

IL NETWORK NAZIONALE DELLA BIODIVERSITÀ: ACCESSO E
CONDIVISIONE DEI DATI IN RETE

STEFANO MARTELLOS*¹, STEFANO DE FELICI², FABIO ATTORRE³,
CARLO FORTUNATO⁴, VALERIO SBORDONI²

1. Dipartimento di Scienze della Vita, Università degli Studi di Trieste,
Via L.Giorgieri 10, 34127, Trieste; 2. Dipartimento di Biologia,
Università degli Studi di Roma Tor Vergata, Via della Ricerca
Scientifica, 00133 Roma; 3. Dipartimento di Biologia Ambientale,
Università La Sapienza, P.le A. Moro 5, 00185, Roma; 4. Ministero
dell'Ambiente e della Tutela del Territorio e del Mare, Via Cristoforo
Colombo 44, 00154 Roma

*Corresponding author. Email: martelst@units.it

L'aggregazione dei dati di biodiversità in grandi database federati è una delle priorità della ricerca nel campo della Informatica della Biodiversità (Biodiversity Informatics). Questa branca della Biologia si è evoluta in modo particolare negli ultimi 15 anni, specialmente in seguito alla nascita del Global Biodiversity Information Facility (GBIF) e del Biodiversity Collections Access Service for Europe (BioCASE). Questi due grandi aggregatori di dati primari di biodiversità veicolano ad oggi quasi 400 milioni di record, sui circa 2,5 miliardi che si stima esistano nel mondo.

L'Italia è uno tra i paesi più ricchi al mondo di collezioni scientifiche, distribuite in numerosi Musei di Storia Naturale, presenti anche in città di medie e piccole dimensioni, nelle Università e in altri Centri di Ricerca. Questa situazione frammentata ha reso indispensabile cercare di aggregare tutti i dati, al fine di renderli fruibili per attività di ricerca e di conservazione.

Per questo motivo, il Ministero dell'Ambiente e della Tutela del Territorio e del Mare (MATTM) ha avviato nel 2005 il progetto

“Sistema Ambiente 2010”, che, successivamente, si è trasformato in un vero e proprio programma, tuttora in corso, articolato in diversi progetti di innovazione tecnologica e scientifica integrati tra loro, inseriti nella Strategia Nazionale per la Biodiversità 2010-2020 (Convenzione delle Nazioni Unite sulla Diversità Biologica - CBD) e nell’Agenda Digitale Italiana (Agenda Digitale Europea - Strategia Europa 2020), in coerenza con le politiche internazionali, comunitarie e nazionali in tema di ricerca e innovazione. Uno dei risultati di questo programma è stato l’istituzione del “Network Nazionale della Biodiversità” (NNB), accessibile in rete tramite il portale “Natura Italia” (un altro dei risultati di “Sistema Ambiente 2010”) all’indirizzo <http://www.naturaitalia.it>.

Il Network, oltre a rendere disponibile documentazione sulla biodiversità e sulla sua tutela nella forma, richiesta dalla CBD e dalla Ue, di un Clearing House Mechanism (CHM), aggrega ed organizza, in due diversi database federati, realizzati, laddove possibile, secondo la Direttiva comunitaria 2007/2/EC (Infrastructure for Spatial Information Information in the European Community - INSPIRE):

- a) dati primari di biodiversità, ovvero record di osservazioni floristiche e faunistiche, e dati di campioni delle collezioni di storia naturale,
- b) dati legati alle specie, ovvero tutte le informazioni che non fanno parte della precedente categoria (dati ecologici, morfologici, sequenze di DNA, ecc.).

I database federati, pur essendo difficili da mantenere e gestire, hanno l’indubbio vantaggio di consentire a ciascun data-provider di mantenere l’assoluto controllo e la proprietà dei dati. Nessuno, a parte il proprietario, infatti, può modificarli o spostarli dal server d’origine.

Il database federato di dati primari utilizza il protocollo BioCASE ed Access Biological Collection Data (ABCD) come Concept Mapping File (CMF).

BioCASE è stato scelto per la semplicità di installazione e configurazione, e per la sua indipendenza dal GBIF, richiesta da

“Sistema Ambiente 2010”. ABCD consente di veicolare dati anche complessi, come quelli associati ai campioni presenti nelle collezioni di Storia Naturale. Questa scelta non preclude la possibilità di una aggregazione dei dati nel network GBIF. Infatti, il network BioCASE è a sua volta un data provider del GBIF.

La ricerca viene eseguita mediante un’interfaccia di interrogazione, che restituisce l’elenco dei record disponibili organizzati per database di origine. In presenza di dati georeferenziati, questi possono essere visualizzati sul sistema di WebGIS integrato nel portale “Natura Italia”, che è reso disponibile dal “Geoportale Nazionale” (INSPIRE) del MATTM.

Il secondo database federato, che organizza dati legati alle specie, non utilizza protocolli e CMF preesistenti, ed è stato sviluppato interamente nell’ambito del programma “Sistema Ambiente 2010”. Il database si basa su un approccio pragmatico, che mette al centro le cosiddette “taxon page”. Ogni database di dati di biodiversità, normalmente, organizza le informazioni per taxon. Quando un utente ricerca delle informazioni, lo fa normalmente immettendo il nome di un taxon, ed il risultato è una scheda a questo dedicata, che riporta le informazioni disponibili nel database. Sfruttando questa caratteristica, il sistema permette di ricercare i metadati delle taxon page presenti in tutti i database federati tramite un’unica interfaccia di interrogazione, ed organizza i risultati per categorie di organismi e di contenuti. Per ogni risultato della ricerca sono riportati i crediti del data provider e degli autori. I nodi federati del sistema sono interrogati direttamente solo quando l’utente decide di accedere alla singola taxon page.

Il “Network Nazionale della Biodiversità”, istituito formalmente nel 2012 ed entrato a regime all’inizio del 2013, organizza, ad oggi, più di un milione di record di dati primari, grazie all’aggregazione di 41 banche dati di diversi data-provider, e circa 250.000 taxon page di dati legati alle specie. Si tratta evidentemente di un risultato rilevante, anche

considerato che a questo scopo sono stati utilizzati finora esclusivamente i database posseduti dai partner del primo progetto di “Sistema Ambiente 2010”, dedicato alla sperimentazione degli apparati tecnologici. Oggi, il NNB deve trasformarsi da sistema sperimentale a concreto strumento a disposizione della comunità scientifica, politico-amministrativa e generalista italiana e internazionale, che potrà attingere alle risorse in esso contenute e contribuire attivamente alla sua crescita.

Nonostante il sistema sia perfettamente a regime, il suo massimo potenziale è ancora ben lungi dall’essere raggiunto. Infatti, le performance in aggregazione dei dati e nella produzione dei risultati, già di buon livello, potranno essere ottimizzate con il perfezionamento di alcuni aspetti tecnologici del sistema, quali, ad esempio:

1. la realizzazione di un sistema di caching per ridurre i tempi di search/retrieval e ovviare ai (momentanei) disservizi dei nodi della rete;
2. il miglioramento del portale di ricerca, che consenta – al pari del portale principale del progetto BioCASE – l’interrogazione tramite un numero di campi molto maggiore dell’attuale.