

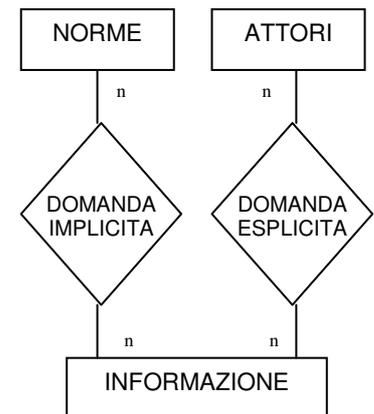
4 MODELLI PER LA CLASSIFICAZIONE DELLA DOMANDA INFORMATIVA

Lo studio del modello logico per la gestione dei dati sulla domanda informativa prende origine da alcuni assunti di base:

La domanda informativa implicita deriva dal sistema dei provvedimenti normativi ed è in sostanza una relazione che lega con cardinalità n-n una norma ad una particolare informazione.

La domanda informativa esplicita è espressione diretta degli attori impegnati nel governo del territorio ed è sostanzialmente una relazione che lega con cardinalità n-n un'attività di gestione effettivamente svolta ad una particolare informazione.

di conseguenza l'entità *informazione* si colloca in una posizione condivisa rispetto alle due direttrici di indagine.



4.1 Il "riferimento semantico" dei dati territoriali e il ruolo dei soggetti

Possiamo evidenziare una sorta di "flusso" con il quale far afferire ogni dato territoriale ad una o più categorie tematiche (macro temi). I temi rappresentano un sistema semantico con cui sistematizzare l'insieme dei dati territorio/ambiente e ci forniscono lo strumento base per la definizione di diversi percorsi di indagine:

ACQUISIZIONE → DATO → INFORMAZIONE → PROCEDURA → ATTIVITÀ → TEMA

Dall'acquisizione, mediante un processo di registrazione, si immagazzina un dato "grezzo", dal quale poi si elabora l'informazione (secondo l'assunto $\text{informazione} = \text{dato} + \text{intelligenza}$); l'informazione viene di conseguenza utilizzata in alcune attività di gestione territoriale attraverso la definizione di procedure/protocolli di utilizzo semplice o integrato; infine ogni attività di gestione appartiene ad uno o più temi secondo il sistema di classificazione di riferimento.

All'interno di questo processo gli attori si occupano del governo del territorio, mediante le attività di cui sono incaricati assumendo un particolare ruolo. In relazione al ruolo rivestito nell'attività di gestione, l'attore formula una domanda di informazione che può essere soddisfatta, parzialmente soddisfatta o non soddisfatta in ragione dell'incrocio tra matrice delle informazioni richieste con la matrice delle informazioni disponibili.

La matrici delle informazioni sono funzionali alla definizione di “pacchetti” di dati orientati ai segmenti di utenza o alle aree tematiche.

4.2 Struttura della domanda informativa

L’operazione di sistematizzazione della domanda di informazione presuppone un’analisi preliminare sugli attributi che ne caratterizzano la natura. È necessario in sostanza capire la “struttura” della domanda al fine di progettare correttamente il modello della base dati necessaria a registrare e classificare gli elementi rilevati sul panorama della richiesta di informazioni.

4.2.1 Classificazione dei contenuti – l’utilizzo di dizionari condivisi

La ricerca di un dizionario terminologico condiviso nasce dalla necessità di individuare un metodo razionale e sistematico per “descrivere contenuti” che sono l’attributo principale sia dell’offerta che della domanda; il “contenuto” che potremmo anche chiamare “tema” è infatti un carattere fondamentale sia di una risorsa informativa (layer GIS, banca dati o altro) sia della domanda di risorse informative. Ne deriva l’inderogabile esigenza di creare un “set” finito di termini utili a tale descrizione che permetta, nell’incrocio dei database della ricerca, la creazione di relazioni intrinseche tra domanda e offerta informativa.

