

UNIVERSITÀ DEGLI STUDI DI TRIESTE
Sede Amministrativa del Dottorato di Ricerca

XX CICLO DEL
DOTTORATO DI RICERCA IN
METODOLOGIE DI BIOMONITORAGGIO
DELL'ALTERAZIONE AMBIENTALE

**ANSwER, Sistema Informativo Ambientale
basato su ontologia e logica *fuzzy***

(Settore scientifico-disciplinare BIO 07)

DOTTORANDO

Davide Licheri

COORDINATORE DEL COLLEGIO

DEI DOCENTI

Chiar.mo Prof. Mauro Tretiach

Università degli Studi di Trieste

RELATORE

Dott. Alfredo Altobelli

Università degli Studi di Trieste

CORRELATORE

Dott. Stefano Sponza

Università degli Studi di Trieste

Yesterday

Yesterday

All those backups seemed a waste of play

Now my database has gone away

Oh I believe in yesterday

Suddenly

There's not half the files there used to be

And there's a millstone hanging over me

The system crashed so suddenly

I pushed something wrong

What it was I cannot say

Now all my data's gone away

And I long for yesterday

Yesterday

The need for backups seemed so far away

I knew my data was all here to stay

Now I believe in yesterday

Les G. Underhill, Direttore della Avian
Demography Unit, Cape Town, SA, 2006

La frase seguente è falsa;

La frase precedente è vera.

AA.VV.

I miei più cari ringraziamenti sono destinati a Gianni Benciolini, Alessandro Scotti e Lorenzo Serra che hanno stimolato, ognuno a suo modo, molte delle idee riferite in queste pagine.

Molto importante ho reputato il sostegno e la fiducia mostratami in ogni occasione dal Prof. Enrico Ferrero e da Fernando Spina.

Grazie a Barbara Amadesi, Chiara Campomori, Barbara Franzetti, Tanya Hubina e Manuela Panzacchi per le competenze GIS e di analisi; Gabriele Facchin, Fabrizio Florit e Massimo Zanetti per la velocità operativa nell'ottenimento dei dati ANSER, Valentina La Morgia per le ricerche bibliografiche e il frigorifero, Paolo Utmar e l'intera "compagnia" di censitori per i dati raccolti su campo.

Grazie "alla Iole e al Franco" per la solita infinita pazienza e a Jungle Julia per i suoi riccioli.

Sommario

1. Introduzione	1
1.1. Scenario e obiettivi	2
1.2. Frammento della realtà naturale	11
1.3. Contesto logico-informatico	19
2. Metodi	32
2.1. Protocolli di raccolta dei dati	32
2.2. Tecnologie di rappresentazione dei dati	34
3. Risultati	35
3.1. Integrazione sintattica	35
3.1.1. Analisi dei modelli attuali	35
3.1.2. Modello proposto	39
3.1.3. Integrazione EML	43
3.2. Inferenza semantica	48
3.2.1. Dimensioni del modello ontologico	48
3.2.2. Condizioni per classi primitive e classi complete	51
3.2.3. Gerarchia inferita dall'agente ragionatore automatico	56
3.3. Modello predittivo <i>fuzzy</i>	63
3.3.1. Variabili I/O e regole IF-THEN	63
3.3.2. Validazione	66
4. Conclusioni	72
5. Bibliografia	75

1. Introduzione

1.1. Scenario e obiettivi

Tra le premesse indicate nel progetto “ANSER” (Programma INTERREG IIIA Transfrontaliero Adriatico), viene sottolineata la necessità di risolvere la contrapposizione tra l'urgenza di tutelare il patrimonio naturale perdurante nelle lagune costiere dell'Alto Adriatico e l'esigenza di offrire un nuovo sviluppo sostenibile alle attività umane esercitate per profitto e fruizione. Legando gli assetti organizzativi di impresa alla salvaguardia delle caratterizzazioni territoriali vocate alla produzione, gli interventi gestionali che possono essere applicati nel tentativo di sostenere e incentivare i livelli di impiego sono individuabili con efficacia solo se non prescindono dalla conoscenza approfondita dell'ambiente in cui verranno praticati. Ulteriormente, la valorizzazione di ecosistemi vulnerabili come le zone umide salmastre si inserisce in una piattaforma internazionale di responsabilità di conservazione nel momento stesso in cui viene riconosciuto il legame strettissimo mostrato dalle popolazioni di uccelli acquatici, che si muovono indifferenti alle giurisdizioni nazionali, verso questi ambienti discontinui e sporadici.

A partire da questi presupposti, la raccolta secondo protocolli standardizzati di dati relativi al censimento della presenza e della densità delle popolazioni di uccelli acquatici, nelle diverse fasi fenologiche vissute nelle aree indagate, è volta a completare le conoscenze indispensabili alla conservazione del sistema complesso delle lagune costiere. Cioè si tenta di contribuire a una migliore definizione delle dinamiche ecologiche in atto, permettendo il confronto di questo corpo di informazioni rispetto alle variabili riferite ai fattori limitanti, abiotici e geografici contestuali. Queste specie animali infatti vivono e interagiscono all'interno di comunità eterogenee e ogni interazione che esse formano presenta un carattere

multifattoriale. Si creano così le “proprietà emergenti”, cioè quei fenomeni caratterizzati dal fatto di essere nuovi e imprevedibili, ossia di non poter essere spiegati (né previsti) sulla base dei principi e delle leggi che regolano il funzionamento delle parti costituenti il sistema complesso. Le proprietà esibite da tali fenomeni vengono definite "emergenti" in quanto emergono dall'organizzazione del sistema ecologico, attraverso le interazioni che si stabiliscono tra i diversi livelli del sistema stesso (Steven 2001). Poiché le proprietà emergenti a loro volta danno vita a ulteriori interazioni, nelle strutture dinamiche degli ecosistemi sussiste un numero ignoto di livelli di interazioni e proprietà emergenti.

Per una gestione ambientale efficace e per un bilanciamento dell'ecosistema, è necessario allora intraprendere ricerche a lungo termine in modo da permetterci di capire meglio gli articolati cambiamenti dell'ambiente e il loro impatto nel tempo. Nemmeno si può trascurare che, accanto all'estinzione derivante dalla componente adattativa connaturata al processo evolutivo dei viventi, a oggi il tasso di estinzione indotto dall'intervento antropico sugli ecosistemi appare almeno mille volte superiore al tasso naturale di estinzione (Primack 2004). Oltre il 95% delle estinzioni animali dal 1600 in poi sono state causate dalla caccia incontrollata, dalla distruzione dell'habitat, dai mutamenti climatici o dalla competizione con specie alloctone introdotte artificialmente.

La comprensione di sistemi ecologici così complessi richiede dati che siano di pertinenza di scale estese capaci di interessare spazi come i paesaggi, le regioni, il globo attraverso scenari temporali anche di secoli. Inoltre, allo stato attuale, i dataset che contengono le misurazioni degli elementi di un ecosistema non riflettono (ancora) sufficientemente le interazioni, sebbene la dimensione di questi archivi possa essere divenuta enorme. Per tradizione infatti gli ecologi non hanno dati condivisi, ma raccolgono, analizzano e pubblicano individualmente o in piccoli gruppi di ricerca. A titolo di esempio si consideri la natura frammentaria delle attività di ricerca sulle dinamiche di popolazione, dove l'analisi di set di dati individuali ha reso difficile formulare una teoria generale e investigare modelli tassonomici o spaziali su larga scala.

Quando invece le scale spaziali e temporali si dilatano e diventano appropriate alle questioni biologiche da risolvere, la proprietà analitica dell'ecologia viene affiancata da una nuova cultura “integrativa” delle modalità di indagine e delle fonti multiple di evidenza (Holling 2000). In tal senso la costruzione di modelli ecologici diventa importante per

approfondire la comprensione e la descrizione dei fenomeni naturali. Un modello efficace lega i dati alle questioni ecologiche e genera una quantità considerevole di predizioni, laddove una percezione del singolo ecologo fallisce con facilità. Oggi il *focus* moderno dell'ecologia richiede quindi di operare su collezioni di dati che sono al di là dell'abilità di assortimento del singolo indagatore o del singolo progetto di ricerca e quindi esige un processo di integrazione delle eterogeneità presenti lungo le scale spaziali, temporali e strutturali che caratterizzano le diverse pianificazioni di raccolta dei dati. Da queste considerazioni nasce la necessità di trovare nuove vie per incorporare la natura vivente dei sistemi naturali all'interno dei sistemi informativi ecologici (Meredith *et Al.* 2000)

Le innovazioni delle tecnologie per la raccolta dei dati, quali il *Global Position System* satellitare, i dispositivi *remote-sensing*, e la disponibilità di computer più rapidi e capaci, si sono legate alla accresciuta disponibilità di grossi e complessi archivi ecologici. A partire dalla nascita dell'*International Biological Program* (IBP, 1964-1974) e successivamente con il *Long-Term Ecological Research Program* (LTER, 1980-oggi), la scienza ecologica ha cominciato a contrapporre agli studi interessati alle singole componenti (dai batteri ai produttori, ai predatori), nuove ricerche sul sistema ecologico nel suo complesso e sui flussi di energia che attraversano gli ecosistemi (Baker *et Al.* 2005). Di conseguenza, tanto appaiono avanzate le tecnologie per la raccolta dei dati, tanto hanno fatto le tecnologie analitiche. Gli approcci informatici, quali le Reti Neurali Artificiali, gli Algoritmi Evolutivi e gli Alberi Decisionali, hanno mostrato una eccellente abilità per il miglioramento della costruzione di modelli ecologici (De'ath & Fabricius 2000).

Come altri domini delle “scienze della vita”, la pratica della scienza sulla biodiversità si è trasformata in una variabilissima rete in rapida espansione di relazioni tra istituzioni e avanzati sistemi informativi. I cambiamenti nella tipologia, scala e complessità delle domande poste dagli scienziati hanno provocato (e in alcuni casi sono stati provocate da) un ingente mutamento nella tecnologia dell'informazione e nei panorami computazionali della rete informatica. Questo nuova branca prende il nome di “eco-informatica” (altrimenti conosciuta come *biodiversity informatics*) e si propone di studiare la struttura inerente alla informazione ecologica al fine di creare e applicare la tecnologia informatica per la sua gestione e analisi. Specificatamente è la scienza che sviluppa basi di dati e algoritmi capaci di promuovere e facilitare la ricerca ecologica su larga scala.

Molti ricercatori in ecologia hanno iniziato a utilizzare le abilità della visione informatica per applicazioni note come i pattern predittivi della ricchezza di specie (Guegan *et Al.* 1998) o come le analisi multivariate per le invasioni biologiche (Kolar & Lodge 2002). Comunque sia, gli approcci eco-informatici non sono ancora largamente presenti nella scatola degli attrezzi dell'ecologo. Una delle ragioni alla base di questa diffusione rallentata sembra essere legata alla mancata esperienza di molti ecologi nel campo informatico, necessaria per operare con alcune implementazioni software (Fielding 1999). Si riscontra quindi una certa esitazione nell'investire il proprio tempo per apprendere un linguaggio e una sintassi del codice di programmazione (Williams & Poff 2006).

L'impostazione eco-informatica presenta però molti vantaggi rispetto ai modelli ecologici tradizionali, quali quelli basati su tecniche varie di regressione, spesso limitati dalla necessità di assunzione della normalità dei dati e delle conseguenti trasformazioni obbligatorie (Lek *et Al.* 1996). I modelli eco-informatici non sono afflitti da questi svantaggi e presentano performance più consistenti rispetto all'approccio tradizionale soprattutto quando analizzano set equivalenti di dati non lineari (Olden & Jackson 2002, Olden & Jackson 2001, Lek *et Al.* 1996).

Essendo un campo interdisciplinare, la ricerca in eco-informatica richiede di fatto uno sforzo di integrazione. La figura 1 descrive la struttura integrativa tra la modellizzazione dei sistemi ambientali, la misurazione e il monitoraggio ambientale e l'eco-informatica.

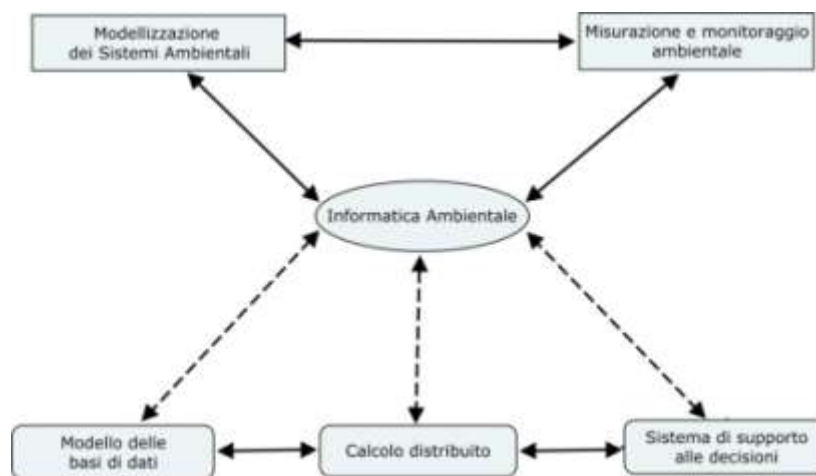


Figura 1

I sistemi di gestione ambientale così progettati generalmente presentano ottime caratteristiche di interattività e dinamicità, garantendo inoltre una profonda flessibilità nell'aggiornamento delle scelte adottate nel tempo, anche se drammaticamente diverse. D'altro canto la complessità di implementazione emerge nella determinazione dei parametri del sistema, nell'assetto delle relazioni interattive, nella formulazione degli approcci modellizzati, nelle interpretazioni dei risultati delle analisi, nell'implementazione delle raccomandazioni gestionali (Huang & Chang 2003).

Un argomento particolarmente critico per l'eco-informatica è l'adozione di uno standard per documentare i dati raccolti. Questa documentazione sui dati è formata da metadati, cioè da "dati sui dati", così tanto essenziali che la più semplice delle analisi ne richiede la presenza. Questo è vero per tutte le discipline, ma lo studio della biodiversità e, in linea di massima, l'ecologia sono particolarmente indebolite da queste circostanze. La maggior parte di noi ha sperimentato la difficoltà di utilizzare anche solo i propri dati, una volta trascorso il momento della raccolta, riconoscendo che, senza un'adeguata documentazione dei dati, questa difficoltà aumenta col passare del tempo. E' infatti ormai assodato che nel tempo la maggior parte delle informazioni vengono perse attraverso un lento processo di "decadimento dei dati" perché, in assenza di metadati, la nostra abilità di localizzare e interpretare il dato diminuisce o si perde del tutto.

A maggior ragione, l'integrazione e la condivisione di metadati in una comunità scientifica diventano necessari se si considera quanto i dati ecologici possono variare nei formati che riflettono le differenti motivazioni di raccolta dei dati stessi, nelle differenti condizioni delle variabili e nei diversi modelli di campionamento nello spazio e nel tempo. I metadati ecologici inoltre possono essere molto variabili in estensione, profondità e qualità di incertezza e accuratezza, potendo consistere di note mentali, appunti scritti a mano sul campo, commenti inseriti in un file di Excel® o altre forme di documentazione varia. L'integrazione delle fonti prevista dalla impostazione eco-informatica prevede quindi necessariamente nuovi metodi non solo per la rappresentazione formale e l'interscambio della conoscenza, ma anche per la gestione delle basi di dati che sia diretta verso assetti di federazione, navigazione, modellizzazione e simulazione guidati dai dati, tecnicamente cioè di *data-driven applications* (Meredith *et Al.* 2000).

Tuttavia la sola traduzione sintattica in un linguaggio formale comune non è sufficiente a definire o, meglio, a registrare tutta la ricchezza dei metadati ecologici. L'approccio eco-informatico richiede la manipolazione di grandi quantità di dati che presentano quelle complesse inter-relazioni descritte prima, proprie dei sistemi ecologici. Si è avvertita subito la necessità di mappare anche un ulteriore livello di metadati sui metadati sintattici attraverso l'implementazione di "ontologie di dominio" capaci di descrivere la semantica dei frammenti della realtà naturale di interesse. Per questo motivo, lo studio e lo sviluppo di formalismi che potessero permettere una rappresentazione strutturata della conoscenza ha conosciuto negli ultimi anni un'attenzione sempre crescente.

Questi formalismi di rappresentazione si sono sviluppati partendo dalla necessità di mappare i modelli delle applicazioni commerciali disponibili, attraverso lo sviluppo dei modelli semantici e concettuali più astratti. Questa convergenza tra i sistemi di gestione di banche dati e i linguaggi di programmazione moderni basati sul paradigma dello sviluppo orientato agli oggetti, ha procurato il progresso dei modelli di dati orientati agli oggetti, tra cui le ontologie di dominio.

In campo filosofico, l'ontologia è lo studio della natura dell'essere e della essenza delle cose. All'inizio degli anni '90 gli informatici, in particolar modo quelli che si occupavano di Intelligenza Artificiale, adottarono il termine, dirottandolo verso un nuovo ma correlato significato. Le moderne ontologie computazionali sono strutture pragmatiche di dati scaturite dalla necessità per i computer di cooperare nella condivisione delle informazioni e nella soluzione dei problemi (White 2005).

Un sistema di integrazione di dati espone ai propri utenti un unico schema su cui porre le interrogazioni. Questo schema è generalmente chiamato "schema mediato" (o schema globale). L'integrazione dei dati, divenuta cruciale per l'avanzamento di progetti scientifici su larga scala nei quali i set di dati vengono prodotti autonomamente da più ricercatori, offre una migliore collaborazione tra agenzie governative, ognuna con la propria fonte di dati, e garantisce una buona qualità di interrogazione attraverso milioni di fonti di dati strutturate sul *World Wide Web* (Halevy *et Al.* 2006). Tra i pionieri, l'*Information Manifold Data Integration Project* seguì la strada di offrire una interfaccia di interrogazione uniforme su una moltitudine di fonti di dati, liberando l'utente casuale dall'incombenza di localizzare le fonti di dati (che

costituiscono il cosiddetto *deep web*), di interagire singolarmente con ognuna e di combinare manualmente i risultati (Levy *et Al.* 1996).

Per rispondere alle interrogazioni usando tutte le fonti di informazione pertinenti, il sistema informativo integrato necessita di una mappatura delle relazioni semantiche tra lo schema mediato e gli schemi delle diverse fonti. Queste mappature costituiscono le componenti principali delle descrizioni delle fonti (Halevy *et al.* 2006) e possono essere formalizzate adeguatamente in una base di conoscenza attraverso lo sviluppo di una ontologia.

Una ontologia rappresenta quindi semplicemente una specifica formale ed esplicita di una concettualizzazione condivisa. La “specificità di una concettualizzazione” consiste nella descrizione della conoscenza che abbiamo di un certo dominio, tramite classi, relazioni fra classi e individui appartenenti a classi. La qualità “esplicita” significa che le classi, le relazioni fra classi e gli individui appartenenti a classi sono sconosciuti, anziché falsi o sbagliati, se non sono definiti e dichiarati esplicitamente (*Open World Assumption*). Inoltre, “formale” implica che l’ontologia sia comprensibile dalle macchine. Da ultimo, “condivisa” significa che l’ontologia cattura conoscenza consensuale e concordata da un gruppo, non solo individualmente. È facilmente prevedibile che, in concomitanza con l’emergenza del “semantic Web”¹, le ontologie giocheranno un ruolo via via più importante all’interno dello sviluppo dei sistemi informatici (Berners-Lee 1998).

La conoscenza così formalizzata è processabile automaticamente da un calcolatore, tramite un “agente ragionatore automatico” che implementa i processi inferenziali e deduttivi. Un agente ragionatore automatico è un software in grado di svolgere dei ragionamenti su basi di conoscenza adeguatamente formalizzate. Le possibilità del ragionatore dipendono strettamente dal linguaggio adoperato per formalizzare la base di conoscenza e possono essere suddivise in validazione, analisi e inferenza. Se il linguaggio adoperato prevede la possibilità di esprimere una conoscenza incompleta, il ragionatore potrà individuare quei fatti che sono impliciti in tutti i possibili modelli che soddisfano la base di conoscenza. La possibilità di effettuare automaticamente l’inferenza è l’elemento che più di

¹ E’ l’evoluzione del Web avvenuta a partire dal 1995 verso un ambiente dove è possibile pubblicare non più solo documenti ma anche informazioni e dati in un formato adatto alla interrogazione, interpretazione e, più in generale, elaborazione automatica.

altri mostra la differenza fra il ragionamento automatico e la semplice computazione (Knublauch *et Al.* 2004). E' opinione comune che, dopo la conclusione della modellizzazione o del disegno basati sui dati disponibili, l'intero lavoro sia in realtà arrivato solo alla metà del compimento. La restante parte consiste infatti nell'esaminare quanta informazione non disponibile, ma presente in quanto conoscenza implicita, sia possibile acquisire attraverso tecnologie dell'informazione innovative come l'intelligenza artificiale e il *data mining* (Huang & Chang 2003).

Si è più volte sottolineato come la raccolta di statistiche ambientali sia condotta con difficoltà legate alle scale ampie dei fenomeni ecologici, all'eterogeneità delle fonti di dati, alle organizzazioni coinvolte, così come alla complessità delle loro caratteristiche temporali e spaziali. Conseguentemente molti dati ambientali sono caratterizzati e soggetti a forte variabilità per quanto riguarda l'affidabilità, la consistenza e la quantità di errori. Nel favorire un incremento per quanto riguarda la qualità dei dati raccolti, è necessario un rigoroso lavoro di validazione durante l'immissione e comunque prima di ogni analisi desiderata (Briggs 1995). Oltre a questa componente di "sviste", cioè di errori che derivano dall'errata messa in pratica di una interpretazione corretta, è necessario confrontarsi con l'eterogeneità e l'incertezza insite nei dati archiviati nelle basi di dati ecologiche e nei sistemi informatici ecologici. I ricercatori infatti registrano e usano le informazioni provenienti da fonti di dati e conoscenza eterogenee: fonti di informazione oggettiva (primariamente quantitativa) come misurazioni o calcoli e fonti di informazione soggettiva (spesso solo qualitativa) come la conoscenza degli esperti e le loro valutazioni soggettive. La conoscenza con cui hanno a che fare gli ecologi solitamente mostra un carattere olistico che spesso non si esprime attraverso un'adeguata formalizzazione (Droesen 1996), provocando in molti campi della ricerca ambientale una mescolanza necessariamente soggettiva di informazione quantitativa e qualitativa.

Il problema della incertezza appare spesso nella modellizzazione ecologica, in particolare per quanto concerne l'incertezza dei dati che deriva dalla conoscenza degli esperti definita vagamente. Molta incertezza inerente ai dati ecologici risulta dalla presenza di variabili random, incomplete o inaccurate, e dalle stime approssimate che sostituiscono le misurazioni per problemi tecnici o finanziari. Infatti non tutti i parametri ecologici sono misurabili (ad esempio il numero assoluto di anatre in una zona umida) e il valore di quei

parametri può essere ottenuto solo attraverso metodi di stima e valutazione spesso di carattere soggettivo. Inoltre i dati ecologici possono anche avere differenti strutture e formati (ad esempio i dati di serie temporali e spaziali).

Esistono numerose vie per affrontare i problemi legati all'incertezza, tra cui le reti di inferenza probabilistica (Pearl 1988) o gli intervalli di confidenza (Shafer 1990). Ma questi metodi convenzionali di classificazione dei concetti che si basano sulla logica booleana ignorano il carattere continuo insito nella natura delle classi di oggetti ecologici, trascurando i confini sfocati delle variabili. Precisamente questa incertezza non-probabilistica dei dati ambientali si lega sia al giudizio dell'osservatore nel valutare un fenomeno, sia al fenomeno in sé.

