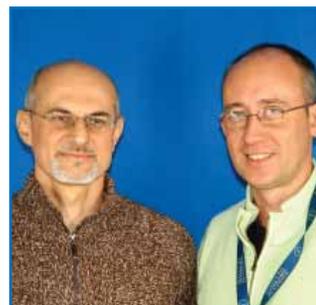


UN DATABASE MANAGEMENT SYSTEM PER L'ANALISI DI DATI IN SANITÀ PUBBLICA: VANTAGGI E CRITICITÀ



Massimiliano Bugarini e Romano Arcieri
Dipartimento del Farmaco, ISS

RIASSUNTO - Nell'articolo è presentato un software, sviluppato dal Reparto di Ricerca Clinica e Farmacologia Sperimentale, Dipartimento del Farmaco dell'Istituto Superiore di Sanità, progettato per il data-management, query, statistiche descrittive e altre funzionalità statistiche. Tale software si propone come ponte tra l'esigenza di informazioni da valutare e le relative uscite di analisi da produrre, distinguendo queste per tipologia. È utile al ricercatore per acquisire i dati sorgente e successivamente per organizzare, creare e mantenere la storia di tutte le analisi effettuate, usando un approccio gerarchico. È utilizzato dal 2005 ed è continuamente aggiornato con nuove procedure e funzionalità.

Parole chiave: studi clinici, statistica, sanità pubblica

SUMMARY (*A data management system software for data analysis in public health: advantages and criticality*) - The aim of this article is to present a software, developed by the Department of Drug Research and Evaluation (Italian National Institute of Health), designed for the data management, queries on data, descriptive statistics analysis and other statistical functions. This software covers the gap between more data sets to be evaluated and the production of several analysis outputs, telling apart them by typology. It helps the researcher to link the source databases and then create, organize and hold story of all analysis steps, using a hierarchic approach. Since 2005 it has been used and enhanced with new procedures and features.

Key words: clinical study, statistics, public health

massimiliano.bugarini@iss.it

In informatica, un Database Management System (DBMS) è un sistema software progettato per consentire la creazione e manipolazione efficiente di database (ovvero di collezioni di dati strutturati) solitamente da parte di più utenti.

I DBMS svolgono un ruolo fondamentale in numerose applicazioni: contabilità, gestione delle risorse umane nei progetti, gestione e ricerca di pubblicazioni dedicate, contesti tecnici come la gestione di applicazioni accessibili a più utenti (1).

In sanità pubblica, gli archivi posti su basi di dati sono stati utilizzati per scopi amministrativi e, successivamente, anche a scopo di studio e di ricerca, al fine di determinare indicatori prognostici di malattia e/o di esito (2).

Anche più propriamente la ricerca clinica, con il tempo, si è sempre maggiormente appoggiata a strumenti informatici, con capacità sempre

maggiori di interrogazione integrata di database multipli, costruiti per la raccolta di informazioni dedicate e specifiche.

Nell'ambito della ricerca clinica, la capacità di fornire uno strumento informatico usufruibile anche da parte di personale medico e paramedico per l'interrogazione e l'accessibilità a grandi archivi di informazioni sanitarie, anche se allocate in database differenti, potrebbe presentare vantaggi non trascurabili per studi di prevenzione e di esito clinico.

Lo sviluppo di nuove tecnologie hardware e l'evoluzione continua dei software hanno consentito la gestione di volumi di dati sempre più grandi e, non meno importante, più compatibili tra loro. A oggi, con opportuni software e un personal computer, è possibile avere una potenzialità di interrogazione e analisi di dati con tempi e costi contenuti.

Il progetto

Il Reparto di Ricerca Clinica e Farmacologia Sperimentale, Dipartimento del Farmaco dell'Istituto Superiore di Sanità, ha sviluppato dal 2004, un software che da un'unica console di lavoro permette di gestire "progetti" di analisi di dati, assegnando a ogni progetto una cartella di lavoro. Per ogni progetto è possibile:

- caricare dati da diversi database, come collezione di tabelle di dati, all'interno del progetto;
- elaborare i dati tramite controlli di qualità, interrogazioni (query SQL e altre procedure);
- salvare i risultati in tabelle momentanee e/o di uscita in vari formati (Excel, testo, Access, ecc.);
- salvare l'intero progetto nella cartella desiderata per averlo disponibile per successive elaborazioni.

Il software gestisce i dati su uno o più database Access, permettendo anche il caricamento da altri formati (Excel, testo). È organizzato in un insieme di moduli interni che facilitano la

creazione di query di interrogazione, creazione di tabelle di varie tipologie, archiviazione e catalogazione delle procedure di elaborazione (anche singolarmente o in gruppi/sequenze di elaborazione).