Esistono alcune esperienze, sia nazionali sia a livello europeo, di realizzazione di glossari o tesauri per l’organizzazione semantica di contenuti; ad alcune di esse si fa diretto riferimento nei documenti dell’Intesa stato regioni, mentre la direttiva INSPIRE si limita a dare indirizzi generali oltre che definire un insieme di 43 categorie tematiche di riferimento.

Come si è detto, nell’ambito del protocollo d’intesa, un ruolo decisivo è svolto dal CNIPA - Centro Nazionale per l’Informatica nella Pubblica Amministrazione - che si occupa direttamente di problematiche fondamentali legate all’utilizzo delle tecnologie informatiche come il Codice dell’Amministrazione Digitale o la Firma Digitale oltre ad numerose tematiche legate all’e-Gov e alla sicurezza informatica. Nell’ambito della realizzazione del “Repertorio Cartografico Nazionale”, con una riunione del 28 febbraio 2006, l’organismo vara un documento che fornisce le linee guida per la definizione di “metadati” a livello nazionale necessari alla condivisione dei dati territoriali. All’interno dei documenti programmatici vengono inseriti due riferimenti a dizionari terminologici per la classificazione del contenuto informativo: “EARTH 2005” e “ThIST”; di “EARTH 2005” inoltre viene citato una non meglio definita

compatibilità con il thesaurus europeo "GEMET". Il testo di INSPIRE adottato dal parlamento europeo non fa invece riferimento esplicito a sistemi di classificazione di contenuti, tuttavia all'articolo 10 riporta: "Le disposizioni di applicazione riguardano i seguenti aspetti dei dati territoriali: / a) sistema comune di identificatori unici per gli oggetti territoriali; / b) rapporto tra oggetti territoriali; / c) principali attributi e corrispondenti tesauri multilingue comunemente richiesti per politiche, che possono avere un impatto diretto e indiretto sull'ambiente; / d) modalità di scambio delle informazioni sulla dimensione temporale dei dati; / e) modalità di scambio degli aggiornamenti dei dati. / 3. Le disposizioni di applicazione sono concepite in modo da garantire la coerenza tra le singole informazioni relative alla medesima località o tra le singole informazioni relative allo stesso oggetto rappresentato a scale diverse. / 4. Le disposizioni di applicazione sono tali da garantire che le informazioni ottenute da set di dati territoriali diversi siano comparabili per quanto concerne gli aspetti indicati all'articolo 9, paragrafo 2 e al paragrafo 2 del presente articolo"⁹. Si definiscono inoltre chiaramente le modalità di ricerca e individuazione delle varie risorse facendo esplicito riferimento all'utilizzo di parole chiave e "criteri di armonizzazione dei dati territoriali"¹⁰. (da valutare il riferimento alle voci tematiche riportati nella direttiva).

Mentre del "ThIST" (Thesaurus Italiano Scienze della Terra) progettato dall'agenzia APAT non esiste ancora documentazione specifica in linea, di EARTH (Environmental Applications Reference Thesaurus - progetto CNR) si legge che "Il contenuto terminologico (...) proviene da diverse fonti, multilingue e monolingue, di terminologia ambientale controllata (GEMET - General European Multilingual Environmental Thesaurus, Thesaurus Italiano di Scienze della Terra, Bollettino Terminologico della Conferenza di Rio su Ambiente e Sviluppo, altri documenti riguardanti settori specifici). Il progetto del CNR ha una prima versione del 2002 che deriva direttamente dal GEMET 2.0 (1999) dove "La compatibilità con il GEMET è completa a livello di concetto e di rappresentazione del medesimo"¹¹; ovvero dove l'identità dei descrittori è garantita dallo stesso sistema di codifica. Il volume 00 relativo alla presentazione di EARTH¹² contiene un capitolo intitolato "Introduzione al GEMET 2.0 -1999" nel quale viene descritta in dettaglio la strutturazione interna del GEMET.