D'altro canto, voler rinunciare alla valutazione della vaghezza significa assumere come preciso un dato che intrinsecamente non può esserlo e propagarne l'errore. Mentre in un database tradizionale la gestione dell'incertezza e della discendenza dei dati può essere considerata marginale, nel processo di integrazione di fonti di dati diviene una necessità. Infatti per loro natura, i dati provenienti da fonti multiple hanno una eterogeneità di accuratezza che può renderli inconsistenti gli uni con gli altri (Halevy *et Al.* 2006).

Lo strumento matematico-informatico di maggiore efficacia per trattare l'insieme dei problemi legati alla incertezza non-probabilistica espressa dai modelli ecologici si richiama alla teoria matematica per trattare dati vaghi e imprecisi, proposta dal Prof. Zadeh nel 1965 con il nome di "logica *fuzzy*". L'approccio *fuzzy* non significa un particolare metodo, bensì l'integrazione dei concetti *fuzzy* nei metodi convenzionali per processare la conoscenza e analizzare i dati. La naturale variabilità dei gradienti ambientali, l'elevato rumore di fondo, gli errori relativi all'operazione di campionamento e misurazione sono conservati lungo i passaggi del processo *fuzzy* e utilizzati come informazione aggiunta (Zadeh 2002).

I dati ecologici o le classi di oggetti ecologici possono essere definiti come insiemi *fuzzy* con limiti non netti, riflettendo meglio il carattere continuo tipico della natura. Poiché gli insiemi *fuzzy* possono essere usati per manipolare il ragionamento "inesatto", la logica *fuzzy* permette di lavorare con la conoscenza incerta sulle relazioni che esistono tra le componenti dell'ecosistema e i modelli basati su questo tipo di informazioni. La modellizzazione ecologica e l'analisi dei dati costituiscono quindi l'area di applicazione principale della teoria degli insiemi *fuzzy* nella ricerca ecologica. L'integrazione dei meccanismi di inferenza *fuzzy* e

delle tecniche dei sistemi esperti ha prodotto negli ultimi pochi anni lo sviluppo di strumenti per i sistemi esperti *fuzzy* e per i modelli dei processi ecologici basati sulla conoscenza *fuzzy* (Salski 1999). La modellizzazione basata su conoscenza *fuzzy* può essere particolarmente utile dove non esistono modelli analitici delle relazioni che devono essere esaminate oppure dove c'è un'insufficiente quantità di dati per le analisi statistiche, o ancora laddove il grado di incertezza di questi dati appare veramente alto (Bock & Salski 1998, Daunicht *et Al.* 1996, Li 1996, Hirota & Pedrycz 1996, Salski *et Al.* 1996, Bardossy & Duckstern 1995, Salski 1992). Infatti in questi casi l'unica base per la modellizzazione consiste nella conoscenza degli esperti, che è spesso incerta e imprecisa (Salski 2006).

Alla luce di queste considerazioni, la presente tesi intende suggerire una serie di esempi di trasformazione dei modelli di gestione dei dati di censimento raccolti nell'ambito del progetto ANSER, organizzandoli in un processo moderno di integrazione, formalizzazione e validazione delle informazioni disponibili.

Questo sviluppo si compie attraverso la realizzazione di tre obiettivi principali.

Il primo obiettivo che si intende raggiungere consiste nella integrazione, tramite una sintassi comune, di tutte le fonti dati rilevanti per il monitoraggio previsto. Il secondo obiettivo si esplica nella creazione di una ontologia di dominio capace di rappresentare i livelli semantici più generalizzati che governano le attività di monitoraggio di Vertebrati, nel caso particolare per le specie di uccelli di laguna. Infine, come terzo obiettivo, un modello predittivo *fuzzy*, basato sull'ontologia di dominio in precedenza descritta, verrà validato dai set integrati delle variabili ambientali e dei dati di monitoraggio del progetto ANSER.

1.2. Frammento della realtà naturale

Il concetto di zona umida è stato definito ufficialmente per la prima volta in concomitanza con una conferenza internazionale tenutasi nel 1971 a Ramsar, sulle coste iraniane del Mar Caspio, e viene espresso nelle dichiarazioni di un trattato intergovernativo che vede, a oggi, la sottoscrizione di oltre un centinaio di paesi. Quando si riferisce al contesto ecosistemico delle zone umide, la Convenzione di Ramsar intende designare *“le paludi e gli acquitrini, le torbe oppure i bacini, naturali o artificiali, permanenti o temporanei, con acqua stagnante o corrente, dolce, salmastra, o salata, ivi comprese le distese di acqua marina la cui profondità, durante la bassa marea, non supera i sei metri. Ai sensi della presente convenzione si intendono per uccelli acquatici gli uccelli ecologicamente dipendenti dalle zone umide”* (Ramsar Convention Bureau 2000).

L'asportazione graduale di suolo e roccia, operata dall'erosione atmosferica e dal dilavamento, viene convogliata dai fiumi verso il mare sotto forma di detriti quali ghiaie, sabbie, silt e argille. Sulla costa la rielaborazione dei clasti e dei sedimenti detritici da parte del moto ondoso e delle maree modella le foci dei fiumi, le spiagge, le dune e le lagune costiere. L'evoluzione geomorfologica del litorale è molto dinamica. La sabbia e la ghiaia sono ridistribuite lungo la linea prossimale di spiaggia mentre i silt e le argille, essendo di granulometria più fine, possono venire trasportati dalle correnti molto più facilmente e quindi sedimentano sul fondo più a rilento e in maniera più estesa. Se il bilancio sedimentario è positivo, cioè se prevalgono i fenomeni deposizionali su quelli erosivi, la costa sarà bassa e il litorale potrà articolarsi in cordoni e scanni sabbiosi che col tempo possono limitare specchi d'acqua più o meno collegati con il mare aperto. E' questa la situazione di laguna costiera nella quale la salinità di specchi d'acqua aperti e poco profondi è influenzata dal mescolamento delle acque dolci portate dai fiumi e di quelle salate che entrano per effetto delle mareggiate e del flusso di alta marea.

Se i collegamenti con il mare scompaiono e le onde di marea si interrompono, la laguna si trasforma in uno stagno retrodunale che, con la dissalatura progressiva delle acque salmastre, lascia il posto a un lago costiero. In alcuni casi la gestione artificiale di canali di collegamento tra il mare e un lago costiero poco profondo ha permesso l'estrazione di sale

marino mediante evaporazione (saline) oppure l'allevamento ittico e la pesca, se i fondali si presentavano più profondi (valli).

Le zone umide si trovano in ogni zona climatica del mondo. Le interazioni tra differenti componenti di idrologia, geomorfologia, climatologia, chimica del suolo e delle acque, risultano nell'espressione di una enorme diversità di habitat (Finlayson & Van Der Valk 1994, Gopal *et Al.* 1990). L'acqua, che con il proprio dinamismo rappresenta il principale elemento motore per tutte le risorse di un ecosistema di zona umida, imprime anche i ritmi di crescita stagionale delle piante e, di conseguenza, sincronizza i tempi di successione ecologica nella comunità animale. Tra le varie tipologie di zone umide, le acque costiere e le zone umide associate probabilmente sono tra le aree più stabili e facilmente prevedibili nel regime tidale giornaliero che fortemente le influenza. Questi cambiamenti del livello di marea sono prodotti dal sole e dalla luna all'interno di un assetto che gli animali possono facilmente interpretare durante tutto l'anno (Weller 1994).

E' curioso però come proprio dove la quantità dell'acqua appare più prevedibile, se da un lato sembra maggiore la disponibilità di cibo e delle altre risorse (Southwood 1977), dall'altro questa prosperità si manifesta più instabilmente (Figura 2). Infatti alcune risorse massive ma limitate temporalmente come le invasioni di invertebrati provocate dalle inondazioni primaverili dei fondali più bassi delle zone umide o le fioriture delle alghe, che sono sincronizzate con la disponibilità di nutrienti nelle acque più profonde, possono rappresentare modelli dal forte carattere stocastico.

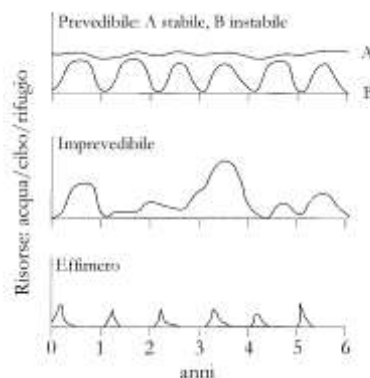


Figura 2, da Weller 1988, modificata

Da queste considerazioni è facile comprendere che gli schemi delle variazioni stagionali di una zona umida sono essenziali. Purtroppo sono pochi gli esempi in cui questi dati sono stati messi in relazione con le densità di presenza degli animali, con le loro caratteristiche di utilizzo dell'habitat e le differenze funzionali che li contraddistinguono.

Numerose specie di pesci, anfibi, rettili, uccelli e invertebrati dipendono da questi habitat dalle differenti caratteristiche idrologiche per alimentarsi, riprodursi, nidificare, anche se limitatamente ad alcuni stadi del loro ciclo di vita. Questi animali possono essere residenti nelle zone umide oppure periodicamente/stagionalmente concentrarsi nelle zone umide, essendo migrati da altri ambienti. Alcune specie, anche quelle domestiche legate all'allevamento, periodicamente usano le zone umide espressamente per pascolarci, eppure non ci risiedono. Il concetto di biota della zona umida quindi dovrebbe legittimamente essere esteso, per esempio, coinvolgendo tutti gli animali che vi dipendono dal punto di vista trofico. Secondo questa visione, la valutazione della biodiversità funzionale delle zone umide considera tutte quelle specie che traggono risorse direttamente o indirettamente dalle zone umide per qualche stadio della loro vita. La loro conservazione infatti non può prescindere dall'applicazione di misure di conservazione degli habitat delle zone umide (Weller 1999).

Il valore delle zone umide per la biodiversità è aumentato ancor di più per il fatto che un ampio numero di specie di uccelli acquatici richiedono due o più aree umide, solitamente di natura diversa, per la loro sopravvivenza durante le migrazioni in e tra continenti.

Le esigenze funzionali del ciclo di vita annuale degli uccelli acquatici spesso possono indurre un grado disuguale di mobilità, essenzialmente per:

- 1- mantenere le contingenze fisiologiche rispetto ai cambiamenti del clima;
- 2- migrare annualmente tra aree di riproduzione, di muta, di svernamento, se distinte;
- 3- ottenere i nutrienti essenziali per la deposizione delle uova;
- 4- alimentare i pulcini nell'area riproduttiva;
- 5- cercare isolamento e protezione durante i periodi di muta post-riproduttiva;
- 6- localizzare aree ricche di cibo durante la migrazione e lo svernamento.

Dopo aver raggiunto un'area in seguito a migrazione sulla lunga distanza, un uccello può continuare a compiere piccoli movimenti localizzati per la costituzione di piccoli territori ricchi di risorse, per l'individuazione di aree di alimentazione difese o indifese e per acquisire

familiarità con un più esteso territorio circostante (*home range*). Inoltre molti uccelli si muovono anche quando la latitudine, la temperatura o le condizioni dell'acqua non lo richiedono espressamente, indicando che altri fattori possono influire oltremodo sulla mobilità. Il più importante di essi sembra essere proprio il cibo, specialmente quando l'apporto nutritivo condiziona il positivo esito dei moduli comportamentali necessari per il ciclo vitale dell'individuo. Di fatto, il soddisfacimento delle necessità alimentari deve essere interpretato come un processo di ottimizzazione che incorpora varie strategie, tra cui adattamenti morfologici e fisiologici per predare, tattiche legate alla mobilità, selezione dell'habitat.

La ragione principale di questa ricchezza di biodiversità è presumibilmente legata alla disponibilità di risorse generata dall'enorme produttività primaria tipica delle zone umide. A sua volta, il mosaico di specie autotrofe che produce questa biomassa proliferata grazie alle polimorfe occasioni di sfruttamento offerte da numerosi micro-habitat a variabile gradiente chimico-fisico delle acque. Infatti le piante differiscono anche nell'apparenza strutturale delle comunità, cioè nei valori di altezza, diametro e robustezza dominanti. Con l'acqua, questa sagoma crea lo schema visivo utile a noi per caratterizzare fisionomicamente il tipo di zona umida e agli uccelli come valore di risorse essenziali per la sopravvivenza.

Dalle specie completamente acquatiche a quelle più terrestri, tutte le forme di vita vegetale risultano dall'adattamento alla presenza, profondità, persistenza, stagionalità, temperatura e chimica dell'acqua. Alcune specie di piante "sommerse" mostrano radici solo durante una fase della loro vita e sopravvivono galleggiando nella colonna d'acqua dopo che si sono staccate dalle radici, altre non presentano mai apparati radicali. Caratteristicamente vanno incontro a una rapida e completa decomposizione, formando un fertile detrito, dal quale trae nutrimento anche il plancton che sfama i pesci e gli altri organismi animali. Le piante sommerse sono specializzate per diverse profondità e substrati del fondo, mostrando una netta stratificazione a seconda della loro capacità fotosintetica in rapporto alla penetrazione della luce e costituendo pattern zonali di valore variabile per le diverse specie di uccelli. Matasse dense di queste piante possono inibire la presenza di uccelli nuotatori; altri invece possono approfittare di questi tappeti in superficie per camminarci sopra e andare a caccia di prede.

Le piante non radicanti che galleggiano sulla superficie, come la lenticchia d'acqua (*Lemma minor*), assorbono i loro nutrienti direttamente dall'acqua attraverso radici sospese e appaiono generalmente più diffuse in settori marginali dove non c'è movimento di acqua o di vento. Al di sotto del tappeto erboso di lemma non si forma alcuna stratificazione vegetale per la bassa penetrazione della luce in profondità. Le idrofite radicanti nell'acqua che raggiungono la superficie con la parte aerea per fiorire, possiedono tuberi appetibili da diverse specie di uccelli. Le specie più grandi producono notevoli quantità di detriti di lenta demolizione che possono sostenere il peso di piccoli uccelli camminatori come le gallinelle d'acqua (*Gallinula chloropus*), le garzette (*Egretta garzetta*) e gli aironi più piccoli, fino alle sgarze ciuffetto (*Ardeola ralloides*).

Il carattere più evidente nella vegetazione delle velme salmastre è la presenza delle salicornie (*Salicornia* sp.). Sono piante carnose terrestri basse, annuali, che alla fine della stagione estiva possono essere riconosciute per un caratteristico diffuso colore cremisi della parte edule. Grazie alla loro capacità di raggrupparsi densamente, trattengono il limo, provocando un lentissimo innalzamento del suolo. Ma la maggior parte della produttività primaria delle zone umide proviene dalle piante emergenti che vivono nel medium aereo ideale per la fotosintesi e attingono i nutrienti dal suolo in cui hanno accresciuto le radici. Esempi sono le tife (*Typha* sp.), i giunchi (*Juncus* sp.), le carici (*Carex* sp.), gli scirpi (*Scirpus* sp.), i falaschi (*Cladium* sp.), le canne (*Arundo* sp. e *Phragmites* sp.) e, fra le essenze arboree, i salici (*Salix* sp.).

La diversità delle piante tipiche delle zone umide impone variazioni nella struttura dell'habitat sia attraverso le differenti altezze e densità caratteristiche, sia attraverso la composizione di specie prodotta da occupazione, competizione e interazione tra le piante. Nei bacini e nelle zone umide, queste strutture di habitat generano zonazioni spesso concentriche verso le parti più profonde (Spence 1982). Inoltre la configurazione del bacino, la composizione del suolo, la successione vegetazionale, la qualità dell'acqua, la presenza di erbivori e molti altri fattori, alterano queste zone e le dotano spesso di caratteristiche peculiari. Le modalità di distribuzione degli uccelli riflettono le caratteristiche di questi microhabitat e variano dalla presenza di una singola specie a quella di più specie nella corrispondente transizione dalla omogeneità alla eterogeneità della biocenosi (Kaminski & Prince 1984, Beecher 1942)

A causa delle considerazioni anzidette, sembra plausibile che molti parametri relativi agli habitat possano essere correlati significativamente con la presenza o l'abbondanza delle specie di uccelli. Malgrado ciò non esiste ancora un sistema semplice per misurare accuratamente e velocemente tutte queste caratteristiche. La distribuzione orizzontale della vegetazione e altre proprietà strutturali costituiscono i parametri più semplici da identificare grazie alla disponibilità di fotografie aeree o immagini satellitari organizzabili in sistemi informativi geografici. Da questi strati informativi è possibile estrapolare il valore di eterogeneità di pattern spaziali come il rapporto acqua/terra, l'estensione degli ecotoni e altri indici geografici che possono essere correlati con la presenza e la densità delle specie. Viceversa, misurazioni e scomposizioni della struttura verticale della diversità degli habitat sono ancora più complesse. Raramente infatti sono state condotte analisi in zone umide capaci di considerare le caratteristiche espresse della colonna d'acqua e dal suolo, tra cui i gradienti di disponibilità del cibo e di qualità delle acque.

Per quanto riguarda la struttura verticale della vegetazione, malgrado si possa riconoscere l'importanza del numero di strati vegetazionali, nonché dell'altezza, del volume e della forma delle foglie delle piante emergenti, bisogna sottolineare che gli uccelli acquatici devono primariamente confrontarsi con un altro strato, la colonna di acqua sottostante, che gli uccelli terrestri non usano (MacArthur & Wilson 1967). Questa condizione poi è complicata dal fatto che le risorse di cibo possono trovarsi a diverse altezze sopra il livello dell'acqua, altrimenti galleggiare sulla superficie, oppure essere distribuite a ogni livello di profondità della colonna d'acqua, magari sul fondo della zona umida come gli organismi epibentici o sotterrati nel substrato come quelli endobentici. Sono queste le condizioni di estrema varietà e ricchezza delle risorse che inducono le specializzazioni non solo morfologiche e anatomiche, ma anche comportamentali che contraddistinguono gli uccelli che frequentano le zone umide.

Utilizzando i pattern fisici e vegetazionali degli habitat come descrittori degli habitat degli uccelli, è possibile tentare di classificare, descrivere e quantificare le specie, la loro distribuzione e la segregazione all'interno delle comunità che usano una zona umida eterogenea. Le caratteristiche dell'habitat utilizzabili per questo tipo di descrizione si riferiscono a quelle condizioni dell'habitat prossime all'uccello, riconoscibili a vista dall'individuo come utili risorse da sfruttare per uno specifico tempo e funzione. Non è

trascurabile che la struttura della vegetazione e le condizioni di allagamento sono i soli parametri pratici e macroscopici utilizzabili, anche perché riconoscibili facilmente dall'uomo.

Alberi, arbusti, cespugli, canneti e ogni forma vegetale condizionano l'altezza sull'acqua dello strato vegetale, la persistenza dell'oscurità prodotta dalla dimensione delle foglie, dei rami, degli steli e dei tronchi che coprono con densità variabile gli specchi d'acqua. Questi fattori influenzano fortemente l'abilità degli uccelli nel frequentare quel particolare microhabitat, nel camminarci attraverso in cerca di cibo, nella decisione di costruirci il nido, nella prospettiva di ripararsi, etc. Questi parametri sono difficili da accertare e misurare anche perché variano nel tempo e nello spazio ma costituiscono un forte interesse dato che permettono di identificare le associazioni utilizzabili per la predizione della qualità degli habitat e il loro utilizzo (Colwell & Dodd 1997). Nell'applicare alcuni dei concetti e dei principi appena descritti, i biologi tipicamente misurano la qualità dell'habitat per gli uccelli acquatici con una varietà di propositi:

- 1- gli ornitologi studiano la biologia della specie o dei gruppi di specie
- 2- gli ecologi studiano come gli uccelli si distribuiscono nel tempo e nello spazio
- 3- i faunisti stimano l'uso dell'habitat per valutare l'efficacia delle forme di gestione
- 4- i biologi conservazionisti determinano le priorità di protezione per le specie e gli habitat
- 5- i gestori valutano i risultati di progetti di costruzione/rinnovamento di zone umide

I dati relativi alla dimensione e alla densità delle popolazioni di uccelli sono costosi da ottenere e spesso presentano errori nelle stime. Una volta ottenuti, è altrettanto importante avere misure affidabili delle caratteristiche dell'habitat in modo tale da poterle correlare con gli indici di popolazione e fornire le scale di "idoneità di habitat" relative alle diverse specie animali. In questo contesto, l'utilizzo dell'habitat si riferisce alla presenza o al tasso relativo di occupazione da parte di uccelli in un particolare habitat, misurato come presenza/assenza, indice di abbondanza o censimento esaustivo di un'area campione (Hall *et Al.* 1997). Ulteriormente, la semplice presenza di un particolare habitat o cibo non necessariamente significa che questo sia disponibile egualmente a ogni animale. Ad esempio, una ricca provvista di bivalvi potrebbe essere localizzata su fondali in cui l'acqua è troppo profonda per permettere alle anatre di superficie di raggiungerla, contrariamente a quanto avverrebbe

in presenza di anatre tuffatrici che potrebbe nutrirsene investendo una maggiore energia nell'immersione. I pesci che si trattengono vicino alla superficie possono essere più facilmente disponibili ai tuffatori come le sterne (*Sterna* sp.) rispetto ai pesci che nuotano sul fondo, meno vulnerabili in quanto meno visibili.

Quindi stabilire la qualità relativa degli habitat è compito arduo perché un numero elevato di relazioni e nessi di causalità legano le caratteristiche e le risorse dell'habitat con l'uso relativo che ne viene fatto dalla specie o dal gruppo di specie oggetto di studio. (Paquette & Ankney 1996; Swift *et Al.* 1984). Una analisi basata sulla incidenza delle specie di uccelli organizzati in gruppi detti “*guild*”, cioè in accordo con le modalità con cui questi usano le risorse trofiche, sottolinea l'importanza della diversità strutturale dell'habitat, delle risorse di cibo e dei regimi dei livelli di acqua (Short & Burhnam 1982). I dati riferiti al ciclo vitale dei membri di questi raggruppamenti sono analizzabili in una “matrice di *guild*” esaminando una serie di caratteristiche dell'habitat interpretabili come risorse per una data funzione, tipo il nutrirsi. Le celle occupate della matrice permettono una rappresentazione visuale macroscopica della complessità di una comunità animale basata sulla varietà delle risorse. La matrice risultante si mostra anche come un utile approccio per documentare i cambiamenti nel tempo che possono derivare sia da eventi naturali che generati dall'uomo. Risulta infine in un modello concettuale della struttura dell'habitat capace di predire la composizione delle specie sulla base di caratteristiche ambientali (Short 1989).

1.3. Contesto logico-informatico

Le discipline “logica” e “informatica” sono due facce di una stessa medaglia. Entrambe si occupano, pur attraverso percorsi e tradizioni peculiari, di formalizzare, elaborare e comunicare la conoscenza. E’ davvero semplice intuirne l’intima relazione quando si paragona metaforicamente il ruolo della logica nei confronti dell’informatica con quello più antico della matematica nei confronti della fisica. La necessità di rappresentare le informazioni con linguaggi simbolici interpretabili senza ambiguità, espressa per la prima volta con gli assiomi di Peano agli inizi del ‘900, generò il passaggio dall’originaria logica tradizionale aristotelica alla “logica matematica”, intesa come “studio combinatorio del contenuto”. Questa definizione soddisfa sia gli aspetti sintattici, volti a descrivere i diversi linguaggi formali, sia i modelli semantici, per l’attribuzione di un significato a tutti gli enunciati di quei linguaggi. Il passaggio da logica a informatica si fa breve se ricordiamo che qualche tempo fa George Boole fondò l’algebra binaria della sua logica sulla bivalenza falso/vero e oggi il “bit” di un elaboratore risolve proprio quella rappresentazione logica attraverso i valori zero/uno.

