

# CLIMAGRI LT: UN META-DATABASE RELAZIONALE PER LA GESTIONE DEI DATI DI ESPERIMENTI AGRONOMICI DI LUNGA DURATA

Zuliani M., Peressotti A., Zerbi G., Zuliani G., Delle Vedove G. e Danuso F.

Dipartimento di Scienze Agrarie e Ambientali, zerbi@uniud.it

## Riassunto

I dati delle prove agronomiche di lunga durata sono potenziali punti di partenza per lo sviluppo di modelli di sistemi colturali volti, tra l'altro, a indagare sugli effetti dei cambiamenti climatici sull'agricoltura. In Italia esistono molte esperienze di lunga durata ma l'utilità dei loro dati è fortemente ridotta dalla loro scarsa disponibilità e dalla mancanza di informazioni sulla loro organizzazione. In questo lavoro viene presentata la filosofia di base e il layout generale di un meta-database (ClimagriLT) progettato per organizzare e mettere a disposizione dei ricercatori interessati dati utili alla modellizzazione degli effetti di variabili (trattamenti, dati pedologici e meteorologici etc.) con particolare attenzione sia alla interfaccia del database che alla politica di accesso alle informazioni.

## Introduzione

L'utilità ai fini modellistici degli esperimenti agronomici di lungo termine è fortemente limitata dalle scarse informazioni sulla loro organizzazione e sui metodi di raccolta dei dati. Mancano altresì standard comuni sulla loro raccolta e gestione ( Hunt 1998 e Hunt et al. 1998). Spesso esistono inconsistenze dovute a cambiamenti in corso d'opera dei disegni e dei protocolli sperimentali, perdite di dati, cambiamenti di procedure etc. La teoria dei database relazionali (Codd, 1970) offre un solido approccio teorico alla soluzione di questi problema, ed è stata adottata in questo lavoro. Si presenta pertanto la filosofia e il layout di un meta-database destinato a conservare e a distribuire dati di un insieme di esperimenti di lungo termine (trattamenti, rese, dati pedologici e meteorologici etc.). L'attenzione è focalizzata anche alla modellizzazione dei dati e alla loro politica di diffusione.

## Esperimenti di lunga durata in Italia

Una analisi degli esperimenti di lungo termine in Italia ha fatto rilevare come in Italia esistano 10 esperimenti di lungo termine (con tale termine si intendono esperimenti che di durata superiore a 12 anni condotti sul medesimo appezzamento). Questi studi sono condotti da Istituzioni Universitarie (Padova, Pisa, Bari, Bologna, Palermo; Perugia) e da Istituti di ricerca del MIPAF (Foggia, Lodi, Bari) e costituiscono un'interessante sorgente di dati per studi agroecologici. La specie più rappresentata è il frumento tenero. I dati sono conservati su supporti molto differenziati: da veri database a quaderni di campagna.

## Il modello di ClimagriLT

Al fine di raccogliere, conservare i dati sono stati progettati e implementati un meta-database e un database integrati in una singola applicazione denominata ClimagriLT.

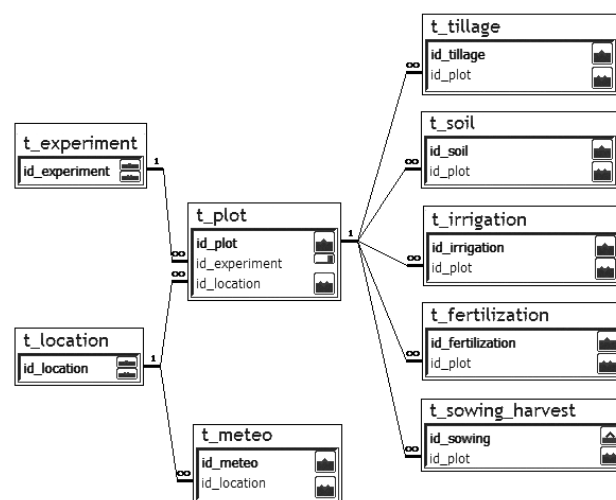
La costruzione del modello generale prevede quattro fasi successive: i) fabbisogni; ii) geometria della struttura concettuale; iii) geometria della struttura logica; iv) normalizzazione (Atzeni et al. 1993, 1996).

Fase I. fabbisogni sono una semplice lista di obiettivi che guidano il disegno del database o gli obiettivi che deve soddisfare.

Fase II. La fase "geometria della struttura concettuale" è l'identificazione delle entità che corrispondono ai gruppi di dati del "mondo reale" e alla individuazione delle relazioni tra le entità. Il concetto chiave del sistema è rappresentato dalla parcella su cui ogni trattamento della singola coltura viene di volta in volta applicato. Questo approccio differisce dalla tradizione agronomica in cui il trattamento è il più importante attributo di un esperimento.

Il passo successivo è stato la partizione in diverse entità di ogni gruppo di eventi che occorre sulla singola parcella. Semina e raccolta, fertilizzazione, irrigazione, lavorazioni, campioni di terreno diventano gruppi di dati diversi e indipendenti. La Fig. 1 riassume il concetto del modello generale.

Fig.1- nucleo del modello dati in ClimagriLT



Durante i test di implementazione del database il modello ha fatto rilevare una buona flessibilità ai diversi formati di dati di ciascun esperimento.

