse, lo strumento terminologico che si vuole realizzare con la presente attività rappresenta una risorsa flessibile, grazie agli aspetti di multilinguismo e multiregistro, e suscettibile di essere usata in molteplici contesti e per molteplici finalità, garantendo un accesso più rapido e agevole all'informazione specialistica di settore.

2. LA METODOLOGIA

Le differenti caratteristiche delle due classi di patologie oggetto di studio hanno visto una leggera differenziazione durante la fase di analisi e di recupero delle risorse disponibili al fine di creare i corpora e le basi di conoscenza utili per la costruzione dei due thesauri. Per tale motivo la fase preliminare della metodologia, ovvero la creazione del corpus, con conseguente analisi delle risorse di riferimento per il dominio e l'estrazione dei termini rilevanti, è stata riportata di seguito in due paragrafi distinti, uno per il thesaurus delle malattie rare e l'altro per il thesarus delle malattie croniche. E' comune al contrario la metodologia utilizzata per la costruzione dei due thesauri. In entrambi i casi si è scelto infatti di utilizzare come supporto alla costruzione del thesaurus il software Multites Pro 2007[1], che risponde ad un insieme di requisiti fondamentali per garantire una corretta gestione di tale risorsa. Tra le caratteristiche più rilevanti del software si possono annoverare:

- Disponibilità delle relazioni standard previste dalla normativa in materia di vocabolari controllati (BT, NT, RT, …) e possibilità di aggiungerne delle altre, definendone la natura: nel caso specifico,

[1] http://www.multites.com/index.htm

ad esempio, la relazione gerarchica è stata ulteriormente specializzata attraverso l'aggiunta di BTG, BTP e delle relazioni inverse;

- Possibilità di definire delle categorie all'interno delle quali collocare i termini inseriti nel thesaurus: le faccette (al livello più generico) corrispondono alle categorie, per ciascuna delle quali è possibile visualizzare i termini che le appartengono;
- Possibilità di ricorrere alla poligerarchia, e quindi di attribuire ad uno stesso termine più di un concetto sovraordinato;
- Inserimento automatico delle relazioni inverse e controllo della coerenza delle relazioni inserite (es. non è possibile associare delle relazioni gerarchiche o associative ai termini non preferiti, ad esclusione della relazione Subject Category – SC che indica la faccetta di appartenenza);
- Nessun limite sulla quantità di termini, etichette di snodo, livelli gerarchici e relazioni di vario tipo che possono essere inserite;
- Possibilità di creare thesauri multilingue, e quindi di definire delle traduzioni parallele in altre lingue per i termini del thesaurus;
- Possibilità di definire e di visualizzare alcune caratteristiche per ciascun termine inserito, tra cui: lo status, per cui un termine può essere *accepted*, *candidate*, *provisional* o *not valid*; il tipo, che permette di stabilire se un termine è un descrittore o un'etichetta di snodo; in quante e che tipo di relazioni è coinvolto il termine; la presenza di eventuali note d'ambito e infine la categoria alla quale il termine appartiene;
- Possibilità di visualizzare il thesaurus tramite 3 diverse modalità di presentazione: alfabetica, gerarchica, e sistematica.

2.1. Metodologia per la creazione del corpus delle malattie rare

Lo Unified Medical Language System (UMLS) è stato lo strumento utilizzato per l'analisi della rappresentazione delle malattie rare all'interno dei principali sistemi di classificazione esistenti a livello internazionale, ed in particolare: il Medical Subject Headings (MeSH), la terminologia di riferimento per l'indicizzazione e la ricerca nella letteratura biomedica; l'Online Mendelian Inheritance in Man (OMIM), base di conoscenza di riferimento sulle malattie genetiche; la Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT), la più vasta terminologia clinica del mondo, e la International Classification of Diseases 10th revision (ICD-10), la classificazione delle malattie, dei traumatismi, degli interventi chirurgici e delle procedure diagnostiche e terapeutiche. Altri studi simili sono stati condotti, si veda ad esempio il lavoro di Milicic et al.[4], che ha portato allo sviluppo di una metodologia per l'identificazione automatica di corrispondenze da Orphanet verso l'UMLS sulla base di tecniche sofisticate per il confronto dei termini il più possibile automatico, riducendo così al minimo il confronto manuale. Questo tipo approccio è prevalentemente dedicato alle risorse presenti nel database di Orphanet ottimizzando in modo specifico il processo di allineamento per questa risorsa poiché Orphanet fornisce, per la maggior parte dei termini in esso contenuti, il riferimento verso i codici di OMIM e quelli della ICD, individuati manualmente dagli esperti del dominio. L'analisi condotta da Merabti et al.,[5] inoltre, ha portato all'implementazione di una metodologia basata sull'elaborazione del linguaggio naturale per la mappatura della terminologia di Orphanet verso il MeSH. I risultati ottenuti con questa metodologia sono stati infine confrontati con quelli ottenuti con la corrispondenza esatta basata su stringhe verso UMLS e ICD utilizzando i riferimenti preesistenti verso gli identificativi dei singoli termini, evidenziando come, con la nuova metodologia implementata, possano essere individuati dei nuovi termini che possano essere aggiunti al database. In questo studio, invece, per poter avere una rappresentazione più ampia della terminologia di settore, sono state prese in considerazione

più risorse terminologiche specifiche per poi mapparle verso i principali sistemi di classificazione in ambito medico, quali il Medical Subject Headings, la terminologia di riferimento per l'indicizzazione e la ricerca nella letteratura biomedica; l'Online Mendelian Inheritance in Man, base di conoscenza di riferimento sulle malattie genetiche; la Systematized Nomenclature of Medicine-Clinical Terms, la più vasta terminologia clinica del mondo, e la International Classification of Diseases 10th revision, la classificazione delle malattie, dei traumatismi, degli interventi chirurgici e delle procedure diagnostiche e terapeutiche. L'analisi terminologica si è svolta a partire da un corpus terminologico costruito dalle basi di dati provenienti dal database di Orphanet, dell'ORDR e NORD. Il corpus è stato mappato attraverso l'UMLS, applicando determinati filtri, per effettuare una prima scrematura in modo tale da comprendere quali terminologie fossero maggiormente rappresentative del dominio oggetto di studio.

È importante evidenziare che l'analisi terminologica è stata condotta sulla terminologia in lingua inglese, poiché non tutti i sistemi di classificazione analizzati hanno la corrispondente traduzione in inglese. La traduzione dei termini per i quali mancava il corrispondente in italiano è stata fatta, quindi, una volta definito il corpus.

2.1.1. ACQUISIZIONE DELLE FONTI

Ottenute le necessarie autorizzazioni per poter acquisire le fonti terminologiche del dominio specifico delle malattie rare per quanto riguarda i dati provenienti dal database di Orphanet, dall'ORDR e dalla NORD, si è dato avvio all'estrazione dei termini e alla creazione del corpus.

- ***Estrazione dei termini di Orphanet***

ID Orphanet	Tipologia termine	Descrizione termine
ORPHA000881	PT	Turner syndrome
ORPHA000095	PT	Friedreich ataxia
ORPHA000586	PT	Cystic fibrosis
ORPHA000846	PT	Alpha-thalassemia
ORPHA000848	SYN	Cooley anemia
ORPHA000848	PT	Beta-thalassemia
ORPHA000262	SYN	Duchenne and Becker dystrophinopathy
ORPHA000262	PT	Duchenne and Becker muscular dystrophy
ORPHA000261	SYN	EDMD
ORPHA000261	SYN	Emerinopathy

TABELLA 1 ESTRATTO DEI TERMINI PROVENIENTI DAL DATABASE DI ORPHANET

Come si evince dalla tabella 1, nella prima colonna sono riportati gli ID dei concetti all'interno del database, nella seconda colonna la tipologia del termine che rappresenta ciascun concetto: PT sta ad

indicare il termine principale (7,715 in totale) e SYN il relativo sinonimo (5,224 in totale), i termini che si riferiscono allo stesso concetto, ovvero alla medesima patologia, sono associati allo stesso identificativo.