Il termine “informatica” presenta contorni poco definiti ed è usato in una varietà di significati spesso associati alla gestione dei dati, scienza del computer, scienza dell’informazione, tecnologia dell’informazione, scienza della comunicazione, interfaccia uomo-macchina e tutte le interazioni sociali inerenti. Astraendo, l’informatica è l’applicazione della scienza dell’informazione nei modi che promuovono la comunicazione, l’incorporazione organizzativa e le interfacce sociali. In un ambiente digitale, uno o più insiemi di bit formano “dati”, i quali necessitano di rimanere collegati alla “informazione sulla Rappresentazione” del particolare oggetto informativo per poterlo dotare di significato e interpretabilità. Arricchire e conservare questi due elementi costituiscono le operazioni fondamentali di cui l’orizzonte tecnologico espresso in seguito si occupa.

Dato che la ecoinformatica si basa su collaborazioni interdisciplinari e sulla gestione di dati diversi, è auspicabile l’utilizzo di strategie che supportano e favoriscono queste attività di collaborazione e gestione dei dati. Ma tutti gli sforzi di integrazione di dati provenienti da fonti eterogenee hanno dovuto fare i conti con la questione della “armonizzazione” dei dati.

L'enorme varietà di modelli di dati adottati dai diversi sistemi informatici ha determinato che questa armonizzazione potesse avvenire solo attraverso la creazione di terminologia comune, dizionari condivisi e software interprete delle diverse fonti, piuttosto che standardizzando i metodi di raccolta e archiviazione già in uso (Toivonen *et Al.* 2005). Questa soluzione permette inoltre alla comunità di continuare a utilizzare hardware, software e strutture dei file familiari.

Nello sviluppo dell'integrazione dei dati avvenuto negli ultimi dieci anni, non è possibile ignorare il ruolo del metalinguaggio *eXtensible Markup Language* (XML, <http://www.w3.org/XML/>). In poche parole, XML ha alimentato il desiderio di integrazione dei dati, offrendo un formato sintattico comune per la condivisione dei dati tra fonti differenti, senza però indirizzare le questioni relative all'integrazione semantica. In questo modo le fonti eterogenee possono anche condividere file XML i cui elementi sono completamente privi di significato al di fuori dell'applicazione (Halevy et al. 2006).

Le caratteristiche principali di XML risiedono nella sua natura nidificata e nella capacità di rappresentare "dati semi-strutturati". Contrariamente ai dati strutturati che sono memorizzati secondo un formato rigido, come le tuple delle tabelle di un "modello entità/relazione", quelli semi-strutturati vengono rappresentati tramite strutture a grafo o ad albero (figura 3) che presentano la potenzialità di variare rispetto a uno schema assegnato.

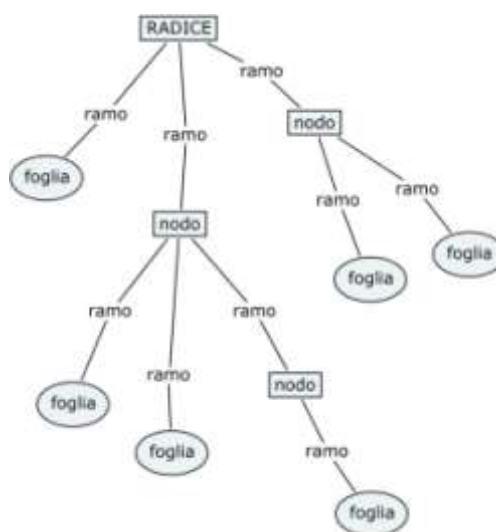


Figura 3

Ad esempio possono mancare alcuni attributi, cioè alcuni rami della struttura ad albero, poiché l'ordine di schema associato permette un alto grado di libertà di aderenza in termini di completezza e battitura (Calvanese *et Al.* 1998). Questo schema inoltre può rimanere implicito per una parte dei dati e richiedere una definizione *a posteriori*. Infine, a causa delle esigenze di gestire l'eterogeneità dei dati, risulta spesso molto più esteso e variabile nel tempo se comparato con gli schemi dei database relazionali (Catarci 1999).

Settare e mantenere i sistemi di integrazione di dati richiede operazioni che manipolano questi schemi e le mappature necessarie tra di essi. Attualmente il metodo per formalizzare e standardizzare il set di concetti che sono essenziali per la descrizione dei dati ecologici si chiama *Ecological Metadata Language* (EML, <http://knb.ecoinformatics.org/software/eml/>). Questo linguaggio, nato all'interno della famiglia di idioletti derivanti dall'XML per un progetto di sviluppo del NCEAS (*National Center for Ecological Analysis and Synthesis*, <http://www.nceas.ucsb.edu/>), si evolve ora grazie alla attività di una comunità *open source* composta da ecologi, manager dell'informazione e sviluppatori di software.

EML è implementato come una serie di tipi di documenti XML che possono essere usati in maniera modulare e flessibile per documentare un insieme di aspetti essenziali delle informazioni ecologiche come i nomi e le definizioni delle variabili, le unità di misura, le date e le localizzazioni, l'identità di colui che ha raccolti i dati, il disegno di campionamento, etc. Attraverso i termini che articolano la sintassi EML è possibile descrivere l'insieme dei metadati che dovrebbero essere inclusi in ogni set di dati ecologici, riducendone l'ambiguità e l'incertezza dei dettagli (Fegraus *et Al.* 2005).

Le categorie di metadati sono organizzate in sette sezioni:

1. informazioni sul dataset
2. informazioni geografiche
3. informazioni temporali
4. informazioni tassonomiche
5. informazioni sui protocolli
6. informazioni sui metodi
7. informazioni sulle tabelle di dati

Le informazioni generali sul dataset sono composte da campi utilizzabili dai motori di ricerca che interrogano vari cataloghi informatici in rete per riconoscerne i propositi. Questi metadati descrivono, oltre al titolo e all'*abstract*, informazioni su eventuali restrizioni nell'utilizzo dei dati, sui contatti con il "creatore" e con il "curatore" degli archivi, sui finanziamenti e patrocini, sulle fonti bibliografiche di pertinenza.

Le informazioni geografiche si riferiscono ai metadati spaziali che caratterizzano i dati, come i campi che contengono informazioni su dove il progetto di ricerca ha avuto luogo, su dove sono avvenuti i diversi campionamenti, su ogni riferimento a un contesto geografico dei dati. EML prevede l'opzione di georeferenziare con i valori di latitudine e longitudine.

Le informazioni temporali sono formate da metadati su quando i dati sono stati raccolti, sia indicando un intervallo di date, ad esempio "da marzo 2007 a ottobre 2007", sia una serie di date specifiche, cioè "28 maggio 2007", "17 luglio 2007". A ogni data si può associare l'ora e informazioni su eventuali buchi di campionamento di qualche variabile.

Le informazioni tassonomiche sono formalizzate con metadati che descrivono le specie coinvolte nell'indagine. E' prevista la descrizione del sistema tassonomico utilizzato per identificare le specie e le diverse categorie quali famiglia, genere, specie, etc.

Le sezioni dedicate alle informazioni sui protocolli e sui metodi sono strettamente correlate e definiscono le situazioni in cui la raccolta dati che si descrive nel file EML partecipa a uno schema progettuale più ampio. In questi casi infatti le stazioni federate dove vengono svolte le attività e il personale coinvolto possono presentare una implementazione parziale dei protocolli standardizzati centralmente perché le condizioni locali lo impongono. Quindi le informazioni sui protocolli definiscono i metadati del disegno di campionamento standard che tutte le stazioni dovrebbero adottare, mentre le informazioni sui metodi specificano le modalità effettive con cui i dati vengono raccolti nella stazione riferita al dataset. Le informazioni sui metodi possono riguardare anche metadati sulle apparecchiature utilizzate per campionare, come pure le strategie di controllo della qualità delle misurazioni e della registrazione dei dati.

Le informazioni sulle tabelle di dati esprimono i metadati sul modello fisico delle tabelle in cui i dati sono archiviati, tra cui il nome del file, la condizione di sensibilità alle maiuscole dei valori, il numero di record e la struttura della tabella. Questa ultima categoria di

informazioni è ulteriormente specificabile con metadati riferiti alle diverse colonne della tabella, cioè agli attributi dei record o campi. Per ogni campo si può definirne il nome breve, l'etichetta che completa il nome breve, la definizione estesa di ciò che la colonna rappresenta, le unità di misura (metro, grammo, etc.) e il tipo di dati (numerico, testo, etc.). Inoltre è possibile registrare quale valore è stato utilizzato per definire il valore nullo, il livello di precisione numerica, la descrizione dei codici, per esempio “AS = *aerial survey*”, e il dominio di validità per i valori presenti nella colonna. La specifica completa di EML è disponibile all'indirizzo <http://knb.ecoinformatics.org/software/eml/eml-2.0.1/index.html>.

Esistono due possibili approcci per costruire la mappatura di un progetto di integrazione: il primo viene definito *local-as-view* (LAV), per il quale una fonte di informazione è descritta come l'espressione di una vista sullo schema mediato, il secondo invece è *global-as-view* (GAV), per il quale lo schema mediato è descritto come una vista sulle fonti di dati (Lenzerini 2002, Levy 2000). La nascita dell'approccio LAV (*local-as-view*) si basa sulla combinazione del potere espressivo delle “*Description logics*”, una branca della rappresentazione della conoscenza, con i linguaggi di interrogazione dei database e fu ispirata dal fatto che le fonti di dati necessitano di essere rappresentate dichiarativamente (Levy & Rousset 1998, Beeri *et Al.* 1997, Borgida *et Al.* 1989).

Le *Description Logics* offrono meccanismi più flessibili per rappresentare uno schema mediato e per l'ottimizzazione delle interrogazioni semantiche che servono in questi sistemi (Calvanese *et Al.* 2006). Da subito infatti venne mostrato che le *Description logics* rappresentavano il candidato adatto per la descrizione delle relazioni tra fonti eterogenee di dati, nella loro qualità di esprimere il compromesso più utile tra la potenza espressiva del linguaggio e la facilità di implementazione offerta (Catarci & Lenzerini 1993). Ma per capire a fondo questo equilibrio mediato dobbiamo premettere quali strade abbia percorso storicamente la ricerca di metodi per descrivere la conoscenza tali da essere utilizzati per costruire applicazioni intelligenti.

Uno dei primi tentativi di rappresentazione strutturata della conoscenza fu intrapreso da Quillian nel 1967 attraverso l'implementazione delle “reti semantiche” che verranno poi rielaborate profondamente da Brachman e Levesque nel 1985.

Una rete semantica rappresenta la conoscenza attraverso un set di nodi che sono connessi l'un l'altro da connessioni di diversa tipologia. Ogni nodo può corrispondere a un concetto o a un oggetto del mondo reale e le connessioni tra nodi rappresentano le relazioni che esistono tra i concetti e/o gli oggetti. Malgrado il loro nome, le reti semantiche non esprimono una semantica precisa (Brachman 1979, Woods 1975). Ciò è dovuto al fatto che sia gli archi che i nodi possono assumere differenti significati. Infatti un nodo può rappresentare un concetto a cui è associato un set di attributi che descrivono tutti gli oggetti del mondo reale che sono istanze del concetto stesso. Similmente, una connessione tra due nodi può contribuire sia alla definizione di un concetto, sia rappresentare una relazione tra due oggetti e costituire quindi una asserzione sulla realtà. Oggi sappiamo che i nodi e gli archi che si riferiscono a concetti fanno parte della conoscenza *intensionale*, mentre quelli riferibili agli oggetti compongono la conoscenza *estensionale*. Questa distinzione è di fatto fondamentale e costituisce la base di tutti i moderni formalismi di rappresentazione della conoscenza.

Un altro importante sviluppo nella rappresentazione della conoscenza fu introdotto da Minsky nel 1975 col nome di “*frames*”. Un *frame* (o modulo) rappresenta un concetto (o classe) con tutti gli attributi associati che compongono le proprietà comuni tra tutte le istanze della classe. Un attributo viene specificato attraverso la definizione di uno “*slot*”, il quale contiene tutte le informazioni rilevanti per quell'attributo: le restrizioni sul numero dei valori possibili (o “*slot filler*”), un valore di default, quale valore scegliere in assenza di informazioni specifiche, le procedure per calcolare il valore quando è stato richiesto ma non è ancora disponibile, le procedure che vengono attivate nel caso di modifica o cancellazione del valore. Inoltre ogni slot possiede un dominio dei propri *filler*, il quale può essere un dominio concreto, ad esempio una stringa o valori interi, oppure consistere in un altro *frame* specificato attraverso il suo identificatore. In aggiunta è possibile specificare che un *frame* è *sub-frame* di un altro e che quindi ne eredita tutte le proprietà, cioè tutti gli *slot*.

Come già notava Hayes (Hayes 1979), molti degli aspetti delle reti semantiche e dei *frames* ammettevano una traslazione verso la logica del primo ordine. Il primo sistema che fu costruito su questa corrispondenza si chiama KL-ONE (Brachman & Schmolze 1985) al quale, per l'alta espressività del linguaggio utilizzato, non fu possibile associare alcuna procedura di ragionamento. Successivamente, sviluppando ulteriormente KL-ONE, si rese

disponibile una varietà di sistemi di rappresentazione della conoscenza chiamati “sistemi terminologici” (o “basati su concetti”), tutti dotati della componente di inferenza (cioè del *reasoner*). Tra i più importanti vanno citati KANDOR (Patel-Schneider 1984), KRYPTON (Brachman *et Al.* 1985), NIKL (Kaczmarek *et Al.* 1986), LOOM (MacGregor & Bates 1987), CLASSIC (Borgida *et Al.* 1989), BACK (Quantz & Kindermann 1990) e KRIS (Baader & Hollunder 1991). Ognuno di questi sistemi terminologici si basa su una specifica *Description Logic*.

Gli elementi base di ogni *Description Logic* sono i concetti e i ruoli. I primi costituiscono le classi, i secondi sono usati per specificare le proprietà delle classi. Ogni concetto è interpretabile come un sottoinsieme del dominio di interpretazione (l'insieme delle istanze del concetto), mentre ogni ruolo è interpretabile come la relazione binaria su quel dominio. Una *Description Logic* è caratterizzata da un insieme di “costruttori” che possono essere usati per costruire concetti complessi ed espressioni di ruoli partendo da un insieme iniziale di concetti e ruoli. I più comuni costruttori di concetti includono i connettivi booleani e i costruttori che permettono di quantificare gli elementi connessi attraverso un ruolo. Per esempio, l'espressione del concetto $Uccelli \sqcap \neg Erbivori \sqcap \forall predare.Molluschi$ denota l'insieme di uccelli che non sono erbivori e che predano solo molluschi.

Il simbolo “ \sqcap ” denota congiunzione ed è interpretato come l'insieme di intersezione, mentre “ \neg ” indica negazione e viene interpretato come il complemento rispetto al dominio di riferimento. L'espressione $\forall predare.Molluschi$ denota l'insieme di oggetti che sono connessi attraverso il ruolo *predare* a tutte le istanze di *Molluschi*. Un'espressione della forma “ $\forall \mathcal{R}.C$ ” viene chiamata “quantificazione universale sui ruoli” ed è usata per specificare le proprietà di tutti gli oggetti connessi dal significato di un ruolo. L'espressione $\exists predare.Molluschi$ denota tutti gli oggetti che sono connessi attraverso il ruolo *predare* ad almeno una istanza di *Molluschi*. Un'espressione della forma “ $\exists \mathcal{R}.C$ ” viene chiamata “quantificazione esistenziale sui ruoli” e può apparire anche nella forma non qualificata “ $\exists \mathcal{R}$ ”. In questo caso viene richiesto solo che almeno un oggetto connesso attraverso il ruolo \mathcal{R} esista, senza specifiche di altre proprietà che quell'oggetto deve soddisfare.

Un costruttore addizionale che si è dimostrato particolarmente utile è quello chiamato “restrizione cardinale” (*number restriction*). Nelle *Description Logics*, le restrizioni cardinali

vengono indicate come $(\geq n \mathcal{R})$ e $(\leq n \mathcal{R})$, dove n è un intero positivo. Questi descrittori consentono di porre restrizioni sul numero di volte che un oggetto può apparire come componente in un ruolo \mathcal{R} . Ad esempio, gli uccelli capaci di predare più di 50 molluschi può essere rappresentato attraverso l'espressione $Uccelli \sqcap \forall predare.Molluschi \sqcap (\geq 50 predare)$. Un notevole incremento nell'espressività è stato ottenuto proponendo l'utilizzo dell'inverso di un ruolo, per il quale si indica ancora più correttamente come $predare^-$, anziché $esserePredatoDa$, il ruolo inverso di $predare$.

Utilizzando una *Description Logic* è possibile costruire una “Base di Conoscenza Terminologica”, la quale consiste tipicamente di due componenti separate, una “intensionale” e una “estensionale”. La componente estensionale contiene le asserzioni circa gli oggetti del mondo reale, specificando quali siano le istanze dei concetti e a quali altri oggetti siano connessi attraverso i ruoli. La componente intensionale, invece, contiene asserzioni generali circa i concetti usati nella parte estensionale e specifica le loro proprietà e le loro relazioni mutuali. Essa è costituita da un insieme di asserzioni che possono essere distinte in due tipologie: “primitive” che hanno forma $\mathcal{C} \sqsubseteq \mathcal{E}$, utilizzate per specificare che l'espressione di un concetto \mathcal{E} è una condizione necessaria perché un oggetto del dominio sia un'istanza del concetto \mathcal{C} . La seconda tipologia è quella delle “definizioni” che presentano la forma $\mathcal{C} \equiv \mathcal{E}$ e permettono di specificare con \mathcal{E} una condizione necessaria e sufficiente per un oggetto di essere un'istanza del concetto \mathcal{C} .

I compiti basilari di inferenza che possono essere condotti sulla componente intensionale della base di conoscenza sono:

1. *consistenza della base di conoscenza*: la base di conoscenza ammette un modello? Questa domanda rimane plausibile anche quando la parte estensionale è vuota, dato che la parte intensionale da sola può già contenere contraddizioni implicite.
2. *consistenza dei concetti*: data una base di conoscenza e un concetto, esiste un modello della base di conoscenza per il quale il concetto possiede un'estensione non vuota? Questa domanda svolge un ruolo fondamentale durante la fase di disegno della base di conoscenza, per eliminare le parti prive di significato.
3. *classificazione dei concetti*: data una base di conoscenza e due concetti, l'estensione di un concetto corrisponde alla estensione dell'altro in ogni modello della base di

conoscenza? La sussunzione (*subsumption*) ci permette di determinare le dipendenze implicite nella base di conoscenza. Essa è la base della classificazione, tipica forma di ragionamento compiuta nei sistemi terminologici che consiste nell'organizzare la tassonomia di tutti i concetti della base di conoscenza.

Considerando la combinazione sia della parte intensionale che di quella estensionale, si rendono disponibili molti altri ragionamenti, come quelli che controllano se la parte intensionale e quella estensionale sono consistenti l'un l'altra, quelli che verificano se un oggetto è necessariamente un'istanza del concetto in tutti i modelli della base di dati (*instance checking*) oppure quelli che trovano tutte le istanze di un concetto in una base di conoscenza (*retrieval*).

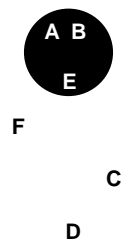
Descrizioni dettagliate per le procedure di ragionamento sulla componente estensionale sono discusse in Baader *et Al.* 1991, Levesque 1986, Schaerf 1994, Donini *et Al.* 1994, Schaerf 1994. Per approfondire il ragionamento intensionale si legga Calvanese 1996.

Gli ecologi spesso usano il linguaggio naturale vago e mal definito per descrivere le loro conoscenze (Slaski 1992, Slaski & Kandzia 1991) e, abitualmente, fanno uso della probabilità per esprimere l'incertezza sul verificarsi di un fenomeno e non l'incertezza intrinseca nel fenomeno stesso (Kosko 1994). Ma l'atteggiamento che considera l'incertezza come qualcosa di indesiderabile e da evitare (Chavas 2000) induce all'errore di considerare il mondo reale artificialmente preciso (Enea e Salemi 2001). La teoria degli insiemi *fuzzy* sviluppata da Lofti Zadeh (Zadeh 1965) è basata su una estensione del significato classico del termine "insieme" e rivoluziona il concetto semantico di "verità". Mentre nella logica booleana un dato oggetto può appartenere all'insieme oppure no (cioè l'appartenenza è bivalente: "tutto o nulla"), la nozione *fuzzy* prevede che l'oggetto possa appartenere parzialmente all'insieme considerato. (Marchini 2004). Analogamente, nella logica classica, una affermazione viene valutata come "vera" o "falsa", mentre nella logica *fuzzy* le viene assegnato un valore di verità parziale.

Quindi la teoria degli insiemi *fuzzy* affronta un sottoinsieme di un dato universo, dove la transizione tra la completa appartenenza e l'estraneità appare graduale. Infatti i confini degli insiemi *fuzzy* non sono netti (Figura 4) Tradizionalmente, un grado di appartenenza

uguale a 1 è assegnato a quegli oggetti dell'universo che appartengono all'insieme, mentre 0 è assegnato agli oggetti che sono esclusi dall'insieme. Nella teoria tradizionale, gli insiemi sono definiti come una collezione di oggetti aventi la stessa proprietà, ad esempio “gli abitanti di Trieste”. La proprietà “gli abitanti alti di Trieste” però non costituisce un insieme nel senso usuale, infatti questa proprietà non offre un criterio di appartenenza definito precisamente. Intuitivamente un'insieme *fuzzy* è una collezione di oggetti che ammettono la possibilità di appartenere all'insieme anche parzialmente.

insieme booleano



insieme *fuzzy*

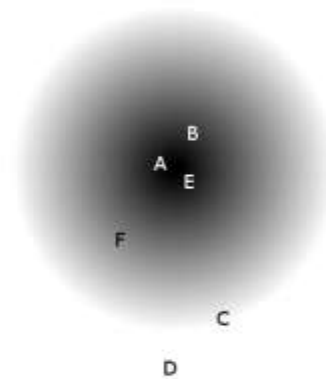


Figura 4

La teoria degli insiemi *fuzzy* è cioè basata su una estensione delle regole della logica convenzionale poiché trascende due dei più importanti principi aristotelici. Dati due insiemi \mathcal{A} e $\neg\mathcal{A}$, “il principio di non contraddizione” sancisce che se un elemento appartiene all'insieme \mathcal{A} , contemporaneamente non può appartenere all'insieme $\neg\mathcal{A}$, formalmente $\mathcal{A} \cap \neg\mathcal{A} = \emptyset$. Il “principio del terzo escluso” invece determina che l'unione dell'insieme \mathcal{A} con l'insieme $\neg\mathcal{A}$ coincide con l'universo del discorso, cioè $\mathcal{A} \cup \neg\mathcal{A} = U$. Ne deriva che un qualunque elemento che non appartiene all'insieme \mathcal{A} deve appartenere all'insieme $\neg\mathcal{A}$. L'applicazione di questi principi nelle sensazioni, nei fenomeni e nei comportamenti del mondo reale produce anomalie e vari paradossi che la teoria degli insiemi *fuzzy* supera attraverso il calcolo dei valori delle conclusioni delle regole, anche qualora le premesse di quelle stesse regole non siano completamente soddisfatte.