Altre esperienze relative alla strutturazione di dizionari terminologici risultano meno strutturate e non orientate al territorio e all'ambiente, oppure strutturate

⁹ Dal testo adottato dal parlamento (art. 10 pag. 16, punti 2,3,4)

¹⁰ art. 8

¹¹ Dal sito "<http://uta.iaa.cnr.it/earth.htm>"

¹² "http://www.t-reks.cnr.it/docs/EARTH_pdf/Vol_00.pdf"

con riferimenti espliciti al GEMET; tra esse possiamo citare: "Eurodicautom"¹³, un database terminologico consultabile dal sito "europa.eu" disponibile in 12 lingue successivamente migrato nel progetto "IATE" (Inter Agency Terminology Exchange) la cui consultazione¹⁴ è ora riservata ad operatori istituzionali; il thesaurus multidisciplinare dell'UNESCO¹⁵ disponibile in inglese, francese e spagnolo organizzato in 7 macro temi (istruzione, scienza, cultura, scienze sociali, informazione comunicazione, amministrazione legislazione economia politica e aree geografiche) e strutturato gerarchicamente.

Le gerarchie terminologiche dei vari dizionari realizzati in ambito comunitario si rifanno quasi totalmente alla norma ISO 12620 del 1999 "Computer applications in terminology - Data categories" dove vengono indicate le modalità di classificazione terminologica. Nella sezione n.9 della ISO 12620 vengono definiti i termini "Thesaurus descriptor", "Top term", "Broader term", "Narrower term", "Related term", "Keyword", "Index header", che vengono diffusamente utilizzati per sistematizzare e strutturare gerarchicamente i vocaboli.

4.2.2 GEMET – General Multilingual Environmental Thesaurus¹⁶

Il "GEMET" è un dizionario terminologico multilingue realizzato a livello europeo dall'agenzia europea per l'ambiente "EEA" che si avvale del network denominato "EIONET" con il quale rende fruibile il contenuto dei documenti relativi alle attività istituzionali.

GEMET nasce con la funzione principale di strumento di gestione informativa dell'ETC/CDS ("European Topic Centre on Catalogue of Data Sources" - Centro Tematico Europeo per il Catalogo delle Fonti Dati) e ha origine dalla fusione del contenuto dei migliori tesauri realizzati in ambito europeo al fine di ottenere una base condivisa per la descrizione delle informazioni. Il progetto è stato gestito dal ministero dell'ambiente della Bassa Sassonia e ha visto la partecipazione di istituzioni italiane, austriache, tedesche e svedesi oltre che la collaborazione con altri stati membri e extra europei.

GEMET include termini provenienti dai seguenti dizionari disponibili nel 1996:

1. Parte dell'"Umwelt Thesaurus" dell'Umweltsbundesamt dell'agenzia tedesca dell'ambiente, 1995;

¹³ "<http://europa.eu.int/eurodicautom>"

¹⁴ "<https://iate.cdt.eu.int/iatenew/>"

¹⁵ "<http://databases.unesco.org/thesaurus/>"

¹⁶ "Introduzione al GEMET 2.0" - dalla documentazione relativa all'EARTH a cura del CNR - volume di presentazione della versione aprile 2002

2. L'intero "Thesaurus Italiano per l'Ambiente (TIA)" del CNR - 1994 disponibile in 4 lingue, completo del sistema interno di classificazione realizzato nel 1995;
3. L'intero "Multilingual Environment Thesaurus (MET)" del Nederlands Bureau voor Onderzoek di Amsterdam che deriva dal "Milieu-thesaurus" olandese disponibile in 8 lingue;
4. L'intero "EnVoc Thesaurus" dell'UNEP Infoterra - 1997 in 6 lingue;
5. L'intero "Thesaurus de Medio Ambiente" del Ministerio de Obras Publicas, Transportes y Medio Ambiente di Madrid - 1995 in 4 lingue;
6. L'intero "Lexique environnement - Planète" del Ministère de l'environnement di Parigi - 1995 in 2 lingue;
7. Singoli descrittori provenienti da documenti importanti dell'Agenzia Europea per l'Ambiente e dal "Thesaurus Eurovoc" del parlamento europeo - 1996.