Un'altra caratteristica molto importante dei termini di Orphanet è che la maggior parte di essi hanno un collegamento con i corrispondenti all'interno dell'ICD-10 e OMIM, ciò è stato di fondamentale importanza per un maggiore controllo in termini di quality assurance in fase di valutazione dei risultati ottenuti durante la fase di mapping con l'UMLS.

- ***Estrazione dei termini dell'ORDR***

I termini dell'ORDR sono stati acquisiti direttamente attraverso il sito web, mantenendo la loro composizione originaria, quella di non essere legati da alcun tipo di relazione, fatta eccezione per quella di sinonimia.
La lista di termini, aggiornata periodicamente dall'ORDR, sul piano semantico è rappresentata da un totale di 6,857 termini principali e 11,803 sinonimi. Per ognuno di essi, tramite il sito web, vengono fornite informazioni dettagliate sulle risorse informative collegate, liberamente accessibili a tutti coloro che vogliano avere quanti più dettagli su ogni singola patologia.
I termini inclusi nella lista rappresentano le malattie rare per cui è stata avanzata specifica richiesta da parte di gruppi di pazienti per ottenere maggiori informazioni presso l'ORDR ed, in particolare, presso il Genetic and Rare Diseases Information Center (GARD), centro finanziato dal ORDR, e presso il National Human Genome Research Institute (NHGRI), oppure sono termini, di diversa provenienza, che rappresentano patologie che nel corso degli ultimi 10 anni sono state definite rare dalla comunità scientifica.

- ***Estrazione dei termini dell'ORDR***

La stessa metodologia applicata ai termini della NORD, è stata replicata per i termini provenienti dalla NORD. Essi, seppur in numero esiguo rispetto alle prime due risorse, sono stati comunque analizzati poiché il campione dei termini rappresenta la terminologia "più vicina" a quella dell'utente finale, il paziente o familiare del paziente affetto da patologia rara. La copertura semantica in questo caso ha registrato un totale di 1,236 termini principale e 4,562 sinonimi.

2.1.2. LE TERMINOLOGIE TARGET: OMIM, SNOMED CT, MESH, ICD10

In questo paragrafo si riporta descrizione dettagliata delle risorse utilizzate per l'analisi della mappatura della terminologia specifica del dominio delle patologie rare. Le terminologie su cui è stata focalizzata l'attenzione sono state Online Mendelian Inheritance in Man (OMIM), SNOMED-Clinical Terms (SNOMED CT), Medical Subject Headings (MeSH), e l'International Classification of Diseases 10th revision (ICD10),

- ***Online Mendelian Inheritance in Man (OMIM)***

La Online Mendelian Inheritance in Man è la terminologia sui geni e fenotipi umani più completa e aggiornata con descrizioni dettagliate sulle patologie genetiche mendeliane e su oltre 12,000 geni. Uno dei punti di forza di OMIM, in particolare, è che evidenzia le relazioni tra fenotipi e genotipi. Il vantaggio di questa terminologia è che viene quotidianamente aggiornata e ciascun record contiene dei riferimenti verso altre risorse genetiche. Il lavoro di costruzione del database iniziò agli inizi del 1960 grazie al dott. Victor A. McKusick, che cominciò a costruire il catalogo delle malattie mendeliane e lo

intitolò Mendelian Inheritance in Man (MIM), che vide nel tempo dodici riedizioni tra il 1966 e il 1998. La versione che diede vita a quella attualmente in uso, fu creata nel 1985 grazie ad una collaborazione, tuttora esistente, con la National Library of Medicine e William H. Welch della Welch Medical Library presso la John Hopkins University School of Medicine, sotto la direzione della dott.ssa Ada Hamosh. La nuova ottica di lavoro fu orientata alla realizzazione di una versione che potesse essere pubblicata sulle prime reti internet, cosa che venne attuata poi nel 1987. OMIM, all'atto della creazione, fu pensato in particolare per l'utilizzo da parte dei medici di base, ma anche per favorire la ricerca genetica, quindi per specialisti del settore, ricercatori, e giovani studenti. Il database oggi è liberamente consultabile da parte del pubblico, senza restrizioni, tuttavia non sempre le informazioni pubblicate sono di immediata comprensione se non da parte del personale estremamente qualificato; non è adatto quindi a fornire risposte immediate a cittadini/pazienti che necessitano informazioni su particolare condizioni medico-genetiche.

- ***Systematized Nomenclature of Medicine (SNOMED-CT)***

  La Systematized Nomenclature of Medicine (SNOMED-CT) è la più vasta terminologia al mondo in ambito clinico creata dalla International Health Terminology Standard Development Organization (IHTSDO) per l'utilizzo all'interno delle cartelle cliniche elettroniche. Questa terminologia copre la maggior parte delle aree di informazione clinica come: patologie, sintomi, procedure diagnostiche, ecc. SNOMED-CT è uno strumento utile per l'indicizzazione e recupero di informazioni, ma anche per l'aggregazione di dati clinici, per ridurre il rischio di errore nei casi di differenti sistemi di codifica utilizzati nel clinical care nella ricerca e per la salute dei pazienti. SNOMED-CT è fondamentale nell'interoperabilità e nello scambio dei dati clinici, l'elevato livello di dettaglio al suo interno contribuisce al supporto decisionale, percorsi di cura, rilevazioni statistiche ecc. Esistono, inoltre, numerose traduzioni che contribuiscono ad aumentare l'internazionalità della classificazione. Le traduzioni in diverse lingue contribuiscono altresì ad aumentare i termini che vengono utilizzati a livello locale nelle singole nazioni. Da un punto di vista quantitativo SNOMED CT contiene più di 311,000 concetti, associati ad un identificativo univoco (ad esempio il concetto 22298006 si riferisce solo ed esclusivamente a Myocardial infarction). Tutti i concetti all'interno della classificazione sono organizzati grazie ad una classificazione gerarchica con una relazione di tipo IS-A, ad esempio:

  Viral pneumonia IS-A Infectious pneumonia
  Infectious pneumonia IS-A Pneumonia
  Pneumonia IS-A Lung disease.

  I concetti possono avere più legami gerarchici all'interno della classificazione, ad esempio, *Infectious pneumonia* è anche un termine più specifico di *Infectious disease*.

  I concetti all'interno di SNOMED CT sono collegati tra di loro con più di 1,360,000 relazioni, mentre ogni singolo concetto contiene dettagli grazie a diversi termini o descrizioni ad esso collegato che si dividono in Fully Specified Names (FSNs), Preferred Terms (PTs), and Synonyms. Ciascun concetto ha uno ed uno solo FSN, che è unico all'interno di SNOMED CT, come anche per il PT, stabilito da un comitato di esperti come il termine più comunemente utilizzato dalla comunità scientifica che allo stesso tempo è anche il più rappresentativo.

  È possibile, per ciascun termine, avere da zero a più sinonimi, questi ultimi non devono per forza avere la prerogativa di essere unici, poiché sono dei rimandi che riportano al concetto principale.

SNOMED-CT è, inoltre, un thesaurus multilingue con alla base un'ontologia; grazie a questo sistema è possibile gestire la sinonimia di termini provenienti anche da diverse lingue, ad esempio:

"Acute coryza", "Acute nasal catarrh",
"Acute rhinitis", "Common cold"

e le relative versioni in spagnolo:

"resfrío común", "rinitis infecciosa"
hanno il medesimo ID 82272006.