Un insieme *fuzzy* \mathcal{A} in un dato universo è caratterizzato da una funzione $\mu_{\mathcal{A}}(x)$ chiamata “il grado di appartenenza di x in \mathcal{A} ”. Dovremmo allora assumere che i valori di $\mu_{\mathcal{A}}(x)$ sono elementi dell'intervallo $[0,1]$, con i gradi 0 e 1 che rappresentano rispettivamente l'estraneità e la completa appartenenza.

$\mu_{\mathcal{A}}(x)$ è la “funzione *membership*” di \mathcal{A} ed è $0 \leq \mu_{\mathcal{A}}(x) \leq 1$. Inoltre l'unione dell'insieme *fuzzy* $\mu_{\mathcal{A}}(x)$ con l'insieme *fuzzy* $\mu_{\mathcal{B}}(x)$ è uguale alla funzione $MAX(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))$ mentre l'intersezione è uguale alla funzione $MIN(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x))$. Ecco che, ritornando ai principi di non contraddizione e del terzo escluso, si ottiene, per la logica *fuzzy*, che $\mu_{\mathcal{A}}(x) \cap \neg\mu_{\mathcal{A}}(x) = MIN(\mu_{\mathcal{A}}(x), \neg\mu_{\mathcal{A}}(x)) \neq \emptyset$ e $\mu_{\mathcal{A}}(x) \cup \neg\mu_{\mathcal{A}}(x) = MAX(\mu_{\mathcal{A}}(x), \mu_{\mathcal{B}}(x)) \neq U$ (Figura 5).

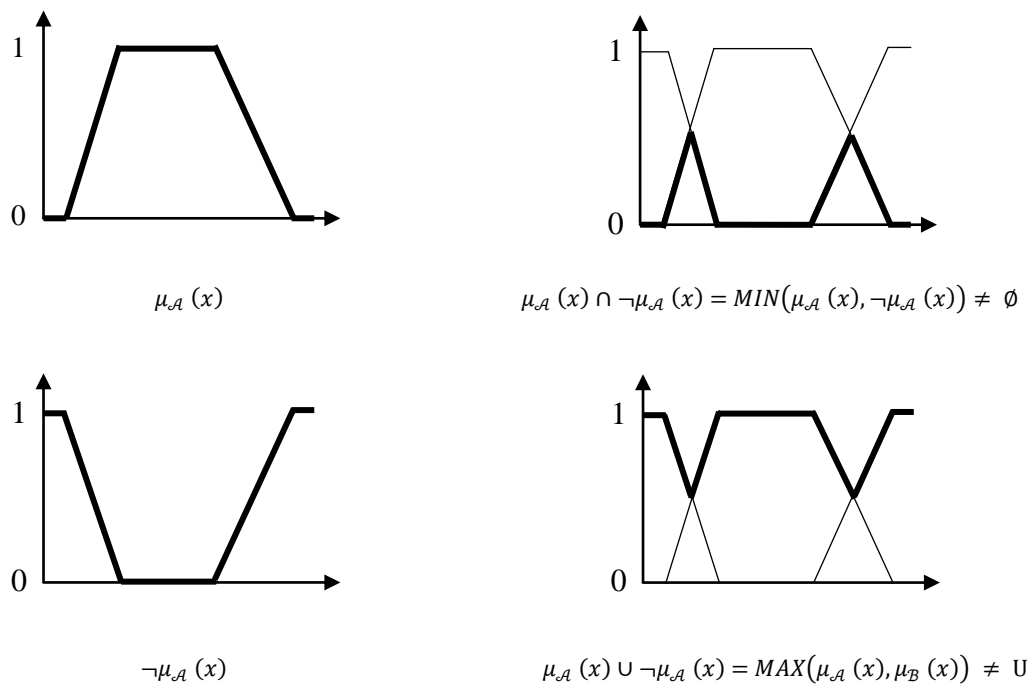


Figura 5

I passaggi fondamentali del modello *fuzzy* sono riassumibili in tre operazioni: fuzzificazione, inferenza e defuzzificazione. La fase di “fuzzificazione” trasforma il valore numerico della variabile di input in un valore di appartenenza all'insieme *fuzzy*, tramite

l'opportuna funzione di appartenenza (funzione *membership*), per valori compresi nell'intervallo [0,1]. Una proprietà importante è la sovrapposizione delle funzioni di appartenenza relative allo stesso parametro: tanto più estesa è la zona di sovrapposizione, tanta più incertezza viene compresa nel modello.

L'inferenza avviene applicando le regole di combinazione tra gli insiemi $fuzzy$. Le regole consistono di semplici espressioni linguistiche che sono convertite formalmente in regole "IF-THEN". Così la descrizione linguistica della relazione tra l'input e l'output di un modello $fuzzy$ può essere rappresentata da un insieme di relazioni qualitative e linguistiche interpretate come asserzioni e regole. Queste relazioni presentano il vantaggio di essere più vicine al pensiero umano, più comprensibili, più facili da modificare e più robuste.

L'output $fuzzy$ è anch'esso un valore di appartenenza nell'intervallo [0,1] che può essere "defuzzificato" in un numero reale detto "*crisp*" utilizzabile in approcci non $fuzzy$ (Silvert 2000).

Da un punto di vista più prettamente implementativo, il processo di creazione di un modello $fuzzy$ basato su conoscenza prevede i seguenti passi:

1. determinazione di una struttura di modello
2. creazione di una base di conoscenza $fuzzy$
3. scelta di un metodo di inferenza $fuzzy$
4. calibrazione e validazione

Il primo passo consiste nella determinazione della struttura del modello, cioè delle variabili di input e di output, del numero di sotto-modelli, delle connessioni tra sotto-modelli, etc. Il passo successivo è la creazione di una base di conoscenza $fuzzy$ che costituisce la parte preponderante del modello $fuzzy$. Essa contiene un set di regole linguistiche e le definizioni degli insiemi $fuzzy$ che corrispondono ai termini linguistici in quelle regole, come "basso", "alto", etc. Il set di regole linguistiche dovrebbe essere completo e così provvedere a fornire una risposta corretta per ogni possibile valore di input. Per soddisfare questa condizione la somma di tutti i valori di input presenti nelle premesse delle regole, vale a dire l'unione degli insiemi $fuzzy$, deve "coprire" l'intero intervallo dei valori della variabile di input. Se questa condizione non è soddisfatta, è necessario aumentare il numero delle regole linguistiche oppure incrementare il cosiddetto grado di "*fuzziness*" degli insiemi $fuzzy$.

La scelta del metodo di inferenza *fuzzy* può dirigersi verso il “modello Mamdani” oppure verso il “modello Sugeno”. Nel primo caso le funzioni *membership* dell’output sono insiemi *fuzzy* che richiedono un processo di defuzzificazione. E’ possibile però, e in molti casi anche più efficiente, utilizzare come output un singolo picco (*singleton*), anziché un insieme *fuzzy*. Questa situazione semplifica le operazioni di defuzzificazione perché utilizza la media pesata di pochi punti al posto di risolvere la funzione bi-dimensionale che identifica il centroide di un insieme *fuzzy*. Il metodo Sugeno adotta questa soluzione ed è preferibile quando le funzioni *membership* di output sono lineari e costanti.

Il processo di calibrazione riguarda qualunque cambiamento all’interno del database, cioè ogni cambiamento delle definizioni dell’insieme *fuzzy*, per adattare meglio le osservazioni con la conoscenza degli esperti. Il processo di validazione, vale a dire ulteriori test di simulazione per un più ampio intervallo di dati indipendenti, fornisce una risposta alla domanda di quanto il modello sia adatto a descrivere la realtà (Salski 1999a).

2. Metodi

2.1. Protocolli di raccolta dei dati

Le attività di censimento previste dal progetto ANSER sono distinte in gruppi a seconda della tecnica di osservazione utilizzata e del protocollo di raccolta dei dati.

Di seguito sono indicate le tre metodologie impiegate per il monitoraggio dell'avifauna:

- 1- censimenti quantitativi (*Daylight Time Count* e *Roost Count*, DTC+RC)
- 2- censimenti da punti fissi (*Point Transect Count*, PTC)
- 3- censimenti da aereo (*Aerial Survey*, AS)

I dati di censimento sono raccolti attraverso la compilazione su campo di schede appositamente strutturate per rispettare i protocolli standard utilizzati durante i censimenti degli uccelli acquatici svernanti svolti nell'ambito dell'*International Waterfowl Census* (IWC, <http://www.wetlands.org/>). Successivamente i dati di queste schede vengono inseriti in tre archivi elettronici la cui struttura è stata definita dal Servizio Tutela Ambiente Naturale e Fauna della Regione Friuli Venezia Giulia (Azione progettuale AT1). Il formato file utilizzato è Microsoft Access®.

La lista degli animali da censire è composta dalle famiglie strettamente legate alle zone umide e da alcune specie di uccelli rapaci che frequentano l'ecosistema della laguna di Grado e Marano e la foce Isonzo, nell'area compresa tra i fiumi Tagliamento e Timavo.

Il protocollo DTC+RC comprende i censimenti durante l'alta marea (DTC) e quelli dei roost delle famiglie *Ardeidae*, *Phalacrocoracidae*, *Accipitridae* e *Strigidae*, sia in alta che in bassa marea. Si svolgono due volte al mese e prevedono la percorrenza di diversi transetti che deve essere completata all'interno delle quattro ore a cavallo dell'istante di picco dell'alta marea, evitando però di effettuare conteggi nella prima e nell'ultima ora di luce.

Il protocollo *PTC* è calibrato per effettuare censimenti da posti fissi della durata di dieci minuti ciascuno con la frequenza di una volta al mese, sempre in condizioni di bassa marea. I conteggi di presenza delle diverse specie vengono separate in due fasce concentriche attorno al punto. Si registrano i contingenti entro i 500 metri dall'operatore suddividendoli a seconda delle attività facilmente interpretabili che gli animali compiono, come il riposo o l'alimentazione. Ugualmente accade per gli individui presenti nella fascia tra 500 e 1000 metri dal punto di osservazione.

I censimenti aerei secondo il protocollo *AS* sono compiuti percorrendo transetti lineari la cui localizzazione varia in funzione dei censimenti a terra, allo scopo di integrarli soprattutto nelle aree più ostiche da raggiungere. I conteggi durante il sorvolo hanno valori variabili a seconda della specie in esame per il differente grado di difficoltà nel riconoscerla.

L'intera area è suddivisa in celle georeferenziate con risoluzione di 1 km e si riferiscono alla griglia definita dalla NATO nel *Military Grid Reference System* (MGRS). Ogni coordinata MGRS utilizzata in ANSER è definita da una stringa alfanumerica costituita da due lettere che designano una macrocella quadrata con risoluzione di 100 km e da un numero a quattro cifre che definisce con la metà sinistra la coordinata Est e con la metà destra la coordinata Nord, riferite entrambe alla macrocella indicata.

2.2. Tecnologie di rappresentazione dei dati

I formalismi utilizzati per raggiungere gli obiettivi prefissati sono:

- 1- Ecological Metadata Language (EML 2.0, <http://knb.ecoinformatics.org/software/eml/>)
- 2- Web Ontology Language, nella sua accezione OWL-DL (W3C Recommendation 10 Feb 2004, <http://www.w3.org/2004/OWL/>).

Inoltre i software utilizzati sono:

- 1- ArcGIS 9.2 (<http://www.esri.com/index.html>)
- 2- CmapTools Knowledge Modeling Kit 4.12 (<http://cmap.ihmc.us/>)
- 3- FaCT++ Reasoner 1.1.9 (<http://www.cs.man.ac.uk/~horrocks/FaCT/>)
- 4- *Fuzzy* Logic Toolbox™ 2.2.7 del framework MatLab 7.0 (<http://www.mathworks.com/products/fuzzylogic/>)
- 5- GeoTrans 2.4.1 (<http://earth-info.nga.mil/GandG/geotrans/>)
- 6- Morpho XML Metadata Client 1.6.1 (<http://knb.ecoinformatics.org/morphoportal.jsp#download>)
- 7- Protégé-OWL 3.4 beta (<http://protege.stanford.edu/overview/protege-owl.html>)
- 8- Zotero 1.0.3 (<http://www.zotero.org/>)

Al fine di mantenere una elevata consistenza di notazione, si è ricorsi alla convenzione di regole utilizzate per nominare gli oggetti dei modelli conosciuta con il nome di “*CamelBack notation*”. Essa prevede, per i nomi composti, l’assenza di spazi tra i termini e l’iniziale maiuscola di ogni parola anche all’interno del nome. Ad esempio, l’entità “classe di distanza” viene nominata ClasseDiDistanza. Fanno eccezione i nomi utilizzati per le proprietà dell’ontologia che hanno la parola iniziale completamente minuscola e utilizzano terminologia anglosassone del predicato (ad esempio, hasTatticaDiAlimentazione e il suo inverso isTatticaDiAlimentazioneOf) . I nomi degli oggetti ereditati dai modelli già in uso non sono stati sottoposti a queste regole e rimangono inalterati per garantire la riconoscibilità ed evitare così l’insorgere di interpretazioni errate o confuse.

3. Risultati

3.1. Integrazione sintattica

3.1.1. Analisi dei modelli attuali

I database nei quali confluiscono tutti i dati delle schede di campo compilate durante le uscite su campo, sono tre e corrispondono alle tre differenti metodologie di censimento di avifauna previste dal progetto “ANSER”. Gli oggetti che costituiscono ogni database sono aggiornati in tre file mdb separati e gestibili tramite Microsoft Access[®]. I dati derivanti dai censimenti effettuati tramite i protocolli *Daylight Time Count* e *Roost Count* sono accodati al file *anser_dtc_rc.mdb*, quelli *Point Transect Count* nel file *anser_ptc.mdb* e quelli raccolti durante i voli aerei nel file *anser_os.mdb*.

La struttura logica dei tre modelli entità/relazioni è analoga per quanto riguarda la gran parte delle relazioni in atto ma distinta negli attributi delle diverse entità. Come si evince dalla figura 6, in tutti e tre i database esiste una relazione 1: \mathcal{N} tra l'entità *DataRecording* e l'entità *Content*. Questa relazione è definita tra l'attributo *DataRecording.id* che costituisce la chiave primaria e l'attributo *Content.idDataRecording* che costituisce la chiave esterna.

L'entità *DataRecording* contiene le tuple che si riferiscono alle informazioni generali legate alla situazione corrente durante l'uscita su campo per quanto riguarda l'identificazione della zona umida o parti consistenti di essa. Altre informazioni definibili sono riferite agli osservatori che compiono i censimenti e alla presenza o meno di disturbo arrecato all'avifauna da parte di differenti cause di origine antropica. Un altro corposo gruppo di informazioni è modulabile, tramite la scelta da un elenco di valori consentiti, attraverso relazioni 1: \mathcal{N} con altrettante “*look-up tables*” relative a descrizioni climatologiche, idrologiche

e metodologiche. Tutti gli attributi della tabella DataRecording sono definiti all'interno di un intervallo di tempo connesso alla data di uscita e, qualora uno o più di essi appaiano mutati, è consentito variarne i valori inserendo una nuova tupla riferita alla stessa area per un intervallo di tempo successivo.

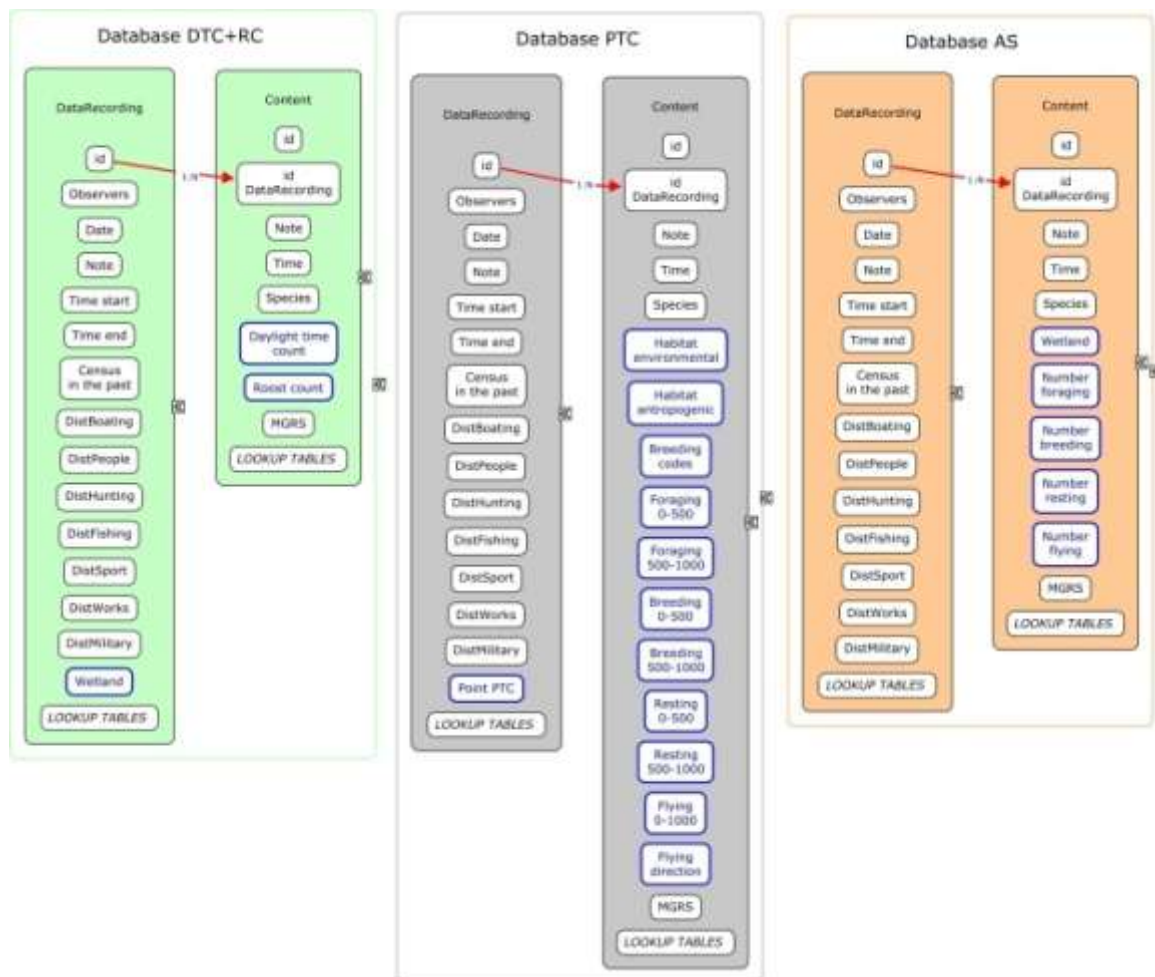


Figura 6

L'entità Content è delegata alla raccolta dei dati quantitativi di presenza delle diverse specie in un particolare punto della zona umida e in un preciso orario. Questi dati quantitativi possono essere suddivisi per la metodologia, per la distanza e per l'attività mostrata dagli uccelli. Grazie alla relazione DataRecording.id : Content.idDataRecording prima descritta, varie tuple di Content possono presentare la stessa tupla di DataRecording, consentendo così di

inserire i molti dati quantitativi delle diverse specie senza dover definire ogni volta le condizioni generali che caratterizzano tutta o parte dell'uscita su campo.

Per documentare con accuratezza questi modelli di dati è necessario considerare l'assortimento fine degli attributi, organizzandoli in due insiemi separati a seconda che siano ubiquitari o meno nelle corrispondenti entità dei tre database. In questo modo si può notare come tutte le entità `DataRecording` presenti nei database siano identiche, tranne che per l'aggiunta della definizione di zona umida (`DataRecording.Wetland`) in DTC+RC e della definizione dei punti di censimento (`DataRecording.Point_PTC`) necessari per i dati PTC. Alquanto diversificata, invece, appare l'entità `Content` nei tre database, dato che la peculiarità dei gruppi di attributi è legata al differente modello di dati quantitativi che devono essere raccolti seguendo i protocolli operativi. Infatti gli attributi comuni nei tre database si "accontentano" di definire quando (`Content.Time`), dove (`Content.MGRS`) e quale specie viene contata (`Content.Species`).

Formalmente, DTC+RC quantifica le presenze con gli attributi `Content.Daylight_Time_count` e `Content.Roost_count`, destinati alle due categorie di censimento. PTC, al contrario, permette di distinguere l'attività svolta dagli animali in due classi di distanze concentriche intorno al punto di osservazione. Così gli attributi disponibili si riferiscono al conteggio di animali in un raggio di 500 metri oppure tra 500 e 1000 metri, che compiono le attività legate al comportamento di alimentazione (`Content.Foraging0-500` e `Content.Foraging500-1000`), di riproduzione (`Content.Breeding0-500` e `Content.Breeding500-1000`) e di riposo (`Content.Resting0-500` e `Content.Resting500-1000`), mentre il volo esclusivamente entro 1000 metri (`Content.Flying0-1000`). A compendio di questi attributi per l'inserimento dei dati quantitativi, sono presenti altri quattro attributi, di cui due per la definizione della categoria di habitat naturale (`Content.Habitat_environmental`) e antropogenico (`Content.Habitat_antropogenic`), uno per la codifica della fase riproduttiva (`Content.Breeding_codes`), uno per la direzione di volo (`Content.Flying_direction`).

AS mostra un'entità `Content` caratterizzata da una ricchezza di attributi intermedia rispetto alle due precedenti. Infatti prevede la possibilità di quantificare le presenze di animali distinti sulla base dell'attività ma senza alcuna indicazione di distanza. La scelta di questo modello è dettata dal fatto che appare intuitivamente inutile una stima di distanze tra il punto di osservazione e gli animali quando l'osservatore stesso compie movimenti di volo. Da

ultimo, il database AS identifica la zona umida censita a questo livello tramite Content.Wetland perché il percorso compiuto con l'aeroplano interessa un'area più grande di una singola zona umida.

Se spostiamo adesso l'osservazione dei tre modelli verso una prospettiva più distante e includiamo anche l'insieme completo delle relazioni verso le *look-up tables*, osserviamo che la complessità dei tre database si fa elevata (Figura 7). Infatti pressoché tutte queste entità che offrono elenchi di descrizioni, nomi e valori consentiti per la definizione delle tuple di DataRecording e di Content, devono essere triplicate per poter partecipare alle relazioni 1:N in cui sono coinvolte. Questo assetto è determinato dalla necessità di distribuire un modello logico alquanto coerente in tre modelli fisici separati, cioè in tre file indipendenti, con conseguente "moltiplicazione" per tre di ogni operazione legata alla sempre possibile variazione di una parte del modello.