Descrizione del software

Lanciando il programma si entra nell'area Gestione Progetti (Figura), dove vengono visualizzati tutti i parametri del progetto correntemente in lavorazione (cartella di lavoro, nome del progetto, data di inizio, database (DB) di dati, DB di uscite, ecc.). Da qui è possibile:

- salvare il progetto corrente nella cartella di lavoro assegnata;
- salvare il progetto in un'altra cartella di lavoro e rinominarlo;
- creare da zero un nuovo progetto (svuotando il progetto corrente) per poi registrarlo nell'archivio progetti, dove sono mantenuti tutti i parametri dei progetti lavorati;
- caricare un progetto esistente (dati, query e tabelle di elaborazione); ▶

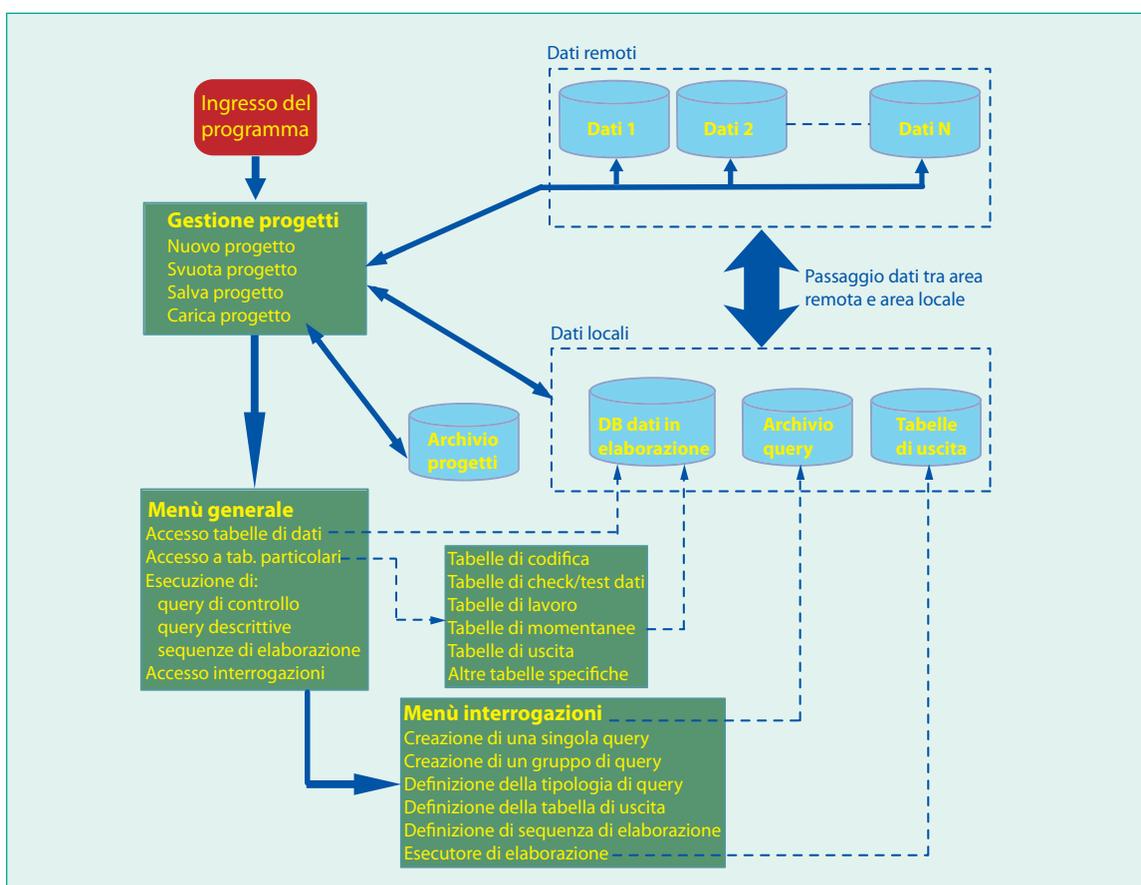
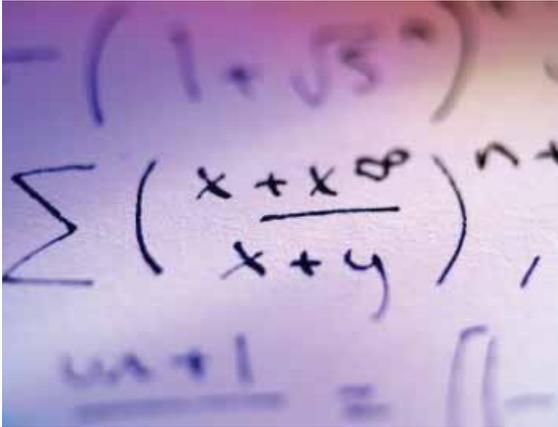


Figura - Schema a blocchi del software



- aggiungere altri dati (tabelle) da altri progetti per effettuare un'analisi trasversale dei dati.