Le primissime versioni di SNOMED avevano una struttura a faccette ordinata sulla base di assi semantici che richiedevano un'organizzazione dei concetti molto più complessa e un differente sistema di codifica. Questo sistema, oltre ad essere più costoso in termini di gestione globale dell'intero sistema, non era abbastanza user-friendly così fu deciso di optare per il corrente sistema classificatorio per non venire meno agli obiettivi base per cui SNOMED-CT è stato creato. SNOMED CT, ad oggi, è gestito e distribuito per il suo utilizzo dall'International non-profit Standards Development Organization, con seede a Copenhagen, Danimarca. L'utilizzo di tale sistema è vincolato dall'acquisto di una licenza. La versione completa di SNOMED CT è disponibile all'interno dell'Unified Medical Language System per gli utenti che hanno sottoscritto l'accordo.

- ***International Classification of Diseases 10th revision ICD-10***

L'ICD-10 è la decima revisione della ICD, ossia la classificazione internazionale delle malattie e dei problemi correlati, proposta dall'Organizzazione Mondiale della Sanità in cui sono classificate ben oltre duemila malattie. La traduzione ufficiale in lingua italiana, è stata effettuata a cura dell'ISTAT e dell'Ufficio di Statistica del Ministero della Salute, ed è stata pubblicata dall'Organizzazione mondiale della sanità a Ginevra nel 2000 e a Roma nel 2001. In Italia la Direzione Centrale Salute Integrazione Sociosanitaria e Politiche Sociali della Regione Friuli Venezia Giulia è riconosciuta – dal gennaio 2010 - come Centro Collaboratore italiano dell'Organizzazione Mondiale della Sanità per la Famiglia delle Classificazioni Internazionali. Il Centro, riconosciuto anche dal Ministero della Salute, si occupa dello sviluppo, implementazione, utilizzo e diffusione delle classificazioni internazionali, in primis ICD e ICF.

- ***Medical Subject Headings***

Il MeSH è un thesaurus costituito da oltre 24.000 descrittori, prodotto e gestito dalla National Library of Medicine (NLM), utilizzato nell'indicizzazione degli articoli delle riviste biomediche di PubMed/Medline ed altri archivi della NLM.
Esso ha una struttura ad albero e i termini accolti al suo interno sono collegati tra loro attraverso un sistema di relazioni semantiche:

- Relazione di equivalenza: questa tipologia di relazione rimanda da una voce non accolta ad una che è considerata, all'interno del dominio specifico di appartenenza, quella più diffusa o maggiormente utilizzata:

Clinical Markers see Biological Markers;

- Relazioni Gerarchiche: questa tipologia di relazione rinvia da un termine gerarchicamente superiore ad uno inferiore e viceversa:

  Apparato digerente
  - *Vie biliari*
  - *Fegato*

- Relazioni Associative: questa tipologia di relazione rimanda da un termine ad un altro correlato:

  *Veins **see** related Phlebography;*
  *Neoplasms **see** related Antineoplastic agents;*

- Altri tipi di relazione: la voce *consider also terms* collega termini linguisticamente correlati, aventi radici latine o greche o anglosassoni, questa tipologia di relazione è usata principalmente per i termini anatomici, ad esempio:

  Liver consider also terms at hepat-

L'aggiornamento dei descrittori all'interno del MeSH viene effettuato con cadenza semestrale, sempre dalla NLM, sulla base dell'evoluzione delle scienze biomediche: gli esperti e specialisti di ogni area della biomedicina effettuano un monitoraggio costante della letteratura scientifica e tengono sotto osservazione le aree emergenti della ricerca. Il thesaurus MeSH è composto da:

- Oltre venticinquemila descrittori principali (main headings);
- Ottantatre sottodescrittori (subheading o qualifiers);
- Oltre centomila voci supplementari (Supplementary Concept Records), che comprendono nomi di sostanze chimiche, numeri di registro CAS etc.

Dal 1999 la NLM ha dato la possibilità di consultare liberamente il Thesaurus MeSH nella sua versione originale senza alcun costo di licenza attraverso le seguenti modalità di ricerca: MeSH Browser che consente la ricerca all'interno del vocabolario attraverso vari punti d'accesso; MeSH Database per la selezione dei descrittori da utilizzare per la ricerca in MEDLINE/PubMED; il Metathesaurus UMLS, in cui convergono, oltre che il MeSH, diversi vocabolari del dominio biomedico.

2.1.3. IL MAPPING

Nonostante diversi vocabolari, terminologie e sistemi di classificazione siano chiaramente necessari, presentano diverse difficoltà soprattutto nella loro implementazione. È possibile, tuttavia, ovviare a questi problemi attraverso il mapping, che significa collegare i contenuti da una terminologia o schema di classificazione ad un altro. Operazione fondamentale del mapping, durante la fase di analisi dei dati, è quella di decidere le regole da applicare, ovvero i parametri da utilizzarsi, che influiranno fortemente sui risultati ottenuti. Il mapping considera scopi differenti, livelli di dettagli e linee guida della codifica source (di partenza) e della codifica target (di arrivo). Il processo utilizza un metodo standard in cui il concetto della terminologia o la descrizione della classificazione sono interpretate fra sistemi. L'analisi dei dati ha previsto due macrofasi: una prima fase è stata dedicata a capire quanto le risorse individuate si sovrapponessero e se fossero più o meno omogenee dal punto di vista della selezione del termine principale. C'è da sottolineare,

tuttavia, che nonostante lo scopo finale sia il medesimo per tutte le associazioni da cui sono state acquisite le basi terminologiche, ovvero fornire informazioni dettagliate agli utenti coinvolti nel mondo delle patologie rare, siano essi professionisti del settore o pazienti, la formazione dei rispettivi database e la conseguente pubblicazione sono avvenute in modalità differenti. È per questo motivo, principalmente, che sussistono delle disomogeneità soprattutto a livello quantitativo.
Tuttavia, anche se la base terminologica della NORD è esigua in termini quantitativi si è rivelata importante nei casi in cui non si avevano informazioni sulle relazioni di sinonimia per termini presenti sia in ORDR che in Orphanet, ma in NORD erano presenti, come mostrato in Fig.5:

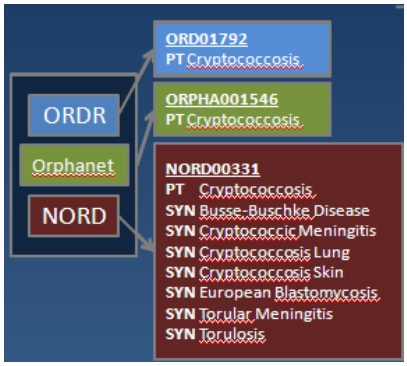

FIGURA 5 RAPPRESENTAZIONE DELLE PATOLOGIE RARE NELLE RISORSE DI PARTENZA

Una volta creata la base terminologica di tutti i termini provenienti dalle tre risorse fin qui esposte, si è proceduto ad effettuare una mappatura verso i termini presenti all'interno dell'UMLS. I termini facenti parte del corpus sono stati utilizzati come discriminante all'interno di tutti i vocabolari e classificazioni esistenti all'interno del database dell'UMLS, una volta effettuata la mappatura, è stato possibile ottenere una matrice di corrispondenza che ha evidenziato quali e quanti termini siano presenti in ogni singola risorsa terminologica [6]. Prima di effettuare la mappatura verso l'UMLS, tuttavia, è stato necessario applicare degli appositi filtri per evitare che ci fosse troppo rumore informativo. All'interno dell'UMLS vi sono numerosi vocabolari altamente specifici, che non potevano, per la loro intrinseca natura, essere sicuramente utili per lo scopo del presente lavoro. Sono stati fissati, quindi, i seguenti obiettivi:

- Sovrapposizione delle terminologie di partenza (per comprendere, in termini quantitativi e qualitativi, quanto le tre risorse di partenza avessero in comune;
- Presenza/non presenza dei termini all'interno dell'UMLS dei termini provenienti dalle risorse di partenza;
- Una volta ritrovato un termine all'interno delle risorse terminologiche, acquisire eventuali termini aggiuntivi che possano alimentare il nuovo thesaurus da creare

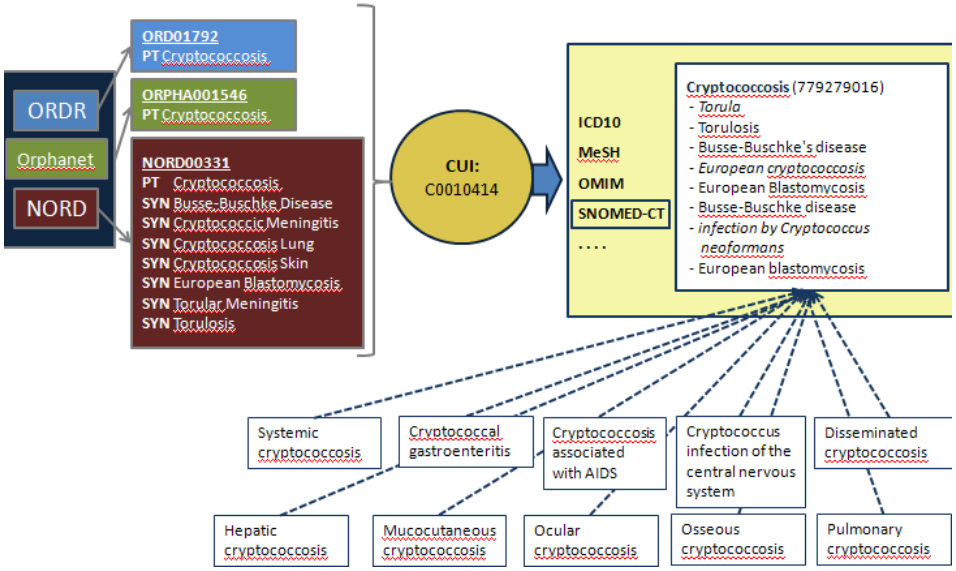

FIGURA 6 MAPPATURA DELLE RISORSE DI PARTENZA ATTRAVERSO L'UMLS

La Fig. 6 mostra l'ipotesi di lavoro appena enunciata. Ad esempio, partendo dal termine Cryptococcosis , si evince che il termine è presente in tutte e tre le risorse di partenza, ma in varia misura: in Orphanet e in ORDR è presente il termine singolo, mentre in NORD, sono presenti diversi sinonimi. Una volta individuato il termine è stata effettuata la mappatura verso l'UMLS, rintracciando, quindi, il termine all'interno di diverse risorse terminologiche. Il risultato ottenuto è che il termine Cryptococcosis è presente in tutte e tre le terminologie, ma è da rilevare come SNOMED CT fornisca maggiori informazioni rispetto a quelle di partenza. Questo esempio rispetto al singolo termine sta a sottolineare l'importanza del lavoro svolto, ovvero, come, si possano ottenere delle risorse terminologiche arricchite che portino, quindi, svolgere funzioni ad alto livello di dettaglio. Poiché i termini di ORDR, Orphanet e NORD si riferiscono a disturbi o patologie (rari), è stato ristretto il campo di ricerca all'interno dell'UMLS (questo perché l'UMLS comprende nomi di farmaci, analisi di laboratorio ecc.) per effettuare la mappatura solo verso uno specifico gruppo di termini. Il filtro semantico applicato fornisce un discreto livello di disambiguazione tra i termini: quando un termine appartenente alle risorse originarie (ORDR e Orphanet), come NF2, viene mappato verso l'UMLS, può trovare corrispondenza in una patologia (neurofibromatosi di tipo 2) oppure in un gene (NF2, cromosoma 22, di cui le mutazioni causano la neurofibromatosi di tipo 2). Applicando il filtro semantico relativo ai Semantic Disorders, si evita cosi di avere come risultato dei termini non di interesse per il presente studio.

Tuttavia, pur applicando tale filtro potrebbe, comunque, persistere un certo grado di ambiguità, quando ad esempio un concetto delle risorse originarie (ORDR, Orphanet e NORD), viene mappato verso più concetti delle risorse presenti nell'UMLS. Un'attenzione particolare è stata rivolta anche agli acronimi presenti in numero non trascurabile all'interno delle classificazioni. Per ovviare a problemi di ambiguità semantica, sono state applicate delle regole di disambiguazione assegnando un "peso" ai termini in forma estesa ed un altro, di misura inferiore, agli acronimi.

L'assegnazione dei valori è stata effettuata con delle regole grammaticali grazie alle quali è stato assegnato valore "0,33" alle parole che avessero almeno una o più lettere tutte maiuscole, considerate, quindi, acronimi; mentre alle parole più lunghe, con carattere minuscolo, è stato assegnato valore pieno, ovvero "1".

Id Orphanet	Tipologa termine	CUI UMLS	Peso
ORPH001	T1	CUI1	1

ORPH001	SYN1	CUI2	1
ORPH001	SYN2 [ACRO]	CUI2	0,33
ORPH001	SYN3	CUI1	1

Il codice ORPH001 è l'identificativo assegnato al termine proveniente da Orphanet, nella seconda colonna è rappresentata la tipologia del termine analizzato:

- T1 : termine principale
- SYN 1 ,3 = sinonimi del termine principale
- SYN 2 [ACRO] = sinonimo del termine principale sotto forma di acronimo

Nell'ultima colonna, è stato inserito il peso associato ad ogni termine. Dalla tabella si può notare come, a partire dallo stesso termine, si ha come risultato l'associazione a due diversi CUI all'interno dell'UMLS, e ciò sembra essere abbastanza improbabile.

La formula che segue è stata applicata come regola di disambiguazione per i risultati ottenuti, la prima volta è stata applicata al CUI1 e la seconda volta al CUI2:

$$WCUI1 = \frac{\sum wCUI1}{\sum wCUI1 + \sum wCUI2} = \frac{2}{2 + 1{,}33} = \frac{2}{3{,}33} = 0{,}60$$

La formula chiarisce come è stato calcolato il peso (W = Weight) di ogni CUI individuato. In questo caso il risultato è di 0,60, quindi, poiché il valore soglia è stato fissato a 0.6, il peso ottenuto stabilisce che in questo caso il CUI1 identificato non genera ambiguità nei risultati ottenuti.

$$WCUI2 = \frac{\sum wCUI2}{\sum wCUI2 + \sum wCUI1} = \frac{1{,}33}{1{,}33 + 2} = \frac{1{,}33}{3{,}33} = 0{,}39$$

Nel secondo caso il valore ottenuto è di 0.39, quindi è considerato ambiguo ed è stato analizzato attraverso revisione manuale.

2.1.4. RISULTATI DELLA MAPPATURA VERSO L'UMLS

Effettuata la mappatura si sono ottenuti risultati di tre tipologie:

1. Nessun CUI associato: nessun termine è stato individuato all'interno dell'UMLS;
2. Un solo CUI: la corrispondenza è stata rintracciata in modo esatto.
3. Più di un CUI: è stata individuata più di una corrispondenza a partire da un solo termine, questi casi sono stati analizzati manualmente per individuare eventuali problemi nella query di partenza o semplicemente per studiarne la casistica.

La risorsa terminologica all'interno dell'UMLS che meglio rappresenta l'ambito delle malattie rare è il MeSH, nella versione del 2014, con il 77%, seguito da SNOMED CT con il 600% e OMIM con il 55%. L'ICD ne contiene solo il 16%.