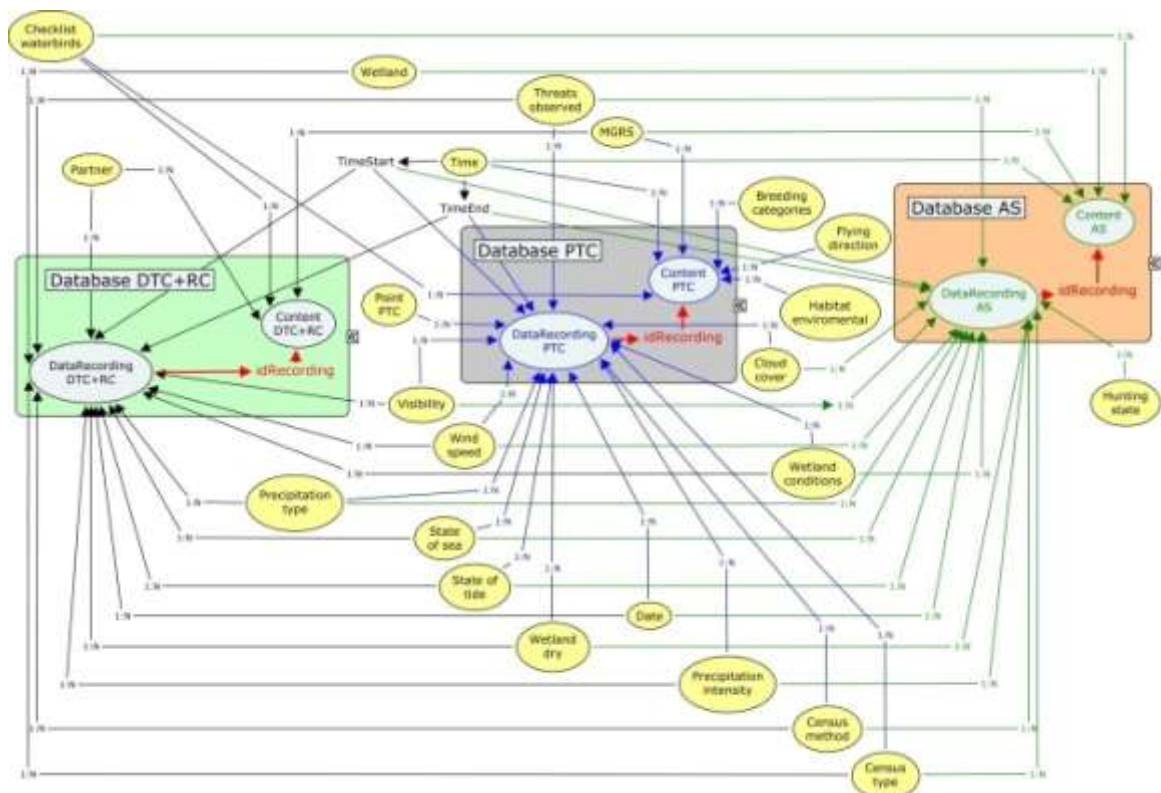


Figura 7

Inoltre la struttura delle relazioni in essere tra le entità più importanti e quelle di *look-up* appare sbilanciata verso uno sviluppo “in orizzontale”, con la conseguente presenza di tabelle formate da numerosi campi che possono assumere valori svincolati e incongruenti tra loro. Una maggiore verticalità delle relazioni tra le diverse entità permette infatti di utilizzare i vincoli di integrità referenziale e di evitare, ad esempio, l’inserimento di una tupla di DataRecording in cui le precipitazioni siano definite “piovose” e la copertura nuvolosa del cielo sia “completamente assente”. Non è trascurabile che tali incongruenze, se meno contraddittorie, possono rimanere trasparenti anche se sottoposte a operazioni manuali di pulizia *a posteriori* dei dati.

3.1.2. Modello proposto

Conclusa l’analisi dei modelli attualmente in uso per la catalogazione dei dati di censimento ANSER e individuate le condizioni essenziali per cui essi si caratterizzano, si è provveduto a definire un modello alternativo capace di rappresentare tutte le informazioni in un unico database. Inoltre l’organizzazione delle relazioni tra entità ha subito un riassetto più consona a quelle che sono le richieste espresse dalla sintassi EML con cui verrà implementato elettronicamente questo modello logico di dati, come descritto nel capitolo seguente (3.1.3. Integrazione EML). A seguito di alcune operazioni di normalizzazione delle relazioni esistenti, è stato necessario aggiungere alcune entità supplementari che nella figura 8 sono rappresentate con una tonalità più scura. Di colore verde appare l’elemento radice del grafo.

La minima generalizzazione massima da cui si può partire per descrivere il modello definisce che tutte le informazioni disponibili fanno riferimento a eventi di incontro tra un soggetto e un oggetto, in un tempo e in un luogo, tramite un metodo. L’entità radice quindi viene chiamata Evento e contiene le tuple costituite dall’associazione di una chiave primaria definita da cinque attributi con un attributo di descrizione facoltativo (Evento.idInformazioniSupplementari).

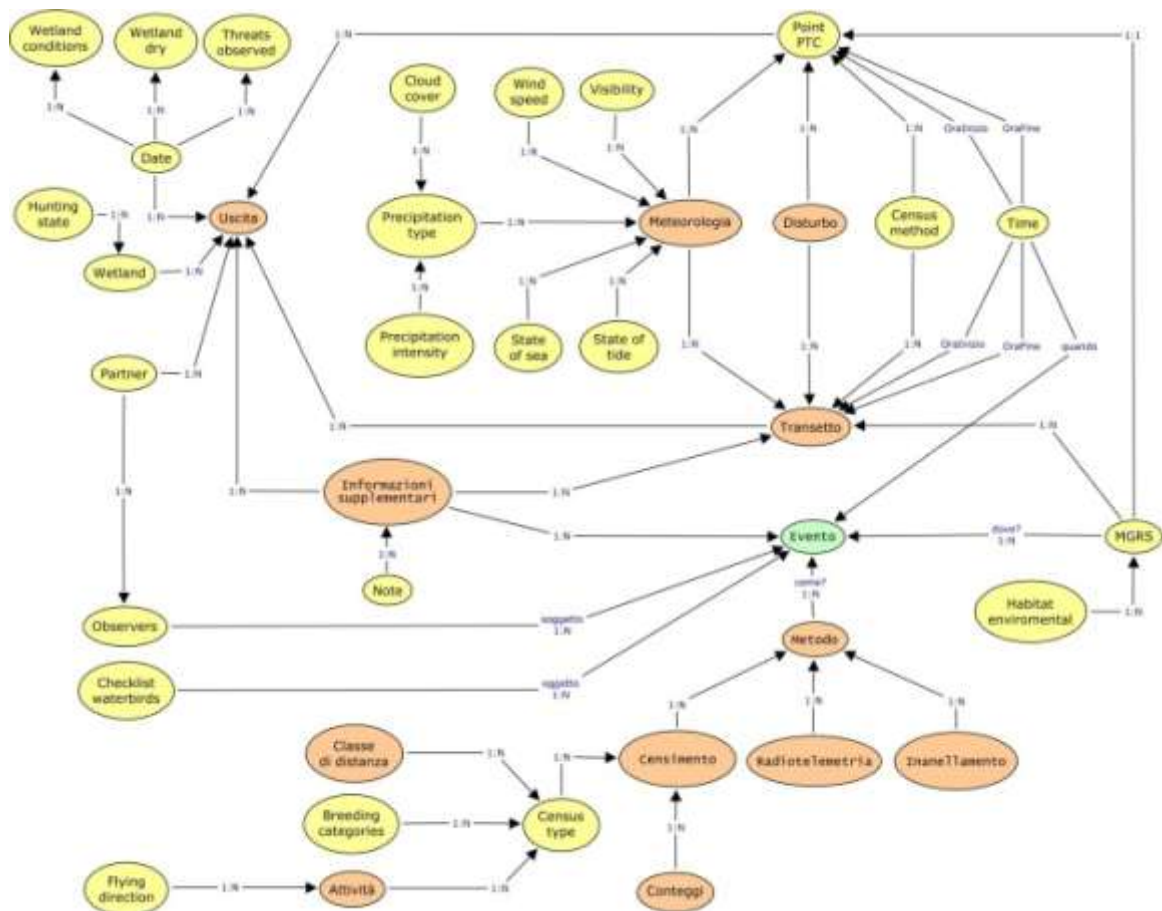


Figura 8

I cinque attributi che determinano la chiave primaria dell'entità Evento sono le chiavi esterne di cinque entità che rispondono alla definizione completa di un evento, e cioè:

- chi?* con l'entità Observers
- cosa?* con l'entità Checklist_waterbirds
- quando?* con l'entità Time
- dove?* con l'entità MGRS
- come?* con l'entità Metodo

Procedendo in ordine, l'entità *Observers* contiene le informazioni relative agli osservatori/operatori che compiono il rilevamento su campo. Essi rappresentano il soggetto dell'evento di incontro. I diversi osservatori/operatori possono essere organizzati in gruppi di appartenenza attraverso una chiave esterna nell'entità *Partner*.

L'entità *Checklist_waterbirds* contiene l'elenco completo delle specie di uccelli che costituiscono la comunità ornitica e rappresenta l'oggetto dell'evento di incontro.

L'entità *Time* definisce l'orario puntuale in cui l'evento di incontro è avvenuto. I valori disponibili sono quelli compresi all'interno di un arco temporale che caratterizza un'uscita su campo.

L'entità *MGRS* definisce la posizione geografica in cui si è determinato l'evento di incontro. I valori disponibili corrispondono a un punto PTC oppure sono compresi all'interno di un transetto percorso durante un'uscita su campo.

L'entità *Metodo* definisce la metodologia di monitoraggio che ha provocato l'evento di incontro. Essa è qui limitata alla definizione di tre chiavi esterne riferibili alle entità *Censimento*, *Radiotelemetria* e *Inanellamento*. L'entità *Censimento* è quella attualmente meglio definita e appare costituita da due attributi. Il primo si chiama *Censimento.idConteggi* e si relaziona con l'entità *Conteggi* in cui viene registrato il numero di individui che compongono l'oggetto dell'evento di incontro. Il secondo attributo è *Censimento.idCensus_type* e condiziona la metodologia di censimento secondo i protocolli *DTC+RC*, *PTC* e *AS*. L'entità *Census_type* viene arricchita dalle informazioni che descrivono, con tre chiavi esterne, le attività compiute dagli uccelli (*Census_type.idAttività*) per i protocolli *PTC* e *AS*, la classe di distanza (*Census_type.idClasseDiDistanza*) e la categoria di riproduzione (*Census_type.idBreeding_categories*) per il solo protocollo *PTC*. In conclusione, l'entità *Attività* presenta un attributo di chiave esterna per definire le direzioni di volo intraprese (*Attività.Flying_direction*).

Arrivati a questo punto, è necessario approfondire la definizione offerta precedentemente per quanto riguarda le domande *quando?* e *dove?* in quanto non si può ignorare che le risposte date per definire l'evento di incontro tramite le entità *Time* e *MGRS* sono in entrambi i casi puntiformi e insufficienti. In realtà infatti gli eventi di incontro si susseguono all'interno di una finestra temporale durante la quale viene applicato un metodo

di monitoraggio e possono avvenire in luoghi raggiunti sequenzialmente percorrendo un tragitto articolato.

Il monitoraggio prevede diverse uscite su campo organizzate in giornate che sono distribuite accuratamente durante la stagione in modo tale da garantire il massimo grado di copertura dell'area da censire e la maggiore contemporaneità dei rilevamenti. Per organizzare le informazioni relative a queste giornate si propone l'entità Uscita che appare composta da attributi relativi alla data in cui viene compiuta (Uscita.idDate), alla definizione della zona umida interessata (Uscita.idWetland), al gruppo di osservatori che la portano a termine (Uscita.idPartner) e alla descrizione di eventuali annotazioni di anomalie e particolari da ricordare (Uscita.idInformazioniSupplementari). L'entità Date possiede a sua volta tre attributi che definiscono, relativamente a quella data, le condizioni della zona umida (Date.idWetland_conditions), il livello di allagamento della zona umida (Date.idWetland_dry) e l'osservazione della presenza di minacce (Date.idThreats_observed). Inoltre l'entità Wetland presenta l'attributo Wetland.idHunting_state per la definizione del regime di attività venatoria previsto per quella zona umida.

Ogni uscita su campo giornaliera prevede la percorrenza di transetti che attraversano diversi punti MGRS per i protocolli DTC+RC e AS oppure il raggiungimento di vari punti fissi per il protocollo PTC. Per questo motivo l'entità Uscita contiene altri due attributi in chiave esterna per la definizione di transetti (Uscita.idTransetto) e di punti fissi (Uscita.idPoint_PTC).

L'entità Transetto e l'entità Point_PTC presentano quasi lo stesso assortimento di attributi. Alcuni di essi definiscono l'orario di inizio e di fine di percorrenza del transetto (Transetto.idTimeInizio e Transetto.idTimeFine) oppure dell'intervallo in cui avviene l'osservazione da un punto fisso (Point_PTC.idTimeInizio e Point_PTC.idTimeFine). Altri si riferiscono al mezzo con cui è percorso il transetto o raggiunto il punto fisso (Transetto.idCensus_method e Point_PTC.idCensus_method), alla presenza di disturbi associati a varie attività antropiche (Transetto.idDisturbo e Point_PTC.idDisturbo) e ai rilevamenti meteorologici (Transetto.idMeteorologia e Point_PTC.idMeteorologia). L'unica differenza tra l'entità Transetto e l'entità Point_PTC è determinata dalla cardinalità della relazione in essere tra i punti MGRS con quelli PTC che è 1:1 (Point_PTC.idMGRS) e quella della relazione tra i punti MGRS e il percorso del transetto che è 1: \mathcal{N} (Transetto.idMGRS).

Da ultimo, l'entità Meteorologia è composta da attributi che descrivono le condizioni in atto durante l'intervallo di tempo già definito attraverso l'entità Time durante la percorrenza del transetto oppure nel periodo in cui avviene l'osservazione da un punto fisso. Essi riguardano il livello di marea (Meteorologia.idState_of_tide), le condizioni del mare (Meteorologia.idState_of_sea), la visibilità (Meteorologia.idVisibility), la velocità del vento (Meteorologia.idWind_speed) e la tipologia di precipitazione (Meteorologia.idPrecipitation_type). A sua volta, l'entità Precipitation_type è completata da due attributi che indicano l'intensità delle precipitazioni qualora presenti (Precipitation_type.idPrecipitation_intensity) e la copertura nuvolosa del cielo (Precipitation_type.idCloud_cover).

3.1.3. Integrazione EML

Le operazioni che sono necessarie per l'integrazione di tutti i modelli attualmente in uso con il modello proposto nel capitolo precedente, sono essenzialmente legate alla scrittura di quattro file in formato EML. Le entità e gli attributi definiti nei metadati EML per DTC+RC, PTC e AS sono perpetuati nel modello integrato mantenendone la coerenza di definizione ma riorganizzando le relazioni in essere anche attraverso la definizione sintattica di entità supplementari.

Nei quattro file EML i metadati delle categorie generali che descrivono i dataset sono identici per quanto riguarda le informazioni geografiche, temporali, tassonomiche e di protocolli. A titolo di esempio viene riportata la rappresentazione sia logica che fisica dei nodi coinvolti per quello che riguarda la porzione che definisce l'area geografica in cui sono raccolti i dati ANSER. Gli schemi della Figura 9 si riferiscono alle seguenti informazioni:

Geographic Coverage:	
Geographic Description:	Pianura deltizia e aree lagunari del Friuli Venezia-Giulia
Bounding Coordinates:	West: 13.0 degrees
	East: 13.55 degrees
	North: 45.5 degrees
	South: 45.34 degrees



Figura 9

Viceversa i nodi EML che contengono i metadati riguardanti il titolo, l'abstract, le parole chiave, i metodi e le tabelle dei dati contengono informazioni differenti a seconda del dataset che il file rappresenta. La definizione completa dei file EML è documentata e scaricabile all'URL www.infs-epe.it/ontology/ANSwER/EML. La sezione che contiene le informazioni sulle tabelle di dati è la più importante perché in essa sono contenuti tutti i metadati che riguardano la struttura logica dei modelli precedentemente analizzati. Le diverse entità presentano i propri dati all'interno di un file di testo con valori separati da virgola. Inoltre ogni entità presenta i metadati che descrivono il relativo file csv per quanto riguarda il nome, la descrizione del contenuto, la dimensione in byte, il formato del testo e il numero di record contenuti. Il nome delle entità riscontrato nei file mdb è stato rigorosamente

mantenuto per i file csv e i nomi degli attributi che compongono l'entità sono sempre espressi nella prima riga del file csv. Gli schemi della Figura 10 si riferiscono alla rappresentazione logica e fisica dell'entità Census_method, rispetto alle seguenti informazioni:

Data Table:	
Name:	Census_method.csv
Description:	Metodologia di censimento
Physical Structure Description:	
Object Name:	Census_method.csv
Size:	51 byte
Number of Header Lines:	1
Record Delimiter:	#x0A
Text Format:	Maximum
Record Length:	column
Simple Delimited:	Field
Delimited:	Delimiter: ,
Number Of Records:	4

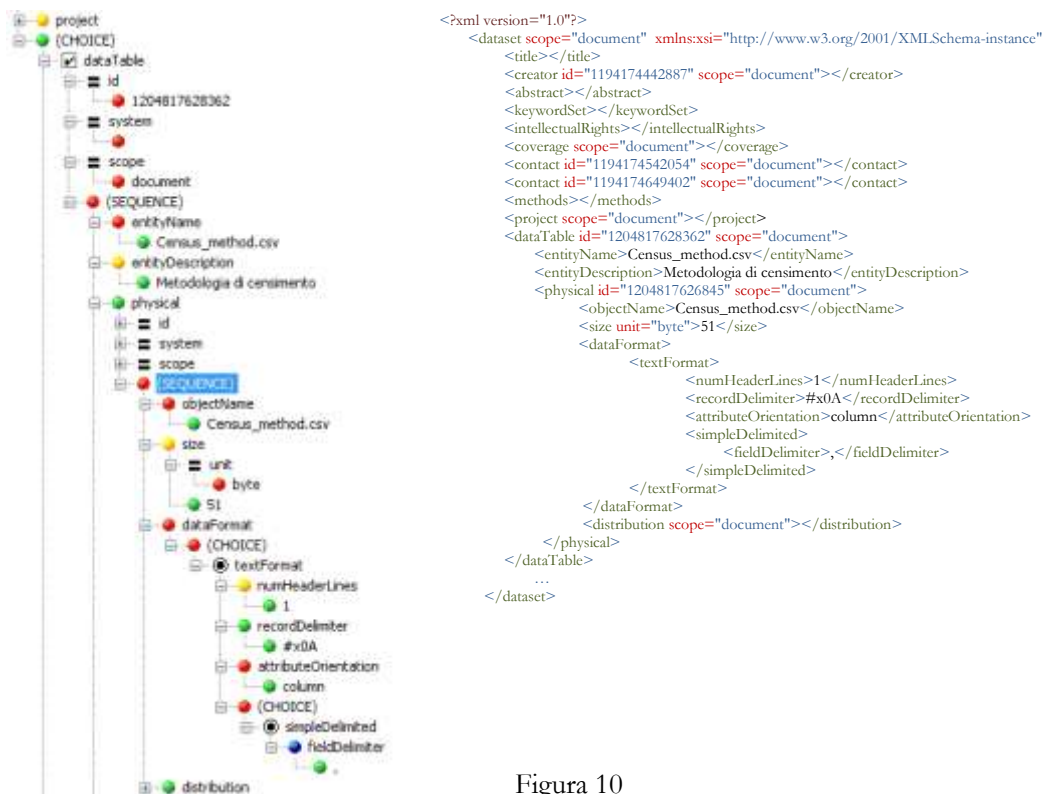


Figura 10

All'interno di ogni entità sono stati descritti gli attributi tramite metadati che ne registrano innanzitutto il nome, l'etichetta e la definizione. Anche nel caso degli attributi si è mantenuto scrupolosamente il nome che essi presentavano nei file mdb originari, mentre tramite l'etichetta si è provveduto a chiarire eventuali ruoli di chiave primaria (prefisso PK) e chiave esterna (prefisso FK). La definizione inoltre ha permesso di esprimere il contenuto dell'attributo più accuratamente, così che sia semplice per chiunque interpretare i dati. Altri metadati registrano il tipo di dati, il sistema che li gestisce, e la categorizzazione, cioè la scala di misurazione adottata per l'attributo. Ad esempio l'attributo ID_wetlands che rappresenta la chiave primaria dell'entità Wetlands ha etichetta PKwetlands, tipo di dati integer, sistema MS Access e categoria di misurazione interval, cioè nella scala a intervalli equidistanti. I dati di tutti gli attributi sono stati ulteriormente definiti attraverso metadati consoni alla scala con cui vengono rappresentati cioè, nel caso di scala interval, tramite l'espressione dell'unità di misura standard, della precisione, del tipo di numero e, se vigenti, dei valori di minimo e di massimo consentiti.

Gli attributi con tipo di dati text e categoria di misurazione nominal sono stati definiti a seconda dei casi come elenchi di valori, di codici, di nomi o di descrizioni. Gli schemi della Figura 11 si riferiscono alla rappresentazione logica e fisica degli attributi definiti per l'entità Wetlands, rispetto alle seguenti informazioni:

Data Table:									
Name:		Wetland.csv							
Description:		Elenco delle zone umide							
Physical Structure Description:									
Object Name:		Wetland.csv							
Size:		1253 byte							
Text Format:		Number of Header Lines:		1		Record Delimiter:		#x0A	
		Maximum Record Length:		column		Simple Delimited:		Field Delimiter: ,	
Number Of Records:		42							
Attribute(s) Info:									
Attribute Name	Column Label	Definition	Type of Value	Measurement Type	Measurement Domain	Missing Value Code	Accuracy Report	Accuracy Assessment	CoverageMethod
ID_wetlands	PKwetlands	Chiave primaria della tabella	integer	interval	Unit dimensionless Precision 1 Type natural				
Wetlands_code	Codice della zona umida	Sistema di codifica dei diversi corpi idrici presenti nell'area di studio ANSER	text	nominal	Def Elenco di codici				
Wetland_name	Nome della zona umida	Nomi attribuiti ai diversi corpi idrici presenti nell'area di studio ANSER	text	nominal	Def Elenco di nomi				
Online Distribution Info:									
ecogrid//knb/ANSWER.53.1									



```

<?xml version="1.0"?>
<dataset scope="document" xmlns:xsi="http://www.w3.org/2001/
XML.Schema-instance">
<title></title>
<creator id="1194174442887" scope="document"></creator>
<abstract></abstract>
<keywordSet></keywordSet>
<intellectualRights></intellectualRights>
<coverage scope="document"></coverage>
<contact id="1194174542054" scope="document"></contact>
<contact id="1194174649402" scope="document"></contact>
<methods></methods>
<project scope="document"></project>
<dataTable id="1204822350796" scope="document">
<entityName>Wetland.csv</entityName>
<entityDescription>Elenco delle zone umide</entityDescription>
<physical id="1204822349345" scope="document"></physical>
<attributeList>
<attribute id="1204822350797" scope="document">
<attributeName>ID_wetlands</attributeName>
<attributeLabel>PKwetlands</attributeLabel>
<attributeDefinition>Chiave primaria della
tabella</attributeDefinition>
<storageType typeSystem="MS
Access">integer</storageType>
<measurementScale>
<interval>
<unit>
<standardUnit>dimensionless</standardUnit>
</unit>
<precision>1</precision>
<numericDomain>
<numberType>natural</numberType>
</numericDomain>
</interval>
</measurementScale>
</attribute>
<attribute id="1204822350798" scope="document">
<attributeName>Wetlands_code</attributeName>
<attributeLabel>Codice della zona
umida</attributeLabel>
<attributeDefinition>Sistema di codifica
dei diversi corpi idrici
presenti nell'area di studio
ANSER</attributeDefinition>
<storageType typeSystem="MS
Access">text</storageType>
<measurementScale>
<nominal>
<nonNumericDomain>
<textDomain>
<definition>Elenco di codici</definition>
<pattern>sigla Provincia +
numero 4 cifre</pattern>
</textDomain>
</nonNumericDomain>
</nominal>
</measurementScale>
</attribute>
<attribute id="1204822350799" scope="document">
<attributeName>Wetland_name</attributeName>
<attributeLabel>Nome della zona
umida</attributeLabel>
<attributeDefinition>Nomi attribuiti ai diversi
corpi idrici presenti
nell'area di studio
ANSER</attributeDefinition>
<storageType typeSystem="MS
Access">text</storageType>
<measurementScale>
<nominal>
<nonNumericDomain>
<textDomain>
<definition>Elenco di nomi</definition>
</textDomain>
</nonNumericDomain>
</nominal>
</measurementScale>
</attribute>
</attributeList>
<numberOfRecords>42</numberOfRecords>
</dataTable>
...
</dataset>

```

Figura 11

3.2. Inferenza semantica

3.2.1. Dimensioni del modello ontologico

L'implementazione dell'ontologia di dominio in linguaggio OWL-DL del sistema ecologico di zona umida indagato dal progetto ANSER ha preso inizio con la fase di analisi di quali fattori dovessero essere coinvolti nel modello in modo tale da consentire la classificazione delle diverse specie di uccelli censiti in gruppi simili per quanto riguarda la selezione dell'habitat, l'attività di foraggiamento e la dieta. Le opzioni attuate dagli uccelli acquatici per queste tre dimensioni sono, come già espresso nell'introduzione, largamente dinamiche durante le differenti stagioni dell'anno e strutturalmente variegata per l'ampia scelta di soluzioni plausibili offerte dall'ambiente. Per questo motivo si è provveduto a migliorare la finezza di alcune definizioni attraverso l'utilizzo di partizioni di valori.

