

LE STRUTTURE DI DATABASE COME ELEMENTO DEL SISTEMA DI QUALITÀ IN AGROMETEOROLOGIA

Luca Bechini

Istituto di Agronomia, Università degli Studi di Milano

Riassunto

Vengono esposti i risultati di una ricerca bibliografica preliminare relativa alla disponibilità di strutture di database esistenti per dati agronomici, a scala di campo e a scala territoriale.

Abstract

The aim of this paper is the analysis of the results of a preliminary bibliographical survey about the availability of standard database structures useful for agronomic data.

Introduzione

Alcuni lavori di ricerca coordinati da docenti dell'Istituto di Agronomia di Milano (prof. Tommaso Maggiore, prof. Renato Benati e dott. Stefano Bocchi) hanno richiesto lo svolgimento di una ricerca sulle strutture di database esistenti per dati agronomici, a scala di campo e a scala territoriale.

In particolare si è di recente conclusa una ricerca bibliografica preliminare sull'argomento che viene presentata in questa sede.

Obiettivi del lavoro

Gli obiettivi finali del lavoro che si sta conducendo presso l'Istituto di Agronomia di Milano si riferiscono a due linee di ricerca:

- ✓ da un lato, si vuole mettere a punto e utilizzare un database per la gestione dei dati raccolti nel corso di prove sperimentali. Gli scopi sono quelli di rendere più efficiente la raccolta, l'immissione e l'elaborazione di tali dati. Un aspetto particolare di tale attività riguarda l'uso di modelli di simulazione; ci aspettiamo che il database faciliti l'interazione tra sperimentatori e utilizzatori di modelli. Tale attività nasce dalla necessità di avere uno standard per la conservazione dei dati di campo, che attualmente non esiste. La conseguenza della mancanza di un formato standard per la conservazione dei dati è che i dati stessi

non sono omogenei fra di loro per molti aspetti (a titolo di esempio si possono citare i codici descrittivi della metodologia utilizzata in campo, le modalità di incolonnamento e le unità di misura) e che si conservano molte informazioni secondarie ricavabili da altre, che è inutile salvare permanentemente nell'archivio. Il risultato finale è che il trasferimento dei dati da/verso un modello di simulazione o da/verso un programma di statistica è lento, soggetto ad errori e farraginoso;

- ✓ dall'altro lato, il Parco Agricolo Sud Milano sta per iniziare un'attività di censimento delle attività agricole nel suo territorio di competenza. L'Istituto di Agronomia sarà coinvolto tra l'altro nella definizione del database, che servirà come fonte di dati per elaborazioni di diverso tipo, da semplici statistiche a indici di sostenibilità basati sull'efficienza dei fattori produttivi a scala aziendale, per arrivare all'applicazione di modelli di simulazione a scala territoriale.

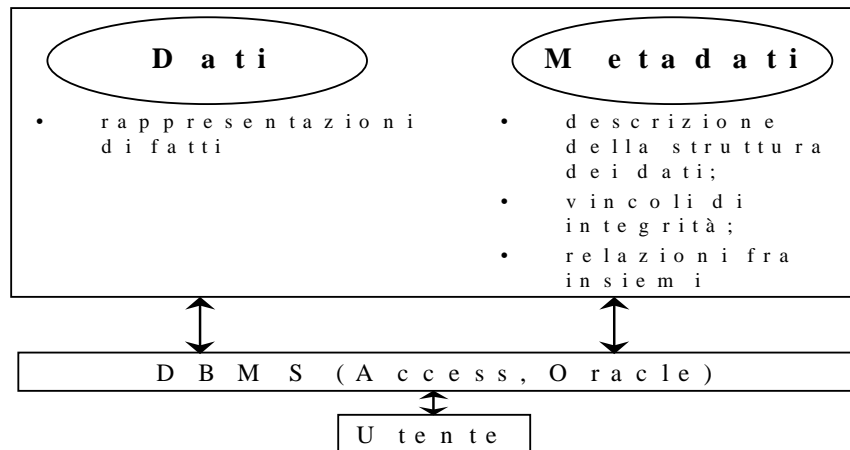
Perché utilizzare un database

Perché abbiamo scelto di utilizzare un database e di non continuare con il formato "libero" usato finora? Per rispondere a questa domanda è utile ricordare che nella terminologia dei database i *dati* vengono sempre distinti dai *metadati*. I dati costituiscono delle "rappresentazioni di fatti", mentre i metadati descrivono la struttura dei dati, pongono vincoli di integrità sugli stessi (restrizioni sui valori ammissibili), e stabiliscono relazioni fra gli insiemi (Figura 1). Immagazzinare i dati sperimentali come si è fatto finora significava concentrarsi sui dati e non sulla loro struttura, generando i problemi già segnalati. Vediamo invece quali vantaggi si conseguono usando un database (Albano *et al.*, 1997):

- ✓ gli stessi dati possono essere usati da diverse applicazioni, senza doverne creare più copie. L'unica copia esistente sarà sempre aggiornata;
- ✓ il DBMS (*DataBase Management System*, il programma che fa da interfaccia tra l'utente e i dati, ad esempio Microsoft Access o Oracle) può effettuare controlli sull'integrità dei dati;

- ✓ le diverse applicazioni che usano i dati vi si riferiscono usando gli stessi nomi (standardizzazione dei dati);
- ✓ oltre che alle applicazioni, i dati sono accessibili all'utente tramite interfaccia;
- ✓ i dati vengono protetti da eventuali avarie del sistema; inoltre ne viene garantita la sicurezza e la condivisione da parte di più utenti;
- ✓ nel caso della modellistica, l'interazione tra i dati e il modello è semplificata perché basta scrivere una volta per tutte un programma che estrae/importa i dati per garantire successivamente una comunicazione costante.