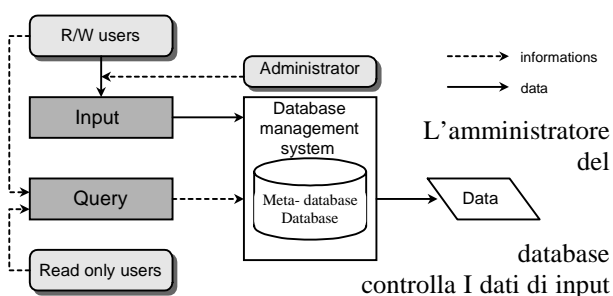
Fase III. Questa fase "geometria della struttura logica" è la traduzione della struttura concettuale in tabelle e relazioni mediante un Database Management System (DBMS). Nel primo sviluppo del progetto è stato usato MSAccess; successivamente il freeware MySQL DBMS.

Fase IV. La normalizzazione permette di controllare se la struttura logica ottenuta è coerente, libera da ridondanze, con un singolo punto di aggiornamento e cancellazione. ClimagriLT è stato normalizzato alla terza forma normale.

## Politica di gestione

ClimagriLT è quindi un'applicazione progettata per condividere dati da fonti diverse. La gestione dell'accesso ai dati è un elemento delicato e richiede una buona organizzazione. Uno schema riassuntivo di questa politica gestionale si trova in Fig. 2.

Fig 2. - Schema generale di gestione dei dati.



L'amministratore del database controlla i dati di input ed è responsabile per l'operazione dell'applicazione. In questa fase egli non è responsabile della qualità dei dati e si limita a certificare che i dati sono stati forniti da una fonte autorizzata. Gli utenti "Read-and-write" cooperano con l'amministratore nell'immissione dei dati e hanno libero accesso ai metadati e ai propri dati. La comunità scientifica (read-only users) hanno accesso ai dati solo dopo aver ottenuto un'autorizzazione dai proprietari dei dati. L'amministratore indirizza le richieste ai proprietari dei dati e, in seguito ad autorizzazione (via lettera, fax, o e-mail certificata), permette l'accesso.

## Infrastruttura attuale del meta-database

L'infrastruttura finale di ClimagriLT è rappresentata in fig. 3. ClimagriLT contiene due database: Il primo raccoglie dati grezzi (BackEnd Database), il secondo serve alle simulazioni (Simulation Server Database, SSD). SSD contiene dati elaborati, ottenuti mediante stime e interpolazioni, al fine di riempire i dati mancanti e correggere quelli fuori range a causa di errori di input. SSD è quindi anche un "mirror database" che ha il fine di proteggere la perdita di dati dovuta a malfunzionamenti hardware. Un server database sarà dedicato a ogni database. Un Web server sarà attivato nella prima fase per svolgere il compito di "database publication" (Web Server 1). Attraverso questo Web Server un utente sarà in grado di svolgere query sui metadata (e sui dati se autorizzati) e comunicare con l'amministratore e i proprietari dei dati. Ogni accesso verrà registrato. Uno switch indirizzerà la richiesta al Backend database o al SSD. I proprietari dei dati avranno accesso al Backend database per l'input e il controllo dei dati. I dati estratti

sono in formato XML che possono essere facilmente trasformati in altri formati.

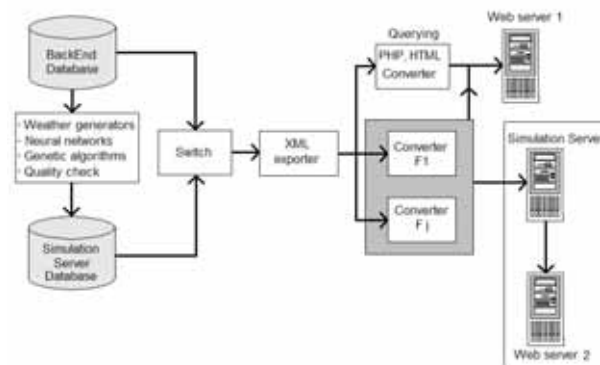


Fig.3 - Struttura generale di ClimagriLT

Al momento ClimagriLT è implementato allo stadio di "standalone database", in Microsoft Access e in MySQL. Dieci esperimenti di lungo termine sono stati già registrati.

## Obiettivi di base e sviluppi futuri

Il fine di ClimagriLT è di fornire qualcosa di utile alla comunità scientifica, producendo una singola fonte di dati capace di fornire un insieme generale e vasto di dati agronomici, pedologici e agroclimatici. L'attuale lavoro è quello relativo alla possibilità di esportare dati in formati diversi capaci di alimentare modelli di sistemi colturali quali CSS (Danuso et al., 1999), Cropsyst (Stöckle et al., 2003) e file ICASA (Hunt et al. 2001).

## Bibliografia

- Atzeni P., De Antonelli V. 1993. *Relational database theory*. Benjamin Cummings Publishing Company Inc. Redwood City. 389 pp.
- Atzeni P., Ceri S., Paraboschi S. e Torlone R., 1996. *Basi di dati*. Edizioni McGraw-Hill, Milano ed. 1999. 605 pp.
- Codd E.F., 1970. *A Relational Model of Data for Large Shared Data Banks* Communications of the ACM, Association for Computing Machinery, Inc. 13:6 377-387.
- Danuso F., Bigot L., Budoi G., Franz D., 1999. *CSS: a modular software for cropping system simulation*. Proceedings of Agroclimatology and Modelling International Symposium "Modelling Cropping Systems", June 21-23, Lleida, Spain
- Hunt L.A., 1998. *Recent attempts to evaluate and apply wheat simulation models, and to simplify the storage and exchange of experimental data*. In 'Wheat: Prospects for global improvement', eds. H.J. Braun, F. Altay, W.E.
- Hunt L.S., White J.W., Hoogenboom G., 2001. *Agronomic data: Advances in documentation and protocols for exchange and use*. Agricultural Systems 70 (2001): 477-492
- Stöckle C.O., Donatelli M., Nelson R., 2003. *CropSyst, a cropping system simulation model*. European journal of Agronomy, 18 (2003): 289-307