Procedendo con ordine e più in dettaglio, il dominio radice del modello considera tre sottoclassi. La prima, Habitat, contiene l'elenco di macrotipologie di habitat disponibili negli ecosistemi di zone umide costiere, secondo la codifica Corine Land Cover (CEC 1991).

La seconda classe ha nome *TatticaDiAlimentazione* ed è intesa a definire con quale comportamento motorio l'animale attua di consuetudine l'attività di foraggiamento. Le azioni individuate sono espresse in cinque classi: Volare, Nuotare, Camminare, Disseppellire e Tuffarsi. Esse sembrano sufficientemente esaustive per rappresentare l'intero assortimento di azioni intraprese dagli individui per conquistare la risorsa alimentare. Se le prime tre descrivono comportamenti diffusi e attuati da numerosi gruppi di specie, non solo di uccelli, l'attività di "disseppellire" le prede endobentoniche nascoste nel fango degli habitat intertidali e quella di "tuffarsi" in acqua intervallando sorvoli di *hovering* per la cattura di prede acquatiche, rappresentano due tattiche peculiari attuate essenzialmente dagli uccelli nelle zone umide. Al fine di consentire una descrizione più articolata di questa dimensione etologica, sono state definite due partizioni di valore, la prima relativa a quanto profondamente i limicoli infilano il becco nel fango per disseppellire le prede (Perforazione), la seconda al fine di classificare la profondità che gli animali riescono a raggiungere sott'acqua (Immersione).

La terza e ultima sottoclasse del dominio si chiama Cibo e in essa è stata organizzata, attraverso le due classi *OrigineAnimale* e *OrigineVegetale*, una tassonomia delle principali risorse alimentari che compongono la dieta delle specie in esame. In questo contesto, il termine “tassonomia” non è da intendersi nella classica e familiare accezione delle scienze della vita, cioè come albero filogenetico di specie viventi. Nel campo dell’ingegneria della conoscenza, infatti, una tassonomia è più semplicemente l’organizzazione di concetti o classi legate da “sussunzione terminologica”. La sussunzione di una sottoclasse da una classe (*sottoclasse* \sqsubseteq *classe*) implica che ogni istanza descritta dalla sottoclasse sia descritta anche dalla classe. Infatti se la classe *OrigineAnimale* è da intendersi come l’insieme di risorse alimentari che hanno un’origine non vegetale, l’aver definito che altre due sottoclassi come *Invertebrati* e *Vertebrati* sussungono da essa, implica che anche gli invertebrati e i vertebrati costituiscono cibo di origine non vegetale. Da un punto di vista formale, la “logica dei predicati del primo ordine” (FOL) esprime $Invertebrati \sqsubseteq OrigineAnimale$ con $\forall x(Invertebrati_x \rightarrow OrigineAnimale_x)$. Così facendo è lecito ridurre ai minimi termini necessari l’architettura dei concetti includendo definizioni meno rigorose ma più semplici e logicamente corrette, come nel caso $InsettiLarve \sqsubseteq OrigineAnimale$. La dimensione legata alla scelta del cibo così come è implementata attualmente nel modello viene descritta dal grafo della figura 12.

Anche per quanto riguarda la classificazione delle preferenze alimentari mostrate dagli uccelli, è stato inserito un arricchimento delle caratterizzazioni disponibili attraverso la classe di partizione di valori denominata *SpettroTrofico* in maniera tale da permettere la definizione di ogni specie come *Generalista* o *Specialista*.

La “assunzione di apertura del dominio” di cui godono le ontologie OWL-DL ha inoltre evitato lo svantaggio di dover considerare il modello come completamente definito, anche perché in ogni caso sarebbe risultato utopico. Anche in questa ontologia quindi, tutti i concetti non ancora espressi nominalmente non vengono considerati come inesistenti. Questa situazione di tolleranza verso la “conoscenza incompleta” conduce al secondo assunto in vigore nelle Logiche Descrittive, detto “assunzione del mondo aperto”. Essa dichiara che un concetto non definito come “vero” non diventa automaticamente “falso”, ma rimane sconosciuto fino al momento in cui una definizione del suo valore di verità verrà aggiunta alla base di conoscenza.



Figura 12

Alla luce di queste considerazioni, le classi di concetti definite nel modello possono ora essere messe in relazione l'un l'altra attraverso le "proprietà", cioè i predicati a due argomenti. Allo scopo sono state definite otto proprietà, ciascuna dotata della relativa proprietà inversa, in modo tale da poter qualificare le condizioni asserite per ogni classe di specie relativamente alle classi di Habitat, TatticaDiAlimentazione, Cibo e tutte le partizioni di valori associate. Due proprietà, hasPreferenza e hasAvversione, servono a condizionare le istanze definendo per ogni specie quali siano le preferenze o le avversioni conosciute verso i differenti habitat. Le relazioni inverse sono denominate rispettivamente isPreferenzaOf e isAvversioneOf. Le tipologie di azione comportamentale preferite da ogni specie sono

condizionate dalla proprietà `hasTatticaDiAlimentazione` (con l'inverso `isTatticaDiAlimentazioneOf`) mentre le scelte alimentari sono condizionate dalla proprietà `hasCibo` (`isCiboOf`). Le proprietà utili alla espressione di quali siano le modalità di disseppellimento, delle capacità subacquee e dell'ampiezza dello spettro trofico sono rispettivamente chiamate `hasPerforazione` (`isPerforazioneOf`), `hasImmersione` (`isImmersioneOf`) e `hasSpettroTrofico` (`isSpettroTroficoOf`). L'ultima proprietà di qualifica condiziona l'appartenenza delle diverse specie alle *guild* ed è stata nominata `hasGuild` (`isGuildOf`). Tutte queste proprietà sono sussunte da una nona proprietà chiamata `hasStrategia` (`isStrategiaOf`).

3.2.2. Condizioni per classi primitive e classi complete

Il popolamento delle conoscenze asserite per tutte le specie presenti nel modello è stato condotto definendo il più completamente possibile le restrizioni note per quanto riguarda tutte le otto proprietà descritte nel capitolo precedente. Queste condizioni sono definite “necessarie” nel senso che sono condizioni necessarie ma non sufficienti. Per esse infatti se qualcosa è membro della classe allora è necessario che soddisfi queste condizioni, senza però poter affermare che se qualcosa soddisfa queste condizioni allora deve essere membro della classe. Molto intuitivamente si tratta di una relazione di sussunzione tra le condizioni necessarie e la classe “primitiva” a cui si riferiscono attraverso l'assioma terminologico *classe* \sqsubseteq *condizione necessaria* cioè, in FOL, $\forall x(\text{classe}_x \rightarrow \text{condizione necessaria}_x)$.

Diversamente invece è stato definito per le classi di *Habitat*, *TatticaDiAlimentazione* e per tutte le classi che definiscono le *guild* caratteristiche degli uccelli acquatici. In questi casi infatti le condizioni sono definite “necessarie e sufficienti” indicando così che qualcosa è membro della classe se e solo soddisfa le condizioni. Questa relazione consiste in una uguaglianza tra le condizioni necessarie e sufficienti e la classe “completa” a cui si riferiscono attraverso la definizione *classe* \equiv *condizione necessaria*, cioè $\forall x(\text{classe}_x \leftrightarrow \text{condizione necessaria}_x)$.

Questo assetto logico permette all'agente ragionatore automatico di controllare la consistenza dei concetti asseriti attraverso le condizioni ma soprattutto di classificare in maniera del tutto automatica le specie di uccelli nelle diverse classi definite. A titolo di esempio, volendo ottenere l'elenco di tutte le specie che si alimentano camminando e

nuotando, è sufficiente creare una classe *CamminatoriNuotatori* in cui vengono definite le seguenti condizioni necessarie e sufficienti:

$$\text{CamminatoriNuotatori} \equiv \left\{ \begin{array}{l} \exists \text{ hasTatticaDiAlimentazione Camminare} \\ \exists \text{ hasTatticaDiAlimentazione Nuotare} \\ (\forall \text{ hasTatticaDiAlimentazione (Camminare} \sqcup \text{Nuotare)}) \end{array} \right.$$

che sintatticamente con il linguaggio OWL si indica:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns="http://www.owl-ontologies.com/Ontology1205364299.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/Ontology1205364299.owl">
  <owl:Ontology rdf:about="" />
  <owl:Class rdf:ID="NuotatoriCamminatori">
  <owl:equivalentClass>
  <owl:Class>
  <owl:intersectionOf rdf:parseType="Collection">
  <owl:Class rdf:about="#UccelliDiLaguna" />
  <owl:Restriction>
  <owl:allValuesFrom>
  <owl:Class>
  <owl:unionOf rdf:parseType="Collection">
  <owl:Class rdf:about="#Nuotare" />
  <owl:Class rdf:about="#Camminare" />
  </owl:unionOf>
  </owl:Class>
  </owl:allValuesFrom>
  <owl:onProperty>
  <owl:ObjectProperty rdf:about="#hasTatticaDiAlimentazione" />
  </owl:onProperty>
  </owl:Restriction>
  <owl:Restriction>
  <owl:onProperty>
  <owl:ObjectProperty rdf:about="#hasTatticaDiAlimentazione" />
  </owl:onProperty>
  <owl:someValuesFrom rdf:resource="#Nuotare" />
  </owl:Restriction>
  <owl:Restriction>
  <owl:someValuesFrom rdf:resource="#Camminare" />
  <owl:onProperty>
  <owl:ObjectProperty rdf:about="#hasTatticaDiAlimentazione" />
  </owl:onProperty>
  </owl:Restriction>
  </owl:intersectionOf>
  </owl:Class>
  </owl:equivalentClass>
  </owl:Class>
</rdf:RDF>
```

Il risultato della classificazione automatica consiste nella inferenza della classe *CamminatoriNuotatori* che contiene le sottoclassi degli uccelli che si alimentano sia camminando che nuotando, vale a dire, per questo modello, *Fischione* e *OcaSelvatica*. Applicando questo sviluppo delle classi definite per tutti i comportamenti previsti dalla classe *TatticheDiAlimentazione* e per tutti gli habitat definiti nella classe *Habitat* è risultato molto semplice rendere in grafo eventuali errori di codifica.

Con la medesima forma, ma a un livello più profondo di definizione, si ritrovano le classi con cui sono state definite tutte le *guild* degli uccelli acquatici. Come esempio, si definiscono qui le condizioni necessarie e sufficienti per due classi complete riferite alla *guild* degli ittiofagi nuotatori e a quella degli uccelli onnivori.

$$\text{IttiofagiNuotatori} \equiv \left\{ \begin{array}{l} \forall \text{ hasAvversione } (\text{Arbusti} \sqcup \text{Barene} \sqcup \text{Dune} \sqcup \text{Canneti} \sqcup \text{Boschi}) \\ \exists \text{ hasCibo Pesci} \\ \forall \text{ hasCibo Pesci} \\ \exists \text{ hasImmersione Profonda} \\ \exists \text{ hasPerforazione Assente} \\ \forall \text{ hasPreferenza } (\text{Canali} \sqcup \text{Valli}) \\ \exists \text{ hasSpettroTrofico Specialista} \\ \forall \text{ hasSpettroTrofico Specialista} \\ \exists \text{ hasTatticaDiAlimentazione Nuotare} \\ \forall \text{ hasTatticaDiAlimentazione Nuotare} \end{array} \right.$$

$$\text{Onnivori} \equiv \left\{ \begin{array}{l} \forall \text{ hasAvversione } (\text{Arbusti} \sqcup \text{Canneti} \sqcup \text{Boschi}) \\ \exists \text{ hasCibo Vertebrati} \\ \exists \text{ hasCibo Invertebrati} \\ \forall \text{ hasCibo } (\text{Invertebrati} \sqcup \text{Vertebrati}) \\ \forall \text{ hasImmersione } (\text{Nulla} \sqcup \text{Superficiale}) \\ \forall \text{ hasPerforazione Assente} \\ \exists \text{ hasPerforazione Assente} \\ \forall \text{ hasPreferenza } (\text{Dune} \sqcup \text{Valli}) \\ \exists \text{ hasSpettroTrofico Generalista} \\ \forall \text{ hasSpettroTrofico Generalista} \\ \exists \text{ hasTatticaDiAlimentazione Camminare} \\ \exists \text{ hasTatticaDiAlimentazione Nuotare} \\ \forall \text{ hasTatticaDiAlimentazione } (\text{Nuotare} \sqcup \text{Camminare}) \end{array} \right.$$

La classificazione dell'agente ragionatore automatico risponde con le classi Cormorano, SmergoMinore e SvassoMaggiore sussunte dalla classe IttiofagiNuotatori (figura 14) e con le classi GabbianoComune e GabbianoReale sussunte dalla classe Onnivori (figura 16).

La sintassi OWL-DL è mostrata nelle pagine seguenti, mentre la documentazione completa delle *guild*, come d'altronde dell'intera ontologia, è navigabile e scaricabile all'URL www.infs-epe.it/ontology/ANSwER/OWL.

Sintassi OWL della classe Onnivori

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns="http://www.owl-ontologies.com/Ontology1205364299.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.owl-ontologies.com/Ontology1205364299.owl">
  <owl:Ontology rdf:about="" />
  <owl:Class rdf:ID="IttiofagiNuotatori">
    <owl:equivalentClass>
      <owl:Class>
        <owl:intersectionOf rdf:parseType="Collection">
          <owl:Class rdf:about="#UccelliDiLaguna" />
          <owl:Restriction>
            <owl:someValuesFrom rdf:resource="#Profonda" />
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasImmersione" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:someValuesFrom rdf:resource="#Nuotare" />
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasTattica" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:someValuesFrom rdf:resource="#Pesci" />
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasCibo" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:allValuesFrom rdf:resource="#Pesci" />
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasCibo" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasPerforazione" />
            </owl:onProperty>
            <owl:someValuesFrom rdf:resource="#Assente" />
          </owl:Restriction>
          <owl:Restriction>
            <owl:allValuesFrom rdf:resource="#Specialista" />
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasSpettroTrofico" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:someValuesFrom rdf:resource="#Specialista" />
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasSpettroTrofico" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:allValuesFrom>
              <owl:Class>
                <owl:unionOf rdf:parseType="Collection">
                  <owl:Class rdf:about="#Canali" />
                  <owl:Class rdf:about="#Valli" />
                </owl:unionOf>
              </owl:Class>
            </owl:allValuesFrom>
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasPreferenza" />
            </owl:onProperty>
          </owl:Restriction>
          <owl:Restriction>
            <owl:allValuesFrom>
              <owl:Class>
                <owl:unionOf rdf:parseType="Collection">
                  <owl:Class rdf:about="#Arbusti" />
                  <owl:Class rdf:about="#Barene" />
                  <owl:Class rdf:about="#Dune" />
                  <owl:Class rdf:about="#Canneti" />
                  <owl:Class rdf:about="#Boschi" />
                </owl:unionOf>
              </owl:Class>
            </owl:allValuesFrom>
            <owl:onProperty>
              <owl:ObjectProperty rdf:about="#hasAvversione" />
            </owl:onProperty>
          </owl:Restriction>
        </owl:intersectionOf>
      </owl:Class>
    </owl:equivalentClass>
  </owl:Class>
</rdf:RDF>
```


3.2.3. Gerarchia inferita dall'agente ragionatore automatico

Concluse le operazioni legate alla componente asserzionale dell'ontologia, rimane l'ultimo compito di attivare l'agente ragionatore automatico, il quale compirà deduzioni sulla consistenza della base di conoscenza, sulla consistenza dei concetti e sulla classificazione per sussunzione, secondo quanto già descritto nel capitolo 1.3. In meno di un secondo, le diverse specie di uccelli vengono classificate in una gerarchia inferita che deriva dalla applicazione integrale delle regole definite nel modello.

Nelle pagine seguenti vengono mostrati i grafi di risposta ad alcune interrogazioni formulate sull'ontologia. Più precisamente, nella figura 13 si possono notare le sussunzioni inferite dall'agente ragionatore automatico tra le diverse *guild* descritte nella base di conoscenza. Non vengono mostrate le classi delle specie per mantenere semplice la lettura.

Le figure successive invece mostrano anche le classi di specie classificate nelle varie *guild*, prima selezionando quelle ittiofaghe, senza e con le classi di habitat interessate (figura 14 e figura 15), quindi quelle dei consumatori di invertebrati (figura 16) e, in ultimo, quelle *guild* che interessano uccelli erbivori, generalisti o consumatori di animali terrestri (Figura 17). La figura 18 intende indicare la complessità generale delle relazioni tra le classi di *guild* e le diverse tipologie di habitat frequentate preferenzialmente.