Figura 1 - Dati e metadati



Strutture di database esistenti

Poiché definire *ex-novo* i metadati per sperimentazioni agronomiche avrebbe richiesto uno sforzo concettuale notevole, si è preferito effettuare un'indagine conoscitiva su approcci esistenti. Il metodo

per la ricerca è stato quello di richiedere informazioni agli esperti; in particolare, si è mandato un messaggio alla lista di discussione AGMODELS-L. Si riassumono brevemente le numerose risposte ricevute.

Le principali tematiche affrontate nella discussione sono state:

- ✓ la struttura sviluppata dall'ICASA (progetto DSSAT);
- ✓ quale formato dei dati utilizzare;
- ✓ altre strutture di database.

Il database ICASA

La struttura ICASA (<http://agrss.sherman.hawaii.edu/icasa/new/ispec.html>) è stata finora sviluppata in formato ASCII come parte dei file di input ai modelli del gruppo DSSAT. Essa comprende alcune informazioni generali sulla sperimentazione (nome dell'esperimento, fattori, personale, istituto, pubblicazioni) e dati di dettaglio. La tabella principale descrive i trattamenti applicati nell'esperimento, indicando dei codici per i livelli dei fattori di variazione del disegno sperimentale. Nell'esempio di Tabella 1 fattori come la cultivar (CU), l'appezzamento (FL), l'analisi chimica del profilo (SA), le condizioni iniziali (IC), i dettagli relativi alla semina (MP), la fertilizzazione organica (MR), le lavorazioni (MT), i trattamenti antiparassitari/erbicidi (MC), le modificazioni ambientali in serra (ME), la data di raccolta (MH) sono costanti per tutti i trattamenti; l'irrigazione (MI) e la fertilizzazione (MF), invece, sono differenti. In particolare, esistono due livelli di irrigazione (codificati con 0 e 1) e tre livelli di fertilizzazione (1, 2, 3). L'incrocio dei livelli dei due fattori genera quindi sei trattamenti.

Ogni livello di ogni fattore è poi descritto in dettaglio in tabelle successive. Ad esempio, per i trattamenti fertilizzanti (Tabella 2) si indicano le date di distribuzione degli stessi, la loro profondità di incorporamento, e la quantità di elementi distribuita per unità di superficie.

I limiti principali di questa struttura, da noi individuati finora, sono:

- ✓ non è sufficientemente documentata; molti degli acronimi non sono descritti nella documentazione ufficiale, non esistono

esempi articolati che possano suggerire la gestione di situazioni complesse;

- ✓ non è chiaro come includere le informazioni sulle rotazioni;

Tabella 1-Esempio di tabella TREATMENTS del database ICASA

*TREATMENTS										
FACTOR LEVELS-----										
@TN>	R>	O>	C>	TNAME.....	CU	FL	SA	IC	MP	MI
MF	MR	MC	MT	ME	MH					
	1	1	1	1	0-N;DRYLAND	1	1	1	1	0
1	1	1	1	1	1					
	2	1	1	1	60-N;DRYLAND	1	1	1	1	0
2	1	1	1	1	1					
	3	1	1	1	180-N;DRYLAND	1	1	1	1	0
3	1	1	1	1	1					
	4	1	1	1	0-N;IRRIGATED	1	1	1	1	1
1	1	1	1	1	1					
	5	1	1	1	60-N;IRRIGATED	1	1	1	1	1
2	1	1	1	1	1					
	6	1	1	1	180-N;IRRIGATED	1	1	1	1	1
3	1	1	1	1	1					

Tabella 2 - descrizione dei trattamenti fertilizzanti nel database ICASA

*FERTILIZERS (INORGANIC)									
@MF>	FYR>	FDOY>	FMCD	FACD	FDEP	FAMN	FAMP		
FAMK	FAMC	FAMO	FOCD						
	1	1981	289	FE001	-99	15	0	-99	-
99	-99	-99	-99						
	2	1981	289	FE001	-99	15	60	-99	-
99	-99	-99	-99						
	3	1981	289	FE001	-99	15	90	-99	-
99	-99	-99	-99						
	3	1982	056	FE001	-99	1	90	-99	-
99	-99	-99	-99						

- ✓ non è chiaro come includere i dati misurati (i risultati della prova);
- ✓ non è prevista l'aggiunta di codici di qualità del dato.