Per quanto riguarda i concetti all'interno della NORD solo il 4% non hanno trovato alcuna corrispondenza, il 25% per ciò che concerne l'ORDR e il 56% per Orphanet.
I casi di ambiguità sono stati revisionati manualmente per individuare le cause della mancata mappatura: è stato riscontrato che i termini che non hanno trovato corrispondenza all'interno dell'UMLS, si riferiscono a patologie estremamente rare e, quindi, non inserite in nessuna classificazione di tipo specialistico, come il caso della *Richieri Costa-Guion Almeida-Cohem syndrome* e la *Lateral body wall complex*.
Nel caso della Levy-Yeboa Syndrome non è stata rintracciata la corrispondenza all'interno dell'UMLS poiché è una sindrome di recente scoperta e quindi non presente all'interno del database dell'UMLS in base agli ultimi aggiornamenti.

2.2. Metodologia per la creazione del corpus delle malattie croniche

Per il corpus relativo alle patologie croniche invalidanti si è proceduto in modo differente rispetto a quanto descritto precedentemente sulle patologie rare. I termini che hanno costituito la base di conoscenza per la costruzione del thesaurus sono stati estratti dalla letteratura scientifica di settore. La costituzione del corpus specialistico di letteratura specialistica di dominio è stata realizzata utilizzando la funzione di *Advanced Search* offerta da Pubmed, il più autorevole database bibliografico di letteratura scientifica biomedica. La ricerca è stata focalizzata sulle tre aree diagnostiche sopramenzionate: patologie tiroidee, patologie metaboliche e ipertensive (pur ricomprendendo fra queste ultime anche le patologie cardiovascolari vere e proprie in quanto diretta conseguenza della condizione ipertensiva).
Sono, pertanto, stati impostati i seguenti criteri di ricerca:

- *Date publication*: "2010/01/01" to "present"
- *Language*: "Italian"/"English"
- *MeSH Terms*: "Diabetes"/"metabolic disorders"/"hypertension"/"endocrine disorders"

Le barre verticali indicano che le stringhe di ricerca sono state usate in maniera alternativa. Nello specifico, si è deciso di focalizzarsi sulle pubblicazioni relative agli ultimi cinque anni e sono state ricercate pubblicazioni scientifiche native in lingua italiana e pubblicazioni scientifiche in inglese (molto più numerose per ovvi motivi). Talvolta è capitato che uno stesso articolo fosse contestualmente disponibile in entrambe le lingue. Mentre per la lingua italiana sono state recuperate tutte le risorse che è stato possibile individuare, proprio perché in numero esiguo, per la lingua inglese è stato invece necessario fissare il numero massimo di pubblicazioni da recuperare. Tale limite è stato stabilito tenendo conto del numero di pubblicazioni in italiano disponibili di modo da avere corpora bilanciati. Per quanto riguarda, invece, l'oggetto della pubblicazione, si è deciso di rintracciare le patologie di interesse utilizzando come chiave di ricerca le voci MeSH, thesaurus specializzato per la letteratura scientifica di ambito biomedico. I termini sono stati usati singolarmente nella ricerca, in quanto precedentemente verificato che se usati in combinazione (onde verificare la co-presenza di una o più patologie di interesse) davano scarsi risultati.
Fra le risorse proposte, impostando di volta in volta i parametri citati, sono state prese in considerazione solamente le pubblicazioni per le quali era disponibile gratuitamente il full text o perlomeno l'abstract. Ciò ha portato allo costituzione di un corpus in lingua italiana contenente 95 abstract e 39 full paper e di un corpus in lingua inglese contente 100 abstract e 50 full paper.
È interessante rilevare una peculiarità emersa nel corso di questa fase del lavoro. Alcune pubblicazioni relative alle patologie metaboliche sono state recuperate usando come chiave di ricerca *MeSH Terms*: "endocrine disorders" e non invece usando "diabetes", come ci si sarebbe aspettato, essendo per di più

termine presente nel titolo. Ciò rivela una indicizzazione non sempre accurata delle risorse ed una necessità di combinare più chiavi di ricerca specifiche per evitare un retrieval eccessivamente ricco di rumore.

2.2.1. ESTRAZIONE DEI TERMINI MEDIANTE IL SOFTWARE T2K

I file pdf delle pubblicazioni selezionate sono stati poi convertiti in formato testo con codifica UTF-8, al fine di essere predisposti alle successive operazioni di estrazione terminologica automatica. A tal fine, si è deciso di adoperare il software T2K (Text to Knowledge), sviluppato dal CNR ILC (Istituto di Linguistica Computazionale) di Pisa, in quanto le funzionalità risultavano familiari perché già adoperato e per via del continuo supporto offerto dagli sviluppatori per la customizzazione e l'utilizzo ottimale del software. Sono stati effettuati diversi test prima di raggiungere una configurazione del term extractor che consentisse di ottenere risultati soddisfacenti. Oltre alle liste dei termini con associata la relativa frequenza, è stato scelto di inserire anche la funzione di lemmatizzazione automatica, di modo da evitare di doverla effettuare successivamente manualmente. È stato necessario ripulire e normalizzare le liste terminologiche ottenute in quanto non esenti da errori, termini non pertinenti al dominio o termini troppo generici. Questo passaggio è stato effettuato manualmente selezionando di volta in volta i termini pertinenti ai tre domini specialistici individuati e predisponendoli così ad essere utilizzati per la creazione del thesaurus.
È interessante rilevare una peculiarità emersa nel corso di questa fase del lavoro. All'interno delle liste di estrazione terminologica, fra le patologie non oggetto di questo studio con più alta frequenza vi sono quelle tumorali. Ciò indica che numerose sono le ricerche e quindi le pubblicazioni relative a questo settore clinico in correlazione con patologie più comuni, quali quelle croniche, di cui la maggior parte della popolazione fa esperienza nel corso della vita.

2.2.2. MAPPING TRA LE DIVERSE RISORSE DI RIFERIMENTO

Per l'integrazione dei termini del thesaurus precedentemente estratti dal corpus di riferimento (si veda paragrafo 2.1) e soprattutto al fine di arricchire il thesaurus di sinonimi, si è scelto di fare riferimento a classificazioni mediche standardizzate, utilizzate quotidianamente per la pratica di codifica dei dati clinici sia nell'ambito delle cure primarie che secondarie. A tal fine sono state eseguite le seguenti attività:

- sulla base delle malattie croniche estratte dal corpus e delle malattie croniche o invalidanti indicate nel "REGOLAMENTO DI INDIVIDUAZIONE DELLE MALATTIE CRONICHE E INVALIDANTI, ai sensi dell'art. 5 comma 1, lettera a) del D. Lgs. 29 Aprile 1998 n. 124", sono state individuate le corrispondenze terminologiche e semantiche (mediante operazioni di mapping), all'interno dei seguenti sistemi di classificazione:
  - ICPC2: Classificazione Internazionale delle Cure Primarie – 2a Revisione;
  - ICD9-CM: Classificazione Internazionale delle Malattie – 9a Revisione Modificazioni Cliniche;
  - ICD10: Classificazione Internazionale delle Malattie – 10a Revisione.

  L'attività è stata effettuata tramite l'uso di transcodifiche esistenti tra le classificazioni (es. tra ICPC2 e ICD10, tra ICPC2 e ICD9-CM, ecc.) e l'uso del tool "UMLS Terminology Services" che permette di fare delle ricerche, in diverse lingue, sia per termine che per codice, all'interno del già citato Metathesaurus UMLS (Unified Medical Language System) che include 139 risorse terminologiche tra vocabolari, thesauri, nomenclature, di ambito biomedico in 21 lingue;
- nel caso di corrispondenza esatta dei termini con le classi delle classificazioni ICPC2, ICD9-CM e ICD10, sono stati identificati i sinonimi utilizzati per ciascuna classe e gli stessi sono stati integrati all'interno del thesaurus;

- il numero di termini individuati è stato esteso attraverso l'identificazione di ulteriori malattie croniche o invalidanti all'interno delle classificazioni ICPC2, ICD9-CM e ICD10, sulla base delle relazioni gerarchiche presenti nelle classificazioni per la specificazione di quei termini troppo generici che sono stati estratti dal corpus di riferimento o presenti sulla lista ministeriale delle malattie croniche.