Dopo il primo accorpamento dei termini il database viene sfoltito eliminando le ridondanze e applicando un sistema di codifica neutro (identificativo "non intelligente"). Successivamente i descrittori vengono organizzati in supergruppi e gruppi che attualmente sono 4 e 32 per la classificazione gerarchica; inoltre viene organizzata una lista di 40 temi per la classificazione tematica. La classificazione gerarchica adotta, secondo le indicazioni della ISO 12620, una strutturazione che partendo da un numero limitato di "Top terms" articola le relazioni tra descrittori di significato più generale ("Broader terms") e altri di significato più specifico ("Narrower term") oltre a descrivere interrelazioni semplicemente associativo ("Related terms").

Per tutte le lingue supportate esiste la traduzione dell'insieme dei descrittori; l'equivalenza semantica è stata verificata singolarmente da esperti delle rispettive nazionalità e la validazione dell'italiano è già effettuata mentre per altre lingue è in corso di ultimazione.

Il GEMET segue le norme ISO sui tesauri monolingui (ISO 2788, 1986) e multilingui (ISO 5964, 1985).

Le relazioni gerarchiche tra descrittori possono essere di due tipi:

1. generico: L'oggetto più specifico contiene tutte le caratteristiche di quello generale con l'aggiunta di alcune caratteristiche specifiche aggiuntive.
p.es. roccia(generale) -> argilla(specifico)
2. partitivo: L'oggetto più specifico è parte dell'oggetto generale.
p.es. parti di un edificio(generale) -> muratura(specifico)

queste relazioni sono di tipo “poligerarchico” ossia i descrittori possono avere un numero indefinito di termini più generali e di termini più specifici.

Le relazioni tematiche sono invece basate sull’insieme dei 40 temi oppure - complementariamente - sull’insieme dei 32 gruppi. Essendo i temi complementari (e non relazionati) ai gruppi la struttura del thesaurus si può anche immaginare sotto forma di matrice tema/gruppo.

Un terzo tipo di relazioni è di tipo “associativo” in quanto i descrittori correlati non hanno dipendenze gerarchiche; questo tipo di relazione ha origine dal lavoro di associazione descrittore-tema e correla ad esempio oggetti e la discipline che li studia, processi e i loro prodotti, significato opposto, causa ed effetto ecc.

L’insieme dei descrittori è un insieme di termini piuttosto efficace ai fini della classificazione dei contenuti informativi; utilizzeremo la strutturazione gerarchica e tematica del GEMET per ottenere informazioni dal database sui contenuti equivalenti e sul grado di “vicinanza” tra termini di input e termini correlati in output.