Nonostante questi limiti, si tratta di uno degli approcci più completi finora reperiti. Per tale motivo, questa struttura è stata implementata in Access da Barry Jacobson (Cargill Central Research, Minneapolis) e viene studiata come una possibile base di partenza (migliorabile) da un gruppo di ricerca all'AB-DLO di Wageningen (Olanda), guidato da Frits van Evert. Noi abbiamo deciso di approfondire la conoscenza di tale struttura e di farne il punto di partenza del nostro lavoro.

Il formato dei dati

La seconda linea di discussione verteva invece sul tema del formato dei dati. Si sono confrontati due approcci:

- il primo prevede di utilizzare il formato ASCII: questo approccio è stato finora seguito dall'ICASA, che ha anche prodotto apposite subroutine FORTRAN per la gestione dei dati in tale formato; tale approccio è stato tuttavia criticato perché la gestione di dati organizzati in uno schema relazionale viene effettuata molto meglio con l'uso di un programma apposito;
- il secondo prevede l'inserimento dei dati in opportuni programmi DBMS, che garantiscono maggiori funzionalità; tuttavia, con tale soluzione lo scambio di dati fra ricercatori che usano DBMS diversi viene reso molto più difficile: Barry Jacobson ritiene quindi che la soluzione ottimale sia quella di usare il formato del proprio DBMS per conservare i dati e di usare il formato ASCII per scambiarli con altri colleghi.

Altre strutture di database agronomici o ambientali

La terza linea di discussione ha riguardato la segnalazione di altre strutture di database agronomici o ambientali, che, pur se meno complete, possono essere utili agli scopi che ci prefiggiamo. Eccone un elenco schematico:

- ✓ Macro-DB, di Nick Jarvis: si tratta di un sistema di supporto alle decisioni per simulare il destino ambientale dei fitofarmaci, che consiste di tre database (suoli, fitofarmaci, clima), collegati a routine di stima dei parametri e al modello di simulazione MACRO. Maggiori informazioni sono disponibili al sito <http://130.238.110.134:80/bgf/Macrohtm/macro.htm>
- ✓ Universal Environmental Database (UED), di Roger Nelson: si tratta di un database ancora provvisorio, che consentirebbe di immagazzinare serie temporali a diverse scale, documentando le unità di misura, la cadenza e la qualità dei dati (dato reale, inaffidabile, previsto, stimato, generato, loro combinazioni). Note esplicative si possono trovare all'indirizzo internet <http://www.bsyse.wsu.edu/~rnelson/ued/ued.htm>;
- ✓ Agroecological data-information system del collaborative agroecosystem research project (a Monaco, Germania). Documentato in tedesco al sito <http://www.gsf.de/FAM/adis.html>;
- ✓ Ecological information system (keris), un sistema per accedere ai dati organizzati in serie temporali o a dati georeferenziati. Documentato in tedesco al sito <http://www.pz-oekosys.uni-kiel.de/keris/>;
- ✓ il formato di file per dati scientifici HDF (Hierarchical Data Format della NASA), sviluppato per immagazzinare e organizzare strutture di dati comuni in campo scientifico. Nonostante il sistema sembri utile, non sono attualmente disponibili applicazioni agroambientali. E' documentato al sito http://daac.gsfc.nasa.gov/REFERENCE_DOCS/HDF/gdaac_hdf.html;
- ✓ ICIS - International Crop Information System è un database standard in corso di sviluppo da parte dei centri di ricerca internazionali del CGIAR (Consultative Group for International Agricultural Research). E' documentato al sito <http://www.cgiar.org/ICIS/chapter27.htm>.

Servizi on-line

Diverso da quelli trattati finora, ma importante per le potenzialità, è il servizio offerto da società private ad agricoltori statunitensi. Tali

servizi gestiscono molti dati ottenuti da fonti diverse e li integrano in modo da fornire via Internet agli agricoltori diverse informazioni localizzate, in tempo reale e fra loro integrate:

- ✓ immagini telerilevate;
- ✓ dati meteorologici, previsioni del tempo;
- ✓ previsioni di crescita delle colture;
- ✓ previsioni di sviluppo di patogeni;
- ✓ riepilogo dei dati per ogni appezzamento (mappe di resa, analisi del terreno, gestione agronomica);
- ✓ assistenza in campo.

Un esempio di tale tipo di servizio è quello fornito da Mpower (<http://www.mpower3.com>).

Conclusioni

Dalla ricerca bibliografica effettuata emerge in modo evidente l'assenza di approcci consolidati per l'archiviazione di dati agronomici. Questo è vero sia per i dati sperimentali (per i quali tuttavia il formato ICASA ha già interessanti requisiti di completezza), sia per i dati territoriali (in modo più marcato).

Tale stato di fatto costituisce un preciso limite per la condivisione dei dati agronomici e più in generale dei dati biologici di rilevanza ambientale fra le strutture di ricerca e di servizio.

Si individua dunque un vasto campo d'indagine e d'azione che merita senz'altro l'interessamento da parte di coloro che si occupano di agrometeorologia.

Bibliografia

Albano, A., G. Ghelli, R. Orsini, 1997. *Basi di dati relazionali e a oggetti*, Zanichelli, 416 pp.