L'integrazione dei termini presenti nel thesaurus, prevede anche l'associazione di termini comuni (anche detti "laici", ossia comunemente usati dai pazienti e cittadini in generale) ai termini preferiti (per lo più standardizzati in sistemi di classificazioni e comunque specialistici) in modo da garantire un accesso multiregistro al thesaurus e quindi, di conseguenza, una sua più vasta adoperabilità. A tal fine si è provveduto a mappare i termini del thesaurus (già integrati con i termini individuati nelle classificazioni) ad un vocabolario medico comune, nello specifico il Vocabolario Medico Comune dei Cittadini (ICMV - Italian Consumer Oriented Medical Vocabulary) [7]. In questo caso è stato semplicemente effettuato un mapping manuale 1:1 tra i termini del thesaurus e i termini dell'ICMV. L'operazione è stata facilitata grazie alla presenza di corrispondenze di vario tipo tra l'ICMV e alcune delle classificazioni standard utilizzate nelle fasi precedenti (ad es. ICPC, ICD10).
Al termine di questo processo è stata infine effettuata la traduzione dei termini in inglese, per garantire l'aspetto multilingue. Anche in questo caso ci si è serviti del tool UMLS già menzionato e delle versioni inglesi delle classificazioni di riferimento.

3. LA CREAZIONE DEL THESAURUS

Il thesaurus, fin dalla sua creazione, ha sempre assolto delle funzionalità di fondamentale importanza che grazie alle nuove tecnologie sono state aumentate di valore e ha consentito, di conseguenza, di poter allargare la sua intrinseca funzione ai diversi contesti d'uso nella gestione dell'informazione in ambiente digitale. Una risorsa che raccolga e strutturi i concetti rappresentativi di un dominio, più o meno specialistico, attraverso relazioni di equivalenza, gerarchiche e associative definite sulle proprietà intrinseche dei concetti stessi, comporta un valore aggiunto per qualsiasi contesto che richieda descrizione, rappresentazione e recupero di informazione e documenti.
Il concetto di thesaurus e le operazioni necessarie alla sua realizzazione sono oggetto di una normativa tecnica aggiornata molto recentemente: la ISO 25964-1:2011, Information and documentation – Thesauri and interoperability with other vocabularies, Part 1: Thesauri for information retrieval, sostituisce infatti le precedenti norme in materia, ormai datate, ovvero la ISO 2788:1986, Documentation – Guidelines for the establishment and development of monolingual thesauri, la ISO 5964:1985, Documentation – Guidelines for the establishment and development of multilingual thesauri.
Nello specifico, le principali funzionalità, strettamente interrelate, del thesaurus possono essere così sintetizzate: controllo terminologico, indicizzazione, supporto nella definizione dei metadati e classificazione sono attività che competono al professionista dell'informazione, mentre navigazione, ricerca ed espansione dei risultati delle ricerche coinvolgono direttamente l'utente come utilizzatore di tale strumento.

3.1. Controllo terminologico

La funzione di controllo viene esercitata prevalentemente per i fenomeni di sinonimia e di polisemia. All'interno di un thesaurus il significato di un termine deve essere assolutamente non ambiguo e la definizione della rete semantica tra i concetti permette di specificare quanto più possibile il significato che si vuole privilegiare in uno specifico contesto.

In tal senso è importante sottolineare anche l'utilizzo di note d'ambito, che consentono di inserire delle definizioni o delle specificazioni ulteriori circa un dato termine, e di qualificatori, che permettono di specificare l'ambito o il sotto-ambito al quale un termine appartiene, contribuisce alla determinazione univoca di un concetto. Allo stesso modo, le relazioni di equivalenza gestiscono la presenza di sinonimi, varianti lessicali, quasi - sinonimi, attraverso l'elezione, sulla base di diversi criteri, di un termine preferito al quale gli altri restano legati in quanto punti di accesso all'informazione. Tale legame permette di essere indirizzati verso il termine preferito e di arrivare comunque all'informazione ricercata, anche se indicizzata tramite i descrittori. Il concetto di controllo si traduce soprattutto in un requisito fondamentale per garantire l'incontro fra lessico dell'indicizzatore e lessico del ricercatore, e cioè la relazione biunivoca fra termine e concetto, fra significante e significato: ciò significa che in un thesaurus un termine esprime sempre uno ed un solo concetto, e che un concetto è sempre espresso da uno ed un solo termine.

3.2. Indicizzazione

L'indicizzazione è la funzionalità principale riconosciuta ad un thesaurus. Tale strumento, infatti, struttura un set di termini che possono essere selezionati ed attribuiti ad un documento in quanto rappresentativi del suo contenuto concettuale. Le relazioni tra i concetti consentono al professionista dell'informazione, in un contesto di indicizzazione manuale, di scegliere i termini in base al livello di specificità che più si adatta alle caratteristiche del corpus documentale, sia questo in forma cartacea o digitale. I descrittori associati ai documenti ne permettono poi il recupero a partire da differenti modalità di accesso all'informazione.

3.3. Navigazione e ricerca

Nei sistemi di content management che prevedono l'integrazione di un thesaurus la presentazione sistematica che mette in evidenza la struttura classificatoria definita può rappresentare il punto di partenza per operazioni di browsing. Anche la presentazione alfabetica può tuttavia essere navigata per orientare la ricerca di informazione da parte dell'utente, soprattutto in contesti in cui il suo bisogno informativo non è perfettamente definito. La visualizzazione delle relazioni tra i termini fornisce un indubbio supporto anche per la migliore conoscenza del dominio, soprattutto per utenti con poche competenze specialistiche. La presenza di relazioni di equivalenza permette principalmente di far fronte alla disomogeneità di competenze e di conoscenze da parte degli utenti, ma anche alla varietà della lingua, che comporta, anche in contesti specialistici, la presenza massiccia di sinonimi e varianti, al fine di recuperare l'informazione indipendentemente dal termine utilizzato per la ricerca o di guidare l'utente verso l'utilizzo del termine scelto come preferito.

3.4. Le relazioni del thesaurus

La metodologia usuale per navigare il formato alfabetico del thesaurus è attraverso l'impiego di un insieme standard di relazioni o tipi di rimandi. Le relazioni utilizzate sono [8]:

USE	Use	Usa	Rinvio da un termine non preferito ad uno preferito

UF	Used for	Usato per	Richiamo da un termine preferito ad uno non preferito
BT	Broader Term	Termine più generale	Rimando a termini dal significato più generale
NT	Narrower Term	Termine più specifico	Rimando a termini dal significato più specifico
RT	Related Term	Termine correlato	Rimando ad un termine che è collegato al primo in modo diverso da BT e NT

Tabella 2 Relazioni standard del thesaurus

L'ordine delle relazioni thesaurali è sempre il medesimo, il primo gruppo di relazioni (USE e UF) è legato al controllo del vocabolario, mentre le altre (BT,NT, e qualche volta RT) si riferiscono alle relazioni gerarchiche nel thesaurus. Le note d'uso (o note d'ambito) sono utilizzate solo nel caso in cui il termine descritto possa avere più significati risultando, di fatto, ambiguo. Molto spesso si tratta di una breve definizione del termine, ma vi sono casi in cui un esempio è necessario per chiarire al meglio il significato e l'inquadramento generale di un termine.