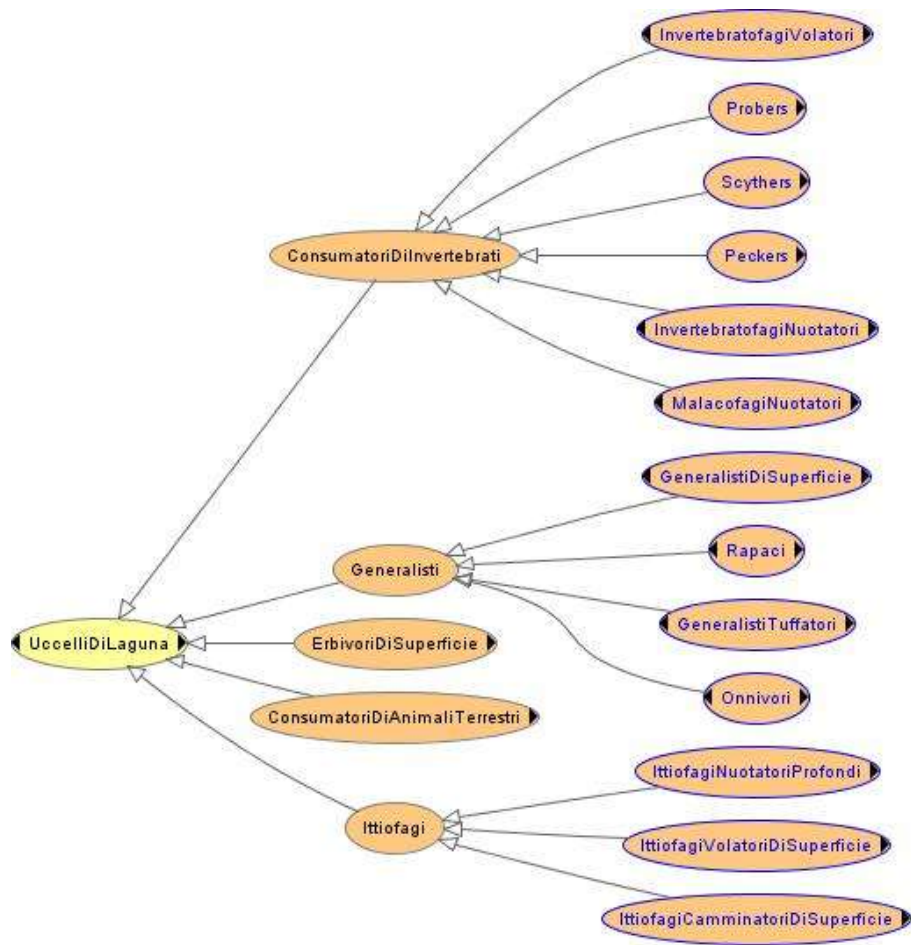


Figura 13

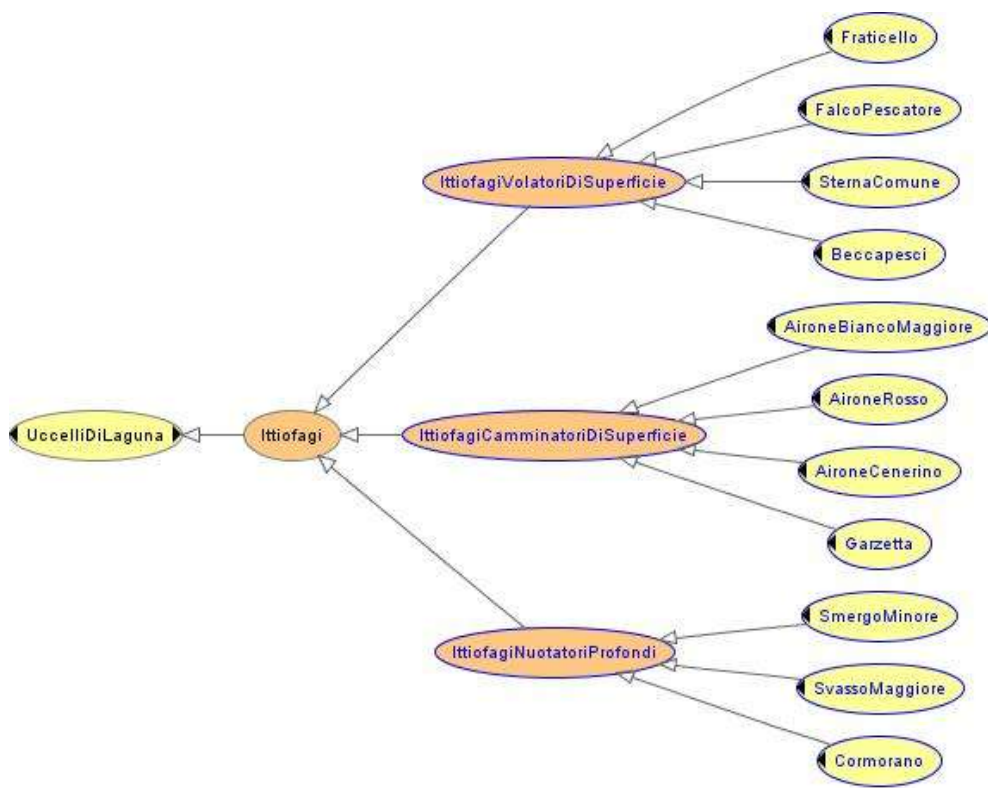


Figura 14

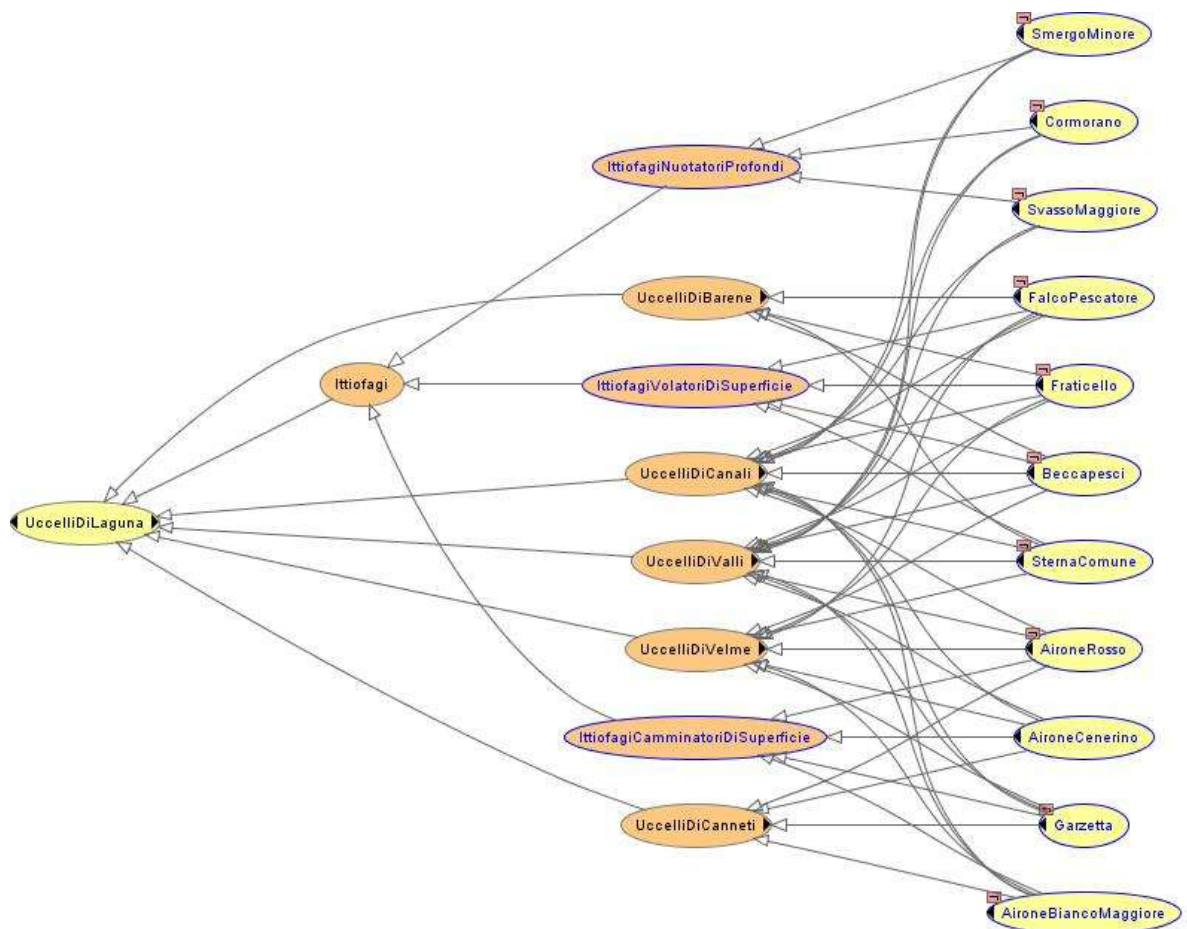


Figura 15

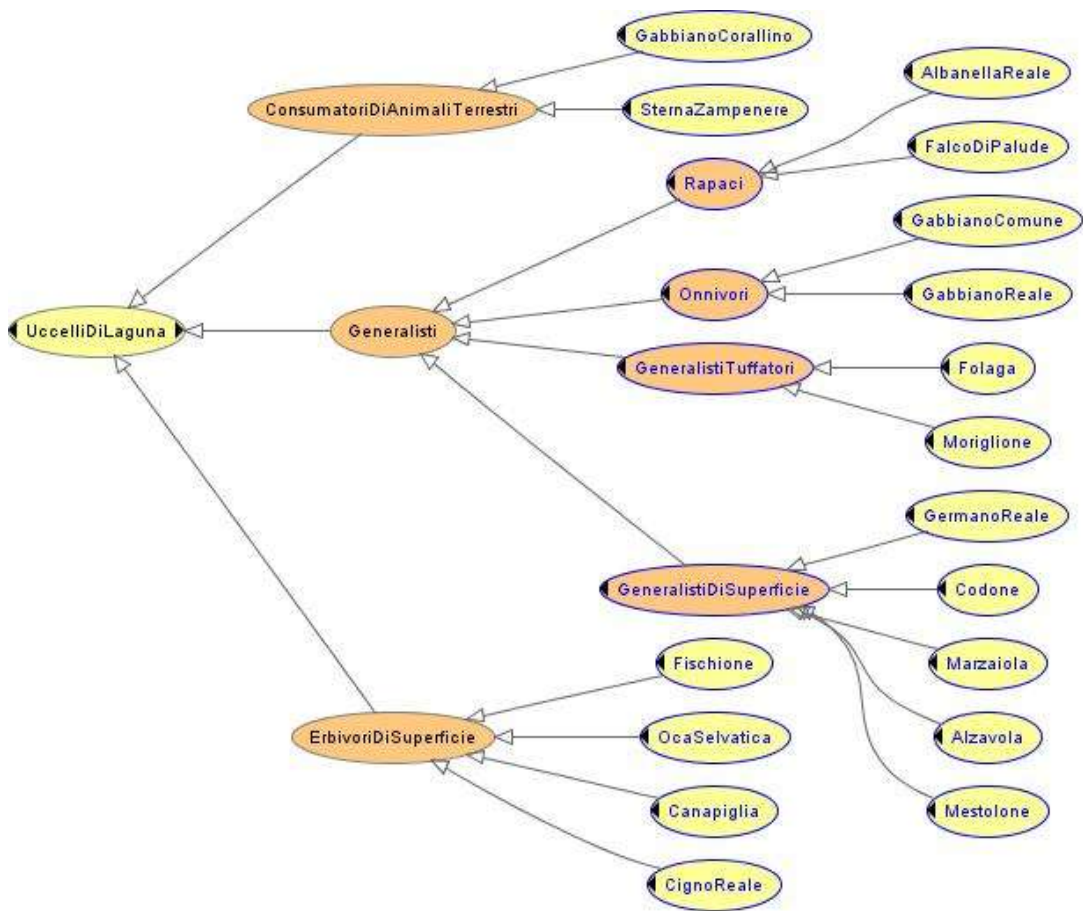


Figura 16

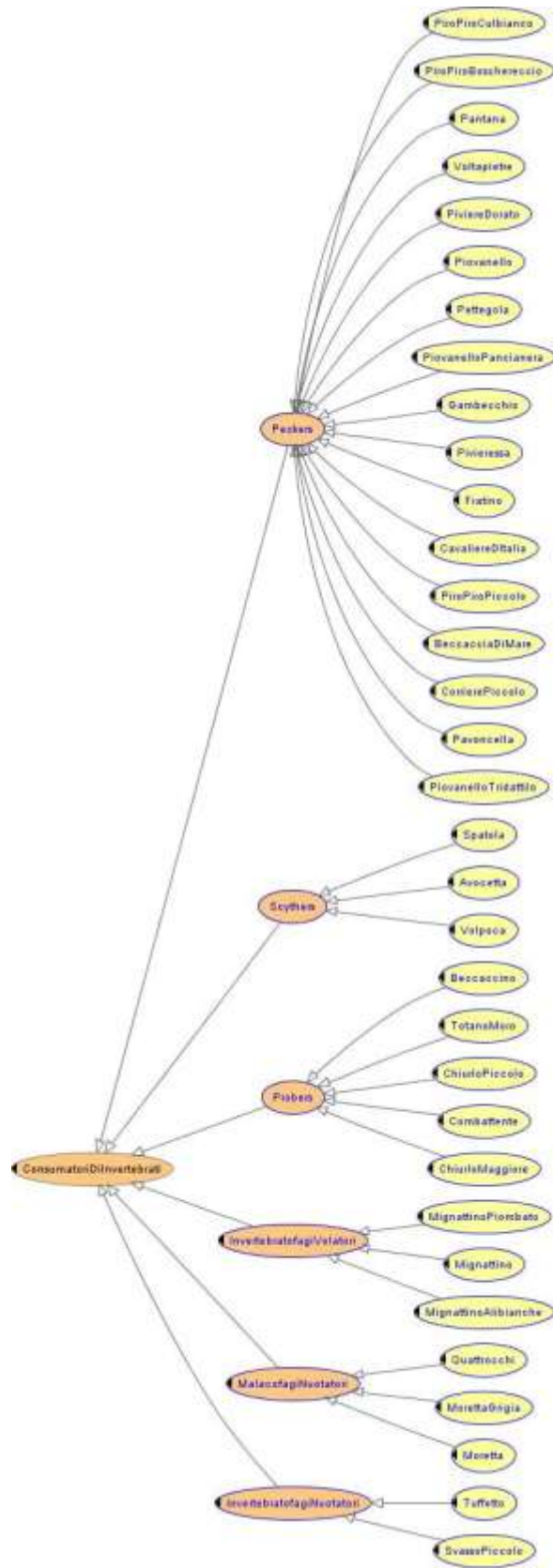


Figura 17

3.3. Modello predittivo *fuzzy*

3.3.1. Variabili I/O e regole IF-THEN

La necessità di valutare se le interpretazioni automaticamente dedotte dalla classificazione svolta sull'ontologia durante il processo di inferenza semantica abbiano o meno una corrispondenza con la reale segregazione spaziale espressa dagli uccelli e desunta attraverso i dati di censimento, ha suggerito l'implementazione di un modello predittivo *fuzzy*. Il presupposto principale consiste nel predire la presenza delle diverse *guild* descritte nel capitolo precedente rispetto alle celle della griglia MGRS, attraverso la quantificazione di una variabile di output di tipo *Mamdani* calcolata grazie ad alcune semplici regole IF-THEN che “ragionano” su dieci premesse. Di queste variabili di input, otto sono relative alla distanza minima esistente tra ogni centroide di cella MGRS e le diverse tipologie di habitat. Gli habitat considerati si suddividono nelle categorie arbusteti, barene, boschi, canali, canneti, dune, valli e velme. Le altre due variabili di input si riferiscono al valore medio di profondità dell'acqua e al dislivello medio del fondale per la cella interessata.

Ogni variabile di input del modello è descritta da due funzioni *membership* che descrivono il variare del grado di verità rispetto alla “prossimità” della cella all'habitat attraverso i due insiemi *fuzzy* vicino e lontano. Esse sono calcolate con la spezzata trapezoidale, cioè con la funzione di un vettore x che dipende dai quattro parametri a , b , c , d e ha forma:

$$f(x; a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}$$

L'unica variabile di output invece quantifica la presenza delle *guild* con tre funzioni *membership* di forma triangolare per esprimere i tre insiemi *fuzzy* bassa, intermedia e alta. La figura 19 ne mostra una rappresentazione grafica, a sinistra per le funzioni input e a destra per l'output.



Figura 19

Per quanto riguarda le due regole IF-THEN che elaborano la variabile di output, la prima implica un più elevato valore di verità se le distanze degli habitat preferiti dalla *guild* sono ridotte e le distanze di quelli avversati sono alte. La seconda regola, viceversa, premette e implica i complementari della prima. Inoltre, per rifinire le regole, si è legata la profondità e la pendenza del fondale a una congetturata ipotesi che fossero gradite o meno dalle diverse *guild*. La tabella 1 mostra quali valori sono stati definiti per l'espressione di tutte le regole.

In base a questo schema quindi si è provveduto a “regolare” il sistema inferenziale *fuzzy*, inserendo, ad esempio, per gli uccelli ittiofagi nuotatori le seguenti implicazioni:

	Arbusti	Barene	Boschi	Canali	Canneti	Dune	Valli	Velme	Acqua profonda	Fondo scosceso
ITTNUO	-	-	-	+	-	-	+	=	+	+

REGOLA 1 1. If (Arbusti is lontano) and (Barene is lontano) and (Boschi is lontano) and (Canali is vicino) and (Canneti is lontano) and (Dune is lontano) and (Valli is vicino) and (AcquaProfonda is vicino) and (FondoScosceso is vicino) then (presenza is alta) (1)

REGOLA 2 2. If (Arbusti is vicino) and (Barene is vicino) and (Boschi is vicino) and (Canali is lontano) and (Canneti is vicino) and (Dune is vicino) and (Valli is lontano) and (AcquaProfonda is lontano) and (FondoScosceso is lontano) then (presenza is bassa) (1)

altrimenti, per i rapaci, queste altre:

	Arbusti	Barene	Boschi	Canali	Canneti	Dune	Valli	Velme	Acqua profonda	Fondo scosceso
RAPACI	-	+	-	=	+	-	-	=	=	=

REGOLA 1 1. If (Arbusti is lontano) and (Barene is vicino) and (Boschi is lontano) and (Canneti is vicino) and (Dune is lontano) and (Valli is lontano) then (presenza is alta) (1)

REGOLA 2 2. If (Arbusti is vicino) and (Barene is lontano) and (Boschi is vicino) and (Canneti is lontano) and (Dune is vicino) and (Valli is vicino) then (presenza is bassa) (1)

I valori *crisp* delle variabili output ottenute per ogni cella MGRS sono stati inseriti in un sistema geografico così da ottenere rappresentazioni grafiche delle predizioni espresse dal modello *fuzzy* facilmente comparabili con quelle che si ottengono con i dati quantitativi dei censimenti.

	Arbusti	Barene	Boschi	Canali	Canneti	Dune	Valli	Velme	Acqua profonda	Fondo scosceso
Ittiofagi nuotatori	lontani	lontani	lontani	vicini	lontani	lontani	vicini	indifferenti	vicini	vicini
Ittiofagi camminatori	lontani	indifferenti	lontani	indifferenti	vicini	lontani	vicini	vicini	lontani	lontani
Ittiofagi volatori	lontani	vicini	lontani	vicini	lontani	lontani	vicini	vicini	indifferenti	indifferenti
Onnivori	lontani	indifferenti	lontani	indifferenti	lontani	vicini	vicini	indifferenti	indifferenti	indifferenti
Invertebratofagi nuotatori	lontani	lontani	lontani	vicini	vicini	lontani	vicini	indifferenti	vicini	vicini
Invertebratofagi volatori	lontani	indifferenti	lontani	indifferenti	indifferenti	lontani	vicini	lontani	vicini	indifferenti
Probers	lontani	indifferenti	lontani	indifferenti	indifferenti	lontani	vicini	vicini	lontani	lontani
Peckers	lontani	indifferenti	lontani	indifferenti	indifferenti	vicini	vicini	vicini	lontani	lontani
Scythers	lontani	indifferenti	lontani	indifferenti	indifferenti	lontani	vicini	vicini	lontani	lontani
Malacofagi	lontani	lontani	lontani	indifferenti	indifferenti	lontani	vicini	vicini	vicini	indifferenti
Consumatori di animali terrestri	lontani	indifferenti	lontani	indifferenti	indifferenti	lontani	vicini	vicini	indifferenti	indifferenti
Rapaci	lontani	vicini	lontani	indifferenti	vicini	lontani	vicini	indifferenti	indifferenti	indifferenti
Erbivori di superficie	lontani	indifferenti	lontani	indifferenti	indifferenti	lontani	vicini	vicini	indifferenti	indifferenti
Generalisti tuffatori	lontani	lontani	lontani	indifferenti	indifferenti	lontani	vicini	lontani	vicini	vicini
Generalisti di superficie	lontani	indifferenti	lontani	indifferenti	vicini	lontani	vicini	vicini	indifferenti	indifferenti

Tabella 1

3.3.2. Validazione

La bontà delle previsioni prodotte dal modello fuzzy è stata valutata confrontando i dati ottenuti dal sistema inferenziale con quelli raccolti sul campo. Per ogni cella MGRS è stato calcolato il coefficiente Kappa di Cohen (Cohen 1960), espresso su una scala di valori compresi tra 0 e 1. I valori del Kappa di Cohen inferiori a 0.4 indicano performance del modello predittivo minime, tra 0.41 e 0.6 moderate, tra 0.61 e 0.8 sostanziali, superiori a 0.81 eccellenti (Landis & Koch 1977). Esso viene calcolato così:

$$Kappa = \frac{(A + D)(A + B + C + D) - (A + B)(A + C) - (C + D)(B + D)}{(A + B + C + D)^2 - (A + B)(A + C) - (C + D)(B + D)}$$

qualora A, B, C e D siano come da tabella 2.

		MODELLO		Tot
		sbagliato	corretto	
MISURE	sbagliato	A	B	A+B
	corretto	C	D	C+D
total		A+C	B+D	A+B+C+D

Tabella 2

La quantità di dati raccolti sinora e disponibili per l'analisi non permette di effettuare la validazione per i censimenti effettuati secondo il protocollo PTC. Per quanto riguarda invece i conteggi effettuati dall'aereo, le performance del modello predittivo sono risultate minime per ogni *guild* considerata (tabella 3). Quest'ultima tipologia infatti era già stata prevista come attività integrativa dei monitoraggi a terra e mostra i propri limiti sia per la copertura spaziale parziale che per il previsto mancato riconoscimento delle specie più piccole o criptiche.

Protocollo AS	K di Cohen	
Ittiofagi nuotatori	0,04	Minima
Ittiofagi camminatori	0	Minima
Ittiofagi volatori	0,01	Minima
Onnivori	0,01	Minima
Invertebratofagi nuotatori	0	Minima
Invertebratofagi volatori	0	Minima
Probers	0,01	Minima
Peckers	0,01	Minima
Scythers	0,02	Minima
Malacofagi	0,01	Minima
Consumatori di animali terrestri	0,01	Minima
Rapaci	0,02	Minima
Erbivori di superficie	0	Minima
Generalisti tuffatori	0,02	Minima
Generalisti di superficie	0,01	Minima

Tabella 3

Al contrario, la validazione rispetto alla somma dei dati di censimento effettuati secondo il protocollo DTC e del protocollo AS ha mostrato i migliori confronti. I dettagli sono espressi nella tabella 4.

Protocolli DTC e AS	K di Cohen	
Ittiofagi nuotatori	0,45	Moderata
Ittiofagi camminatori	0,57	Moderata
Ittiofagi volatori	0,56	Moderata
Onnivori	0,68	Sostanziale
Invertebratofagi nuotatori	0,69	Sostanziale
Invertebratofagi volatori	0,57	Moderata
Probers	0,11	Minima
Peckers	0,03	Minima
Scythers	0,07	Minima
Malacofagi	0,09	Minima
Consumatori di animali terrestri	0,12	Minima
Rapaci	0,78	Sostanziale
Erbivori di superficie	0,09	Minima
Generalisti tuffatori	0,41	Moderata
Generalisti di superficie	0,17	Minima

Tabella 4

Le figure di seguito illustrano i modelli relativi alle *guild* degli uccelli ittiofagi camminatori (figura 20), degli onnivori (figura 21), degli invertebratofagi nuotatori (figura 22), dei probers (figura 23), dei malacofagi (figura 24), dei rapaci (figura 25) e dei generalisti di superficie (figura 26). I simboli verdi rappresentano il modello, quelli arancio i conteggi.

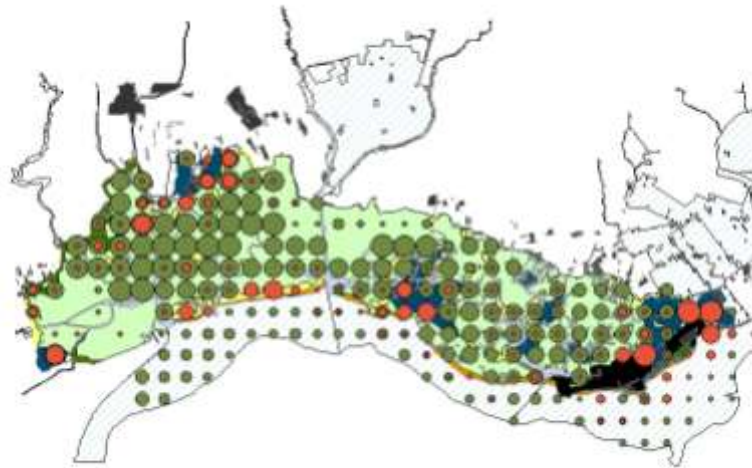


Figura 20 - Modello per gli ittiofagi camminatori (prestazione moderata)

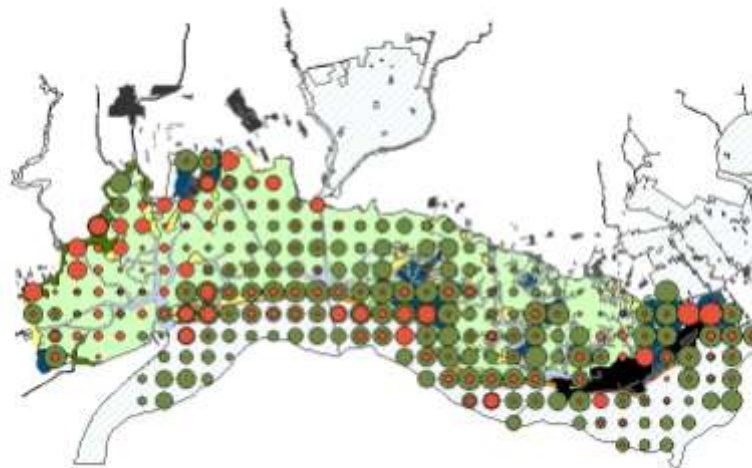


Figura 21 – Modello per gli uccelli onnivori (prestazione sostanziale)

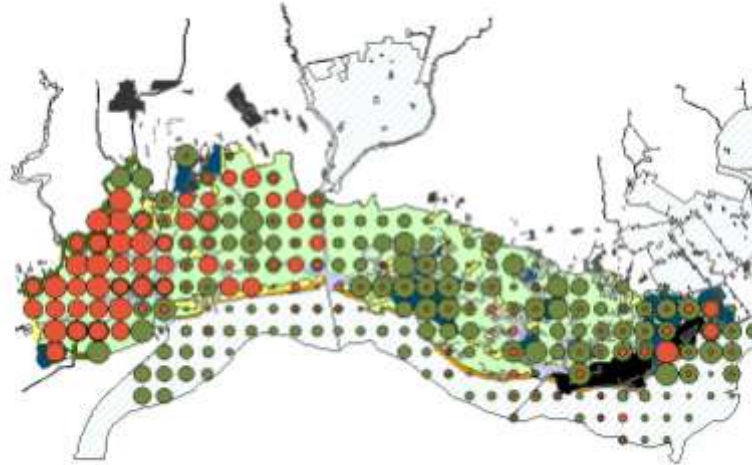


Figura 22 – Modello per gli invertebratofagi nuotatori (prestazione sostanziale)