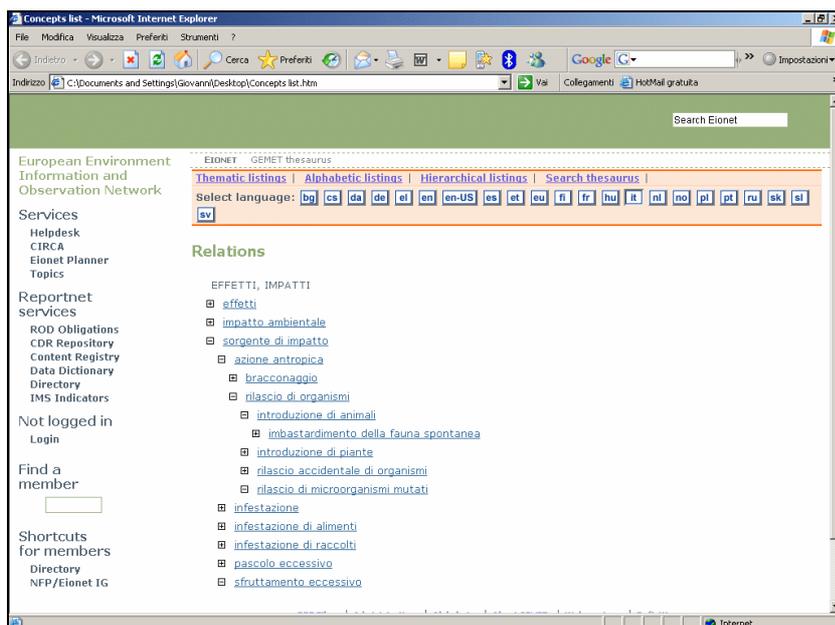


Figura 5: L'interfaccia di navigazione on-line di GEMET

Dal sito ufficiale è possibile ottenere l’intera banca dati completa della struttura gerarchica e tematica in formato xml che, con opportuno trattamento, è possibile importare nei più diffusi sistemi dbms esistenti sul mercato. La base dati della ricerca include perciò il contenuto informativo del GEMET con lo scopo di utilizzare non solo l’insieme dei vocaboli con la relativa codifica europea, ma anche l’insieme delle relazioni semantiche molto utile per ricavare, senza

La lista risultato fornisce una stringa dalla quale è possibile desumere la collocazione del termine ricercato nei vari contesti in cui è stato inserito nel thesaurus in quanto riporta "Broader terms" e "Narrower terms", "Related terms", gruppo e tema. Questo prospetto permette di individuare inoltre eventuali altri descrittori più vicini al reale significato ricercato.

4.2.4 Problematiche inerenti la classificazione di attributi spaziali

Oltre ad essere caratterizzate dal "contenuto" o "tema", le risorse informative possiedono attributi di tipo spaziale che possono essere sintetizzati in due categorie:

1. estensione geografica o "copertura" dell'informazione contenuta; ovvero la parte di territorio descritta dal contenuto della risorsa informativa;
2. grado di copertura; che quantifica quanta parte dell'estensione geografica è descritta dall'informazione; possiamo avere una copertura totale o parziale (a macchia di leopardo o con esclusione di fasce o isole).

Dalle informazioni relative ad ogni risorsa informativa è possibile dedurre quasi sempre le caratteristiche spaziali. L'oggetto della ricerca è tuttavia quello di andare oltre il trattamento indifferenziato di queste informazioni per definire, attraverso opportuni sistemi di classificazione o riferimento, un linguaggio comune sintetico ed efficace per confrontare gli attributi spaziali dei vari tipi di offerta e di domanda. La criticità dell'operazione è sempre determinata dalla disomogeneità delle entità trattate: dati da sensore, banche dati alfanumeriche, cartografie vettoriali o raster sono tipologie relativamente diverse di offerta informativa; quest'ultima si deve confrontare con la domanda che normalmente viene espressa con modalità assolutamente diverse.

L'efficacia della base dati integrata domanda offerta è dunque strettamente connessa ai sistemi di classificazione ivi compresi quelli di natura spaziale.

Per i dati da sensore l'informazione spaziale quasi sempre si identifica con la risoluzione spaziale che comunque viene descritta con due valori (minimo e massimo) che definiscono un "range" entro il quale è sensato l'utilizzo di quel dato. Per giacimenti e risorse informative invece esistono in sostanza due tipi di dato: il primo, cartografico, e di conseguenza dotato di una scala di rappresentazione, il secondo di tipo non cartografico per sua natura avulso da qualsiasi caratterizzazione geografico/spaziale. Essendo comunque piuttosto

semplice pensare ad un sistema di corrispondenza tra scala in pixel e scala frazionale, appare evidente la problematica di individuazione di scale di utilizzo per risorse informative come ad esempio basi dati alfanumeriche, pur quando provviste di dati geograficamente contestualizzabili come le coordinate cartografiche o i riferimenti toponomastici. Abbiamo inoltre un'informazione sul grado di omogeneità spaziale che indica se il dato copre con lo stesso grado di dettaglio l'intera area di estensione o solo una parte della stessa.