3.5. Pubblicazione sul web e integrazione del thesaurus in applicazioni esistenti

La disseminazione del thesaurus come risorsa disponibile sul web o come sistema integrato in altre applicazioni, e di conseguenza i formati e i protocolli da usare dipendono dall'obiettivo che la costruzione del thesaurus stesso si pone. Se, infatti, un thesaurus ha lo scopo di essere utilizzato per il recupero di informazioni (Information Retrieval) allora esso deve poter essere propriamente e completamente integrato in sistemi in cui hanno luogo le funzioni di indicizzazione, navigazione e di ricerca. Alcuni sistemi di indicizzazione o ricerca (es. Content management systems) hanno già integrato un modulo di mantenimento del thesaurus. Se questo è utilizzato anche per lo sviluppo del thesaurus, non è richiesto alcun import o export del thesaurus. Ciò nonostante, se il sistema integrato non ha meccanismi di esportazione del thesaurus in un formato standard ci possono essere difficoltà nel momento in cui si presenta il bisogno di cambiare sistema di gestione del thesaurus o di rendere il thesaurus disponibile in altre applicazioni. Inoltre Il sistema di indicizzazione e ricerca che utilizza il thesaurus deve essere capace di importare il thesaurus nello stesso formato.

Nel caso specifico, il software utilizzato per la costruzione del thesaurus MultitesPro 2007 permette sia l'import che l'export in formati standard, in particolare XML e SKOS/RDF, CSV, HTML. L'export in HTML crea un semplice sito web per l'accesso al thesaurus, mentre per il deployment e la pubblicazione del thesaurus su un server Multites prevede l'uso del pacchetto MultiTes EDK[2].

[2] http://www.multites.com/productsEDK.htm

Un'altra possibilità per la disseminazione del thesaurus è quella di renderlo disponibile in modalità stand-alone, come risorsa a se stante, distribuita su CD-ROM, o su un website, un'Intranet o più in generale su Internet.

A meno che la gestione del thesaurus non sia già integrata in un sistema con le applicazioni di indicizzazione e ricerca, il primo requisito quindi è quello di esportare i dati del thesaurus dal thesaurus management system (nel nostro caso MultitesPro) al sistema di information retrieval che integrerà il thesaurus, rispettando i formati e i protocolli del caso. L'uso di un formato di scambio comune garantisce l'interoperabilità tra applicazioni diverse. Tra i formati di scambio più comuni, quello più utilizzato negli ultimi anni è SKOS (Simple Knowledge Organization Systems), standard W3C particolarmente utile per la rappresentazione di risorse per il Web Semantico, codificato in XML e RDF (Resource Description Framework). L' esportazione del thesaurus in formato SKOS permette la possibilità di pubblicare il thesaurus come Linked Data (LD), favorendo cosi la condivisione e il riuso dei dati del thesaurus nonché l'integrazione con altre risorse del Linked Data cloud.

Sia nel caso in cui il thesaurus venga pubblicato in rete, che nel caso dell'integrazione in applicazioni di information retrieval è necessario anche utilizzare un protocollo standard per lo scambio dei dati o di un sottoinsieme di essi.

Oltre a protocolli general-purpose, esistono protocolli sviluppati specificatamente per l'interrogazione diretta di thesauri ai fini di indicizzazione e/o recupero delle informazioni, e quindi possono essere impiegati per presentare e applicare i thesauri, i concetti che ne fanno parte, i termini e le relazioni, per descrivere il significato dei termini e facilitare l'interoperabilità semantica. Sebbene alcuni siano molto usati e più conosciuti, la selezione e l'uso di un protocollo piuttosto che un altro dipende dai bisogni dell'applicazione, soprattutto lo scopo e l'ambiente software.

Tra i protocolli specifici per thesauri indicati dallo standard ISO 25964-1:2011, quelli che potrebbero essere utilizzati nel caso specifico del thesaurus delle malattie rare e del thesaurus delle malattie croniche sono:

- SWAD-E SKOS API – delle web service API che sono disegnate per fornire l'accesso a thesauri e altri sistemi di organizzazione della conoscenza via web[3]. Queste API, create in Java, definiscono un core set di operazioni per l'accesso e l'interrogazione programmatica di un thesaurus. L'uso di queste API è adatto per thesauri in formato SKOS (e come detto in precedenza MultitesPro permette l'export del thesaurs in SKOS), ma possono essere adattate anche ad altri formati. Alcuni esempi di chiamate via web service fatte tramite le API SKOS possono essere: getConcept (uri); getConceptsMatchingKeyword/Regex (string); getAllConceptRelatives (concept); getSupportedSemanticRelations; getAllConceptRelatives (concept, relation); ecc.

- Altre API create ad hoc o adattate – Esistono vari adattamenti delle API SKOS su varie piattaforme. Vi sono inoltre altri protocolli simili ma non relativi ai thesauri, sia basati su piattaforme SOAP che su REST XML RPC o JSON-RPC.

Come accennato in precedenza la scelta dipenderà dallo scopo per cui i due thesauri dovranno essere utilizzati. E' auspicabile ad ogni modo la pubblicazione degli stessi come Linked Data.

[3] http://www.w3.org/2001/sw/Europe/reports/thes

4. CONCLUSIONI

La raccolta e l'analisi della terminologia propria dei domini specialistici oggetto del presente studio sono state le fasi che hanno richiesto un lungo e costante impegno nel corso dello svolgimento dell'attività. Convinti che un'accurata selezione delle risorse sia punto di partenza imprescindibile per la realizzazione di un prodotto realmente usabile, grande attenzione è stata dedicata a comprendere quali risorse, sia propriamente terminologiche che di letteratura scientifica, fossero più significative all'interno dei domini di riferimento. Sono state effettuate delle scelte metodologiche limitando da una parte, per le malattie rare il campo di analisi alle principali terminologie specialistiche di settore, e dall'altra per le malattie croniche invalidanti, sono stati analizzati i principali sistemi di classificazione delle malattie esistenti ed è stato fatto un censimento della letteratura scientifica presente su PubMed, il principale riferimento bibliografico in ambito biomedico.

Nello specifico delle malattie rare l'obiettivo di partire da termini specialistici per ritrovare le corrispondenze nei principali sistemi di classificazione al fine di costruire una base di conoscenza arricchita, ha incontrato delle criticità, in particolare, sui termini per i quali, durante la mappatura, non è stata rintracciata alcuna corrispondenza. Ciò può anche trovare giustificazione nei seguenti casi: la patologia in questione potrebbe essere estremamente rara e, quindi, non essere stata introdotta nelle risorse terminologiche analizzate; le patologie rare recentemente scoperte sono state introdotte nelle liste terminologiche specialistiche dell'ambito specifico delle malattie rare, poiché aggiornate di continuo, ma non nelle risorse terminologiche analizzate che comprendono in generale tutto il dominio biomedico (ICD, SNOMED, OMIM, MeSH), che hanno, nella maggior parte dei casi aggiornamenti semestrali o annuali. Un'altra ragione può dipendere anche dalla definizione della malattia rara in sé. Come è stato già detto, la definizione di "rara" può dipendere da fattori geografici, alcune patologie possono essere considerate rare in alcune aree e frequenti in altre. I risultati ambigui ottenuti possono essere parzialmente spiegati se si pensa al differente orientamento che i termini possono avere all'interno di un sistema di classificazione. Ad esempio, il concetto di Orphanet con identificativo 30 "*Oroticaciduria*" viene mappato verso tre concetti nell'UMLS: C0268128, C0220987 e C0268131. Nel database di Orphanet il concetto *Oroticaciduria* risulta avere quattro relazioni di sinonimia con altri termini, mentre nell'UMLS, i diversi termini che fanno parte sempre dello stesso concetto, sono isolati e, quindi, non aggregati, come mostrato di seguito:

Orphanet 30	**CUI 1 C0268128**	**CUI 2 C0220987**	**CUI 3 C0268131**
Oroticaciduria	Orotic aciduria	/	/
Orotic aciduria hereditary	/	Hereditary orotic aciduria	/
Orotidylic decarboxylase deficiency	/	/	Hereditary orotic aciduria, type 2
Uridine monophosphate synthetase deficiency	/	/	/

TABELLA 3 RELAZIONI DI SINONIMIA TRA ORPHANET E UMLS

Per quanto riguarda le malattie croniche, invece, le criticità riscontrate dipendono anche dalla definizione stessa della condizione di "cronico", poiché molto spesso tale condizione dipende dal quadro clinico generale del paziente e dall'insieme di sintomi che, nel tempo, definiscono la caratteristica di cronicità della patologia

da cui è affetto. Proprio per questi motivi è importante che ci sia un'integrazione delle diverse denominazioni delle singole patologie, a partire dalle diverse risorse disponibili, ed è maggiormente importante che ciò venga fatto fra risorse nate per essere usate in contesti differenti, quali il Vocabolario Medico Comune dei Cittadini (ICMV - Italian Consumer Oriented Medical Vocabulary) e l'International Classification of Diseases 9th revision – Clinical Modifications, perché il thesaurus che si vuole realizzare possa essere suscettibile di diverse destinazioni d'uso. Come mostrato in tabella 4, nei termini presenti nella lista ministeriale delle patologie croniche invalidanti, si perdono informazioni circa la specificità della patologia poiché si utilizza il termine appartenente ad un livello semantico più generale.

Termini Comuni & ICMV	**Termine ICPC**	**ICPC termini di inclusione**	**Termine ICD9-CM**	**Termine ICD10**	**Lista ministero**
Soffrire di pressione bassa; Ipotensione	Ipotensione posturale		Ipotensione cronica	Ipotensione idiopatica	Affezioni del sistema cardiocircolatorio
Soffrire di pressione bassa; Ipotensione	Ipotensione posturale	ipotensione ortostatica	Ipotensione ortostatica	Ipotensione ortostatica	Affezioni del sistema cardiocircolatorio
Soffrire di pressione bassa; Ipotensione	Ipotensione posturale	Ipotensione idiopatica, ipotensione ortostatica	Ipotensione non specificata	Altra ipotensione	Affezioni del sistema cardiocircolatorio
Malattia cardiaca cronica, Cardiopatia; Cardiopatia ischemica	Cardiopatia ischemica con angina		Malattia cardiopolmonare cronica, non specificata	Cardiopatia Ischemica, non specificata	Affezioni del sistema cardiocircolatorio

TABELLA 4 CORRISPONDENZE TRA ICMV, ICPC, ICD9-CM, ICD10 E LISTA MINISTERIALE

Nel thesaurus creato, le informazioni vengono aggregate tramite un reticolo semantico che rappresenta i concetti in modo non ambiguo, favorendo quindi la standardizzazione della terminologia clinica. Al fine di testare la validità dello strumento realizzato nel dominio di applicazione, sono stati effettuati alcuni test da parte di un esperto di settore, che ha particolarmente apprezzato l'aspetto multiregistro del vocabolario. Tuttavia, sarà necessario effettuare un puntuale controllo sistematico del thesaurus in tutte le sue parti.

È doveroso sottolineare, tuttavia, come la costruzione di un sistema di organizzazione della conoscenza, nel caso specifico di un thesaurus, richieda, un margine di soggettività che si manifesta nelle scelte compiute per la selezione dei termini, per la loro disposizione al suo interno e per le modalità di applicazione dei principi propri alla metodologia e via dicendo. Il supporto di una base di conoscenza strutturata e il rispetto della norma in materia garantiscono, tuttavia, che lo strumento creato sia comunque un forte strumento di supporto per la refertazione, la ricerca e l'indicizzazione delle risorse nell'ambito specialistico di riferimento.

5. APPENDICE 1 - ESTRATTO DEL THESAURUS DELLE MALATTIE RARE

Aarskog, sindrome di
SC: H Malattie genetiche
L Anomalie dello sviluppo durante l'embriogenesi
UF: Displasia faciogenitale
ENG: AArskog Syndrome

Aase-Smith di tipo 1, Sindrome di
USE: Aase-Smith, Sindrome di
ENG: AAse Syndrome, type 1

Aase-Smith, Sindrome di
UF: Aase-Smith di tipo 1, Sindrome di

Acalasia-microcefalia
SC: G Malattia neurologiche
L Anomalie dello sviluppo durante l'embriogenesi
ENG: Achalasia - microcephaly

Aceruloplasminemia
NT: Aceruloplasminemia congenita
Neurodegenerazione con accumulo cerebrale di ferro
ENG: aceruloplasminemia

Aceruloplasminemia congenita
BT: Aceruloplasminemia
ENG: aceruloplasminemia congenital

Acrocefalosindattilia
BT: Craniosinostosi sindromica
NT: Sindrome di Apert
Sindrome di Pfeiffer
Sindrome di Saethre-Chotzen
ENG: acrocefalosindattilie syndromes

6. APPENDICE 2 - ESTRATTO DEL THESAURUS DELLE MALATTIE CRONICHE INVALIDANTI

Ghiandole paratiroidee
USE: Paratiroidi
ENG: parathyroid glands

Gozzo
BT: Processi patologici del sistema endocrino

Insufficienza surrenalica
BT: Processi patologici del sistema endocrino

Iperparatiroidismo
BT: Processi patologici del sistema endocrino
NT: Iperparatiroidismo secondario
RT: Paratiroidi
ENG: hyperparathyroidism

Iperparatiroidismo secondario
BT: Iperparatiroidismo
ENG: secondary hyperparathyroidism

Ipersurrenalismo
BT: Processi patologici del sistema endocrino

Ipogonadismo
BT: Processi patologici del sistema endocrino

Ormoni paratiroidei
RT: Paratiroidi

Ormoni tiroidei
RT: Processi patologici del sistema endocrino

Paratiroidi
UF: ghiandole paratiroidee
RT: Iperparatiroidismo
Ormoni paratiroidei

7. RIFERIMENTI BIBLIOGRAFICI

1. Orphan Drug Act. Public Law 97-414- Jan.4, 97th Congress, 1983
2. O'Halloran JF, Miller GC and Britt H. Defining chronic conditions for primary care with ICPC-2. Family Practice 2004; 21: 381–386.
3. WHO, Global status report on noncommunicable diseases, 2010
4. M. Milicic-Brandt, A. Rath, A. Deverau, S. Ayme, *Mapping Orphanet Terminology to UMLS*, in AIME 2011 Proceedings of the 13th conference on Artificial intelligence in medicine, pp. 194-203, Spring-Verlag:Berlin, Heidelberg, 2011.
5. T. Merabti, M. Joubert, T. Lecroq, A. Rath, S.J. Darmoni, Mapping biomedical terminologies using natural language processing tools and UMLS: mapping the Orphanet thesaurus to the MeSH, IRBM, Vol. 31, Issue 4, September 2010, pp. 221-225.
6. E. Pasceri, Analyzing rare diseases in biomedical terminologies, in JLIS.it, vol.3, No.1 8 (2012).
7. E. Cardillo, A Lexi-ontological resource for consumer healthcare: the Italian Consumer-oriented Medical Vocabulary, 2011.
8. V. Broughton, Costruire Thesauri: strumenti per indicizzazione e metadati semantici, (a cura di) P. Cavaleri, (traduzione di) L. Ballestra e L.Venuti, Milano, EditriceBibliografica, 2008, p.88.

<fine del documento>