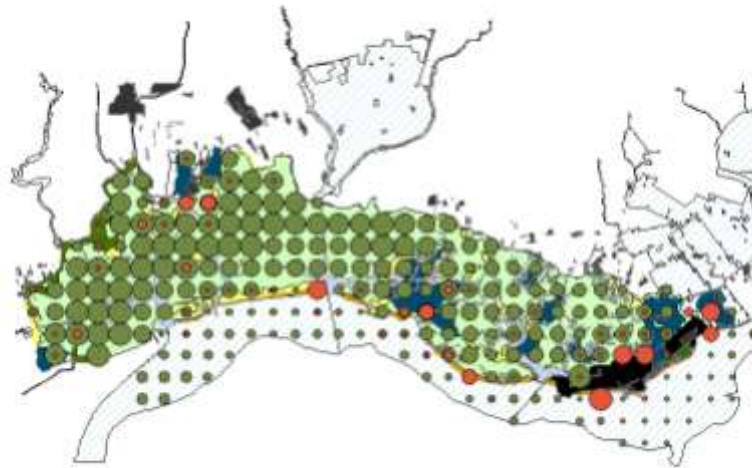


Figura 23 – Modello per i limicoli Probers (prestazione minima)

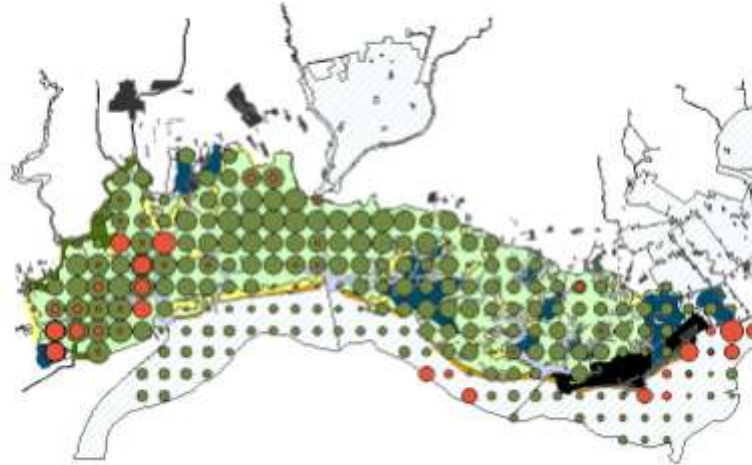


Figura 24 - Modello per i malacofagi (prestazione minima)

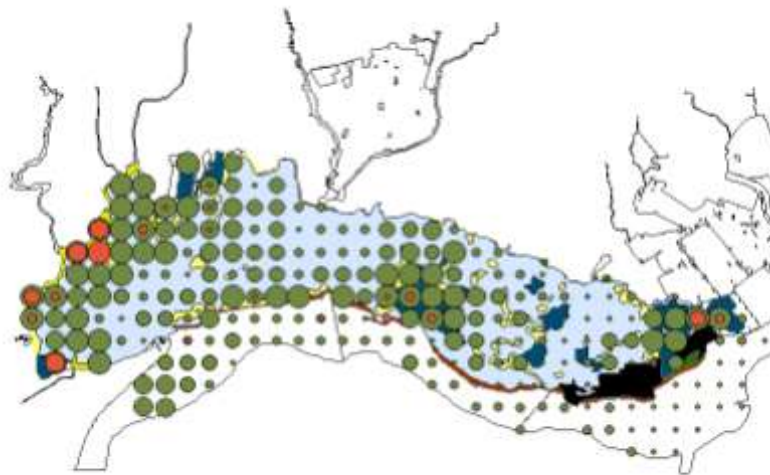


Figura 25 – Modello per gli uccelli rapaci (prestazione sostanziale)

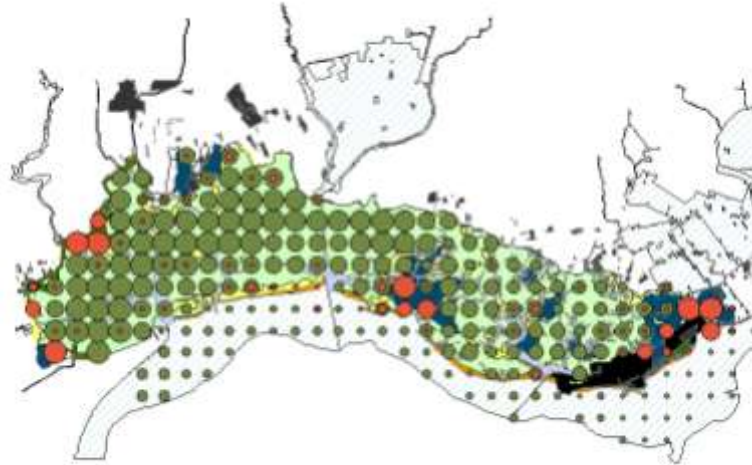


Figura 26 – Modello per i generalisti di superficie (prestazione minima)

4. Conclusioni

Il raggiungimento progressivo degli obiettivi prefissati ha consentito di risolvere i problemi legati essenzialmente alla gestione dei dati raccolti all'interno di una piattaforma progettuale ampia e complessa come quella rappresentata dal progetto ANSER.

Essi possono venire così descritti:

- L'eliminazione delle eterogeneità tra i dataset ha permesso la coesistenza di dati provenienti da differenti attività di monitoraggio dell'avifauna di palude. Quando saranno disponibili dati derivanti dall'attività di inanellamento e radiotelemetria, il modello proposto consentirà coerentemente anche la loro gestione per le relazioni tra gli attributi delle diverse entità;
- Il modello sintattico è espresso in EML. Viene così scongiurato il rischio di perdita della compatibilità con i sistemi di dati che si stanno diffondendo in questi anni e contemporaneamente con quelli più obsoleti. Il modello proposto è, tra le altre cose, fortemente integrabile con il Web, rendendo più lineare l'integrazione dello stesso con un'applicazione Web di *front-end*;
- L'integrazione in un unico modello fisico dei dati semplifica notevolmente ogni operazione di modifica della struttura;
- lo sviluppo verticale del modello garantisce la possibilità di sviluppare un sufficiente numero di quei vincoli di integrità referenziale che garantiscono un livello di qualità elevato dei dati inseriti;
- il modello semantico dei concetti coinvolti nel modello EML è tradotto in una ontologia sviluppata in linguaggio OWL-DL e si affida alla *Description logic* per ottenere una profonda espressività dei termini e una integrale autonomia di ragionamento;

- la classificazione delle classi di interesse per ANSER avviene attraverso l’inferenza logica gestita da un agente ragionatore sull’ontologia, svincolando l’utente dalla necessità di controllare la consistenza del modello a seguito di un aggiornamento delle proprietà caratterizzanti le diverse classi. Ciò consente inoltre la “leggerezza” di intraprendere anche scelte azzardate senza, per questo, correre il rischio di smarrirsi per sempre nelle proprie azioni;
- la disponibilità di una nuova classificazione, ora eseguibile con un solo click, rende utilizzabile il modello *fuzzy* contestualmente. L’aggiunta di classi, la modifica delle proprietà caratterizzanti, finanche il completo stravolgimento del modello, sono operazioni istantaneamente in vigore nel sistema inferenziale *fuzzy*;
- seppur quantitativamente limitati, i dati di censimento raccolti secondo il protocollo DTC hanno reso possibile una prima serie di evidenze legate alla definizione del modello ontologico. Infatti i valori di output del modello predittivo *fuzzy* sembrano spesso sovrastimati, soprattutto per quanto riguarda le zone di velme, indicando chiaramente la necessità di aumentare le classi di definizione coinvolte;
- malgrado l’aggiunta di due variabili di input legate alle condizioni batimetriche, quali la profondità e la pendenza medie del fondale, si nota che il loro effetto sulla previsione di distribuzione non è sensibile, soprattutto delle *guild* legate alle zone intertidali. Probabilmente è necessario considerare l’aggiunta di misure di diversificazione più accurate capaci di esprimere la distribuzione spaziale dei profili batimetrici delle zone interne alle lagune, anche in considerazione dell’effetto di cambiamento drammatico che provoca l’onda di marea quotidianamente;
- miglioramenti nella capacità predittiva potrebbero essere inoltre legati alla necessità di adattare le funzioni *membership* di input delle diverse variabili attraverso tecniche di apprendimento basate su algoritmi genetici. La funzione di *membership* per ogni variabile verrebbe individuata come la soluzione a maggiore fitness, dopo l’evoluzione, per un certo numero di generazioni successive, di una popolazione di soluzioni.

Da ultimo, ma non per importanza, l'architettura proposta in queste pagine offre la possibilità a ognuno di immaginare e realizzare, anche in completa autonomia, sviluppi futuri del modello proposto senza per questo "reinventare la ruota" dal principio. Questa considerazione sottolinea la prerogativa rilevante di riuso delle funzionalità implementate nella ragionevolezza di contribuire alla indagine scientifica attraverso il più grande, libero e variegato numero di interventi e collaborazioni.

Bibliografia

- Baader F, Bürkert H-J, Heinsohn J, Hollunder B, Müller J, Nebel B, Nutt W, Profitlich H-J
1991 “Terminological knowledge representation: a proposal for a terminological logic”
Tech. Rep. TM-90-04, Deutsches Forschungszentrum für Künstliche Intelligenz
DFKI, Kaiserslautern, Germany
- Levesque HJ 1986 “Knowledge representation and reasoning” Annual Review of Computer
Science 1:255-287
- Baader F, Hollunder B 1991 “A terminological knowledge representation system with
complete inference algorithm” In Proc. of the Workshop on Processing Declarative
Knowledge PDK-91, no.567 in Lecture Notes in Artificial Intelligence, Springer-
Verlag, pp 67-86
- Baker KS, Jackson SJ, Wanetick JR 2005 “Strategies supporting heterogeneous data and
interdisciplinary collaboration: towards an ocean informatics environment” System
Sciences. Proceedings of the 38th Hawaii International Conference on System
Sciences, pag. 219
- Bardossy A, Duckstein L 1995 “Fuzzy Rule-Based Modelling with Applications to
Geophysical, Biological and Engineering systems” CRC Press. Boca Raton. Florida.
- Beecher WJ 1942 “Nesting birds and the vegetative substrate” Chicago, IL, Chicago
Ornithological Society; Weller MW, Spatcher CE 1965 “Role of habitat in the
distribution and abundance of marsh birds” Special report No.43. Ames IA, Iowa State
Agriculture and Home Economics Experiments Station. Pagg. 1-31

- Beeri C, Levy AY, Rousset MC 1997 “Rewriting querce using views in description logics” In Proceedings of the ACM Symposium on Principles of Database Systems (PODS). Tucson, Arizona. Pagg.99-108
- Berners-Lee T 1998 “Semantic Web Road map: An attempt to give a high-level plan of the architecture of the Semantic WWW”. Draft, <http://www.w3.org/DesignIssues/Semantic.html>
- Bock W, Salski A 1998 “A fuzzy knowledge-based model of population dynamics of the Yellow-necked mouse (*Apodemus flavicollis*) in a beech forest” Ecological Modelling 108:155-161
- Borgida A, Brachman R, McGuinness D, Resnick L 1989 “CLASSIC: a structural data model for objects” In Proceedings of the ACM SIGMOD Conference. Portland Oregon. Pagg.59-67
- Brachman RJ 1979 “On the epistemological status of semantic networks” In Associative Networks, NV Findler, Ed. Academic Press
- Brachman RJ, Levesque HJ 1985 “Readings in knowledge representation” Morgan Kaufmann, Los Altos
- Brachman RJ, Pigman GV, Levesque HJ 1985 “An essential hybrid reasoning system: knowledge and symbol level accounts in KRYPTON” In Proc. of the 9th Int. Joint Conf. On Artificial Intelligence IJCAI-85
- Brachman RJ, Schmolze JG 1985 “An overview of the KL-ONE knowledge representation system” Cognitive Science 9:171-216
- Briggs DJ 1995 “Environmental statistics for environmental policy: genealogy and data quality” Journal of Environmental Management 44: 39-54
- Calvanese D 1996 “Unrestricted and finite model reasoning in Class-based Representation formalisms” PhD thesis, Dipartimento di Informatica e Sistemistica, Università di Roma “La Sapienza”
- Calvanese D, De Giacomo G, Lembo D, Lenzerini M, Rosati R 2006 “Data complexity of query answering in description logics” Proceedings in KR. Pages 260-270

- Calvanese D, De Giacomo G, Lenzerini M 1998 “What can knowledge representation do for semi-structured data?” In Proc. of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference AAAI 98 IAAI 98
- Catarci T 1999 “Web-based information access” IFCIS International Conference on Cooperative Information Systems CoopIS. Pagg. 10-19
- Catarci T, Lenzerini M 1993 “Representing and using interschema knowledge in cooperative information system” Journal of Intelligent and Cooperative Information Systems. Pagg. 55-62
- Commission of European Community, 1991 “CORINE Biotopes manual, habitats of the European Community. A method to identify and describe consistently sites of major importance for nature conservation” EUR 12587/3.
- Cohen J 1960 “A coefficient of agreement for nominal scales” Educ. Psychol. Meas. 20:37-46
- Colwell MA, Dodd SL 1997 “Environmental and habitat correlates of pasture feeding by non breeding shorebirds” Condor 99:337-344
- Daunicht W, Salski A, Nöhr P, Neubert C 1996 “A fuzzy knowledge-based model of the annual production of Skylarks” Ecological Modelling 85:67-74
- De'ath G and Fabricius KE, 2000 “Classification and regression trees: a powerful and yet simple technique for ecological data analysis”. Ecology 81, 3187-3192
- Donini FM, Lenzerini M, Nardi D, Schaerf A 1994 “Deduction in concept languages: from subsumption to instance checking” Journal of Logic and Computation 4:423-452
- Droesen WJ 1996 “Formalisation of ecohydrological expert knowledge applying fuzzy techniques” Ecological Modelling 85:75-81
- Fegraus E, Andelman SJ, Jones MB, Schildhauer M 2005 “Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation” Bulletin of the Ecological Society of America 86:158-168

- Fielding SH Ed. 1999 "Machine learning methods for ecological applications". Kluwer Academic, Dordrecht, The Netherlands
- Finlayson CM, Van Der Valk AG 1994 "Classification and Inventory of the World's Wetlands". *Vegetatio* 118:1-192
- Guegan JF, Lek S, Oberdorff T 1998 "Energy availability and habitat heterogeneity predict global riverine fish diversity" *Nature* 391: 382-384
- Gopal B, Kvett J, Löffler H, Masing V, Patten BC 1990 "Definition and classification" In Patten BC, Jorgensen SE, Dumont HJ, Gopal B, Koryavov P, Kvet J, Löffler H, Sverizhev Y, Tundisi JG Eds., *Wetlands and Continental Shallow Waterbodies, Vol. 1. Natural and Human Relationships*, SPB Academic Publishing, Amsterdam. Pagg. 9-15
- Halevy A, Rajaraman A, Ordille J 2006 "Data integration: the teenage years". VLDB06, Seoul, Korea. 9-16
- Hall LS, Krausmann PR, Morrison ML 1997 "The habitat concept and a plea of standard terminology" *Wildlife Society Bulletin* 25: 173-182
- Harold ER, Means WS 2001 "XML in a nutshell: a desktop quick reference" Petrycki L & Posner J Eds. O'Reilly & Associates, Sebastopol, CA, USA
- Hayes PJ 1979 "The logic of frames" In *Frame Conceptions and Text Understanding*. Metzger D Ed. Walter de Gruyter and Co.
- Hirota K, Pedrycz W 1996 "Directional fuzzy clustering and its application to fuzzy modeling" *Fuzzy Sets and Systems* 80:315-326
- Holling CS 2000 "Two cultures of ecology" *Conservation Ecology* 2 <http://www.consecol.org/vol2/iss2/art4/>
- Huang GH, Chang NB 2003 "Perspectives of Environmental Informatics and System Analysis" *Journal of Environmental Informatics* 1:1-6
- Kaczmarek TS, Bates R, Robins G 1986 "Recent developments in NIKL" In *Proc. of the 5th Nat. Conf. On Artificial Intelligence AAAI-86*

- Kaminski RM, Prince HH 1984 “Dabbling duck-habitat associations during spring in Delta Marsh, Manitoba” *Journal of Wildlife Management* 10:37-50
- Kolar CS, Lodge DM 2002 “Ecological predictions and risk assessment for alien fishes in North America” *Science* 298:1233-1236
- Kosko B 1994 “The probability Monopoly” *IEEE Transactions on Fuzzy System* 2:32-33
- Knublauch H, Ferguson RW, Noy NF, Musen MA 2004 “The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications”. Third International Semantic Web Conference, Hiroshima, Japan. Pagg. 1-15
- Landis JR, Koch GG 1977 “The measurements of observer agreement for categorical data” *Biometrics* 33: 159-174
- Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S 1996 “Application of neural network to modelling nonlinear relationships in ecology”. *Ecological Modelling* 90: 39-52
- Lenzerini M 2002 “Data integration: a theoretical perspective” In *Proceedings of the ACM Symposium on Principles of Database Systems PODS*
- Levesque HJ 1986 “Knowledge representation and reasoning” *Annual Review of Computer Science* 1:255-287
- Levy AY 2000 “Logic-based techniques in data integration” In Jack Minker Ed. *Logic-based Artificial Intelligence*. Kluwer Academic Publisher. Pagg. 575-595
- Levy AY, Rajaraman A, Ordille JJ 1996 “Querying heterogeneous information sources using source description” In *Proceedings of the International Conference on very large databases VLDB*
- Levy A, Rousset MC 1998. “Combining Horn rules and description logics in carin”. *Artificial Intelligence* 104: 165-209
- Li BL 1996 “Fuzzy modelling in ecology” *Ecological Modelling*. Special Issue 90
- MacArthur RH, Wilson EO 1967 “The theory of island biogeography” Princeton, NJ, Princeton University Press

- MacGregor R, Bates R 1987 “The Loom knowledge representation language” Tech. Rep. ISI/RS-87-188, University of Southern California, Information Science Institute, Marina del Rey, Cal.
- Marchini A 2004 “Logica fuzzy: prospettive di applicazione nella ricerca ecologica” In *Ecologia. Atti del XIII Congresso Nazionale della Società Italiana di Ecologia* (Como, 8-10 settembre 2003). Casagrandi R & Melià P Eds. Aracne, Roma
- Meredith AL, Edwards JL, Nielsen ES 2000 “Biodiversity Informatics: the challenge of rapid development, large databases, and complex data” *Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt.* 729-732
- Minsky M 1981 “A framework for representing knowledge” In *Mind Design*. Haugeland J Ed. The MIT Press
- Olden JD, Jackson DA 2001 “Fish-habitat relationships in lakes: gaining predictive and explanatory insight by using artificial neural networks” *Transactions of the American Fisheries Society* 130: 878-897
- Olden JD, Jackson DA 2002 “A comparison of statistical approaches for modelling fish species distributions” *Freshwater Biology* 47: 1976-1995
- Paquette GA, Ankney CD 1996 “Wetland selection by American Green-winged Teal in British Columbia” *Condor* 98:27-37; Swift BL, Larson JS, DeGraaf RM 1984 “Relationship of breeding bird density to habitat variables in forested wetlands” *Wilson Bulletin* 96: 48-59
- Patel-Schneider PF 1984 “Small can be beautiful in knowledge representation” In *Proc. of the IEEE Workshop on Knowledge-Based Systems*
- Pearl J 1988 “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”. San Mateo, California, Morgan Kaufmann Publishers, Inc.
- Primack R 2004 “*A Primer of Conservation Biology*” Third Edition . Sinauer Associates, Sunderland, MA
- Quantz J, Kindermann C 1990 “Implementation of the BACK system version 4” Tech Rep. KIT-Report 78, FB Informatik, Technische Universität Berlin, Berlin, Germany

- Quillian MR 1967 “World concepts: a theory and simulation of some basic capabilities”
Behavioral Science 12: 410-430
- Ramsar Convention Bureau 2000 “The Ramsar Convention on Wetlands”
<http://www.ramsar.org/>
- Salski A 1992 “Fuzzy knowledge-based models in ecological research” Ecological Modelling
63: 103-112
- Salski A, Fränzle O, Kandzia P 1996 “Fuzzy Logic in Ecological Modelling” Ecological
Modelling, special issue 85(1)
- Salski A 1999 “Ecological modeling and data analysis” In Practical Applications of fuzzy
technologies. Hans-Jürgen Zimmerman Ed. Kluwer Academic Publisher,
Massachusetts, USA
- Salski A 1999a “Fuzzy logic approach to data analysis and ecological modeling”. Proc. of
European Symposium on Intelligent Techniques (ESIT'99), Orthodox Academy of
Create, Greece: CD-ROM BC-01
- Salski A 2006 “Ecological applications of fuzzy logic”. In Ecological informatics: scope,
techniques and applications. Recknagel F Ed., Springer, Berlin, Germany
- Schaerf A 1994 “Reasoning with individuals in concept languages” Data and Knowledge
Engineering 13:141-176
- Schaerf A 1994 “Query answering in concept-based knowledge representation systems:
algorithms, complexity and semantic issues” PhD thesis, Dipartimento di Informatica
e Sistemistica, Università di Roma “La Sapienza”
- Shafer, G. (1990), “The Unity of Probability,” in Acting Under Uncertainty: Multidisciplinary
Conceptions, von Furstenberg G Ed., New York: Kluwer, pp.95-126
- Short HL, Burhnam KP 1982 “Technique for structuring wildlife guild to evaluate impacts
on wildlife communities” Special Scientific Report – Wildlife 44. Washington, DC, US
Fish and Wildlife Service

- Short HL 1989 "A wildlife habitat model for predicting effects on human activities on nesting birds" In Freshwater wetlands and wildlife. Sharitz RR, Gibbons JW Eds. Washington, DC, Department of Energy. Pagg.957-973
- Silvert W 2000 "Fuzzy indices of environmental conditions" Ecological Modelling 130:111-119
- Southwood TRE 1977 "Habitat, the template for ecological strategies?" Journal of Ecology 46:337-365
- Spence DHN 1982 "The zonation of plants in freshwater lakes" Advances in Ecological Research 12:37-125
- Steven J 2001 "La nuova scienza dei sistemi emergenti. Dalla colonia di insetti al cervello umano, dalla città ai videogame e all'economia, dai movimenti di protesta ai network" Garzanti, Milano
- Swift BL, Larson JS, DeGraaf RM 1984 "Relationship of breeding bird density to habitat variables in forested wetlands" Wilson Bulletin 96: 48-59
- Toivonen T, Vieno M, Kumpulainen R 2005 "Sharing and Exchange of observational data sets: exchange standards and good databasing practices". European Network for Biodiversity Information ENBI. <http://enbi.utu.fi>
- Weller MW 1988 "The influence of hydrologic maxima and minima on wildlife habitat and production values of wetlands" In Proceedings of the National Symposium on Wetland Hydrology. Kusler JA, Brooks G Eds. Association of State Wetland Managers, Berna, NY. Pagg. 55-60
- Weller MW 1994 "Freshwater marshes: ecology and wildlife management" 3rd edn. University of Minnesota Press, Minneapolis, MN
- Weller MW 1999 "Wetland birds: Habitat Resources and Conservation Implications". University Press, Cambridge, UK. Pagg. 1-271
- White S 2005 "Better computational descriptions of science" Scientific Computing World, http://www.scientific-computing.com/features/feature.php?feature_id=37

Williams JB, Poff NL 2006 “Informatics software for the ecologist’s toolbox: a basic example” *Ecological Informatics* 1:325-329

Woods WA 1975 “What’s in a link: foundations for Semantic Networks” In *Representation and understanding: studies in cognitive science*, DG Bobrow and AM Collins, Eds., Academic Press

Zadeh LA 1965 “Fuzzy sets” *Information and Control* 8:338-353

Zadeh LA 2002 “Toward a perception-based theory of probabilistic reasoning with imprecise probabilities” *Journal of Statistical Planning and Inference* 105:233–264