4.2.5 Problematiche inerenti la classificazione di attributi temporali

Anche per quel che riguarda le caratteristiche temporali del dato, è necessario ridurre l'indeterminatezza delle descrizioni libere definendo opportune scale di classificazione che consentano il confronto degli elementi di offerta e domanda. Nuovamente raggrupperemo le varie possibilità di articolazione dei caratteri temporali in voci sintetiche:

1. data di origine;
2. scala temporale;
3. ultimo aggiornamento.

La data di origine indica il momento in cui il dato viene prodotto per la prima volta e molto spesso non è determinante ma ne chiarisce il quadro di riferimento; la scala temporale invece fornisce indicazioni sugli intervalli con i quali il dato viene aggiornato o prodotto nuovamente mentre l'informazione sull'ultimo aggiornamento è utile per verificare l'utilizzabilità del dato in funzione dell'applicazione.

Per quanto riguarda i dati da sensore abbiamo quasi in tutti i casi a che fare con processi di produzione e aggiornamento sistematico e/o automatico; il dato della risoluzione temporale è chiaro ed utilizzabile direttamente per incrociare le basi dati. Per ciò che riguarda i giacimenti e risorse informative invece siamo in presenza di un'estrema varietà nei tempi e nei modi di produzione/riproduzione e manutenzione per cui le informazioni sulla data di creazione e di aggiornamento e della frequenza degli aggiornamenti, oltre a quella dell'omogeneità, devono essere necessariamente elaborati per ottenere un criterio semplificato e omogeneo di classificazione.

4.2.6 Problematiche inerenti la classificazione di attributi di formato

La classificazione del formato di un dato presume una preventiva chiarificazione del significato stesso di "formato". Non è da escludere infatti che, in ambiti disciplinari diversi, l'uso del vocabolo "formato" possa assumere valenze di varia natura. In ambito strettamente informatico ci si potrebbe limitare al "formato" come "protocollo di registrazione" del dato sul supporto di memorizzazione, tuttavia, volendoci spingere più al di dentro delle problematiche relative alla semantica dell'informazione territoriale, potremmo invece analizzare il modo in cui l'informazione è organizzata andando ad individuare le varie modalità di strutturazione dei contenuti geografici e di quelli non geografici. Se ci soffermiamo ad esempio sui dati geografici, la questione potrebbe spostarsi sui criteri di distinzione tra le varie modalità di georeferenziazione, oppure tra la geometria pura e l'organizzazione topologica o infine, tra informazione discreta e uniformemente distribuita.

Con buona probabilità, il "problema del formato" è tanto più interessante quanto più riguarda, non tanto il formato di memorizzazione del dato, quanto la possibilità di classificare la struttura interna della risorsa informativa in modo da permettere all'utente di capire se un determinato dato è quello che realmente serve alla sua applicazione. Potremmo citare il classico esempio dei "file delle strade"¹⁷ che, sia il netturbino, sia il portalettere, sia l'operaio delle fognature affermano di possedere, ma che risultano poi essere tre risorse informative assolutamente diverse in virtù dei diversi ambiti operativi dei tre soggetti: come poter quindi discernere tra: 1) un disegno dei sedimi stradali, 2) un grafo stradale con gli indirizzi, 3) Un archivio di sezioni stradali, se tutti vengono denominati "file delle strade"?

Il problema è forse più evidente per il dato geografico ma esistono anche questioni riguardanti gli archivi alfanumerici legate alle infinite modalità di inserire informazioni relative ad un determinato fenomeno in una base dati strutturata.

È probabilmente necessaria una fase di analisi delle fasi e delle metodologie di progettazione dei modelli logici allo scopo di individuare possibili "classi formali" mediante le quali sistematizzare le risorse informative dal punto di vista della loro struttura interna.

¹⁷ Robert Laurini - Università di Lione - Ontologie per le applicazioni geografiche (<http://lisi.insa-lyon.fr/~laurini/iuav-pdf/II-8.pdf>)