Dall'area "Gestione Progetti" si può passare al Menù generale del progetto, che si configura automaticamente in base alle informazioni contenute, specifiche per ogni progetto. Tale menù si divide in due pagine di lavoro distinte: una per l'accesso alle tabelle e un'altra per l'accesso alle query di interrogazione. Riassumendo in sintesi, qui troviamo:

- **accesso tabelle di dati:** menù contenente l'elenco di tutte le tabelle di dati presenti nel DB. Sono disponibili anche altri menù per l'accesso a sottoinsiemi definiti dall'utente (gruppi di tabelle);
- **accesso a tabelle particolari:** menù per l'accesso a tabelle di particolare tipologia, come ad esempio listati di dati, tabelle di codifica/decodifica delle modalità (valori possibili) previste per le variabili desiderate, tabelle di test/check necessarie durante la fase di analisi di qualità, tabelle di metadati, tabelle momentanee di elaborazione, tabelle di uscita, ecc.

Esecuzione di:

- **query di controllo:** tipologia di query assegnata a tutte quelle query che estraggono informazioni riguardo la qualità dei dati registrati. Il Menù generale dà un accesso diretto alle query di tale tipologia, permettendone l'esecuzione, selezionandole da un'unica lista;
- **query descrittive:** tipologia di query assegnata alle singole query di estrazione/filtraggio dati e creazione delle tabelle di uscita. Il Menù generale dà un accesso diretto alle query di tale tipologia, permettendone l'esecuzione, selezionandole da un'unica lista.

Le query suddette possono far parte anche di una sequenza di elaborazione;

- **sequenze di elaborazione:** sono gruppi di query che vengono eseguite automaticamente in successione, secondo un ordine progressivo definito dall'utente. Il Menù generale dà un accesso diretto alle sequenze di elaborazione disponibili nel progetto, permettendone l'esecuzione, selezionandole da un'unica lista.

Dal Menù generale del progetto, si può accedere al Menù Interrogazioni, che è il motore centrale del programma. Da questa area si gestiscono tutte le procedure di elaborazione del progetto e vengono avviati i processi di analisi dei dati. Nel suddetto menù si trovano le seguenti sezioni:

- **query di prova:** sono query generiche, gestite da un generatore visuale (*user-friendly*) tramite il quale l'utente può comporre l'interrogazione selezionando dati dalle tabelle locali. Una volta creata, e testata, la query di prova può essere salvata all'interno dell'*archivio query*, divenendo così parte integrante dei dati del progetto;
- **archivio query:** è possibile creare, modificare e cancellare query di interrogazione e/o di creazione tabella. Per ciascuna di esse vengono definite informazioni relative a: nome identificativo della query ed eventuale nome del gruppo di appartenenza; tipologia di query e nome della tabella di uscita (tabella prodotta); proprietario/creatore della query, data e ora di creazione; testo della procedura da eseguire (query SQL o funzione/procedura interna);
- **testo della query:** è un'area di testo dove l'utente può scrivere i propri comandi, o più semplicemente inserirvi una query di prova (vedi sopra) dopo averne accertato e testato il corretto funzionamento.



I comandi che in questa area si possono eseguire sono i seguenti:

- comandi SQL di selezione e/o creazione tabella;
- comandi SQL di aggiornamento, inserimento o cancellazione di record;
- cancellazione o clonazione di tabelle;
- cancellazione di tabelle per gruppo o stessa tipologia;
- esecuzione di funzioni di calcolo (in dotazione nel software);
- esecuzione di query di Access e/o di query catalogate (Archivio Query);
- esecuzione di gruppo/sequenze di elaborazione;
- messaggi a video e/o su file (per commentare le fasi di elaborazione).

Conclusioni

Il programma è stato ideato per sopperire alle problematiche che ogni ricercatore clinico incontra nel percorso tra la gestione e organizzazione di dati originali (grezzi) e la produzione di uscite di statistica descrittiva e di analisi qualitativa.

L'elaborazione e la manipolazione dei dati è organizzata liberamente dall'utente, che può "frammentare" l'analisi in vari moduli differenti e indipendenti (specializzati per tipologia di analisi) oppure dare una struttura gerarchica con vari moduli di analisi, migliorandone così la leggibilità e l'affidabilità anche in caso di "variazioni di percorso", che determinano una modifica del piano di analisi originale, per l'avvenuta insorgenza di ulteriori approfondimenti (feedback statistico).



Tra i vantaggi offerti dal programma vi è quello di gestire "tabelle di metadati", costruite nella prima fase di analisi come tabelle di lavoro a cui fare riferimento per le successive fasi, e quello di utilizzare "tabelle momentanee" che vengono cancellate al termine di ogni singola procedura.

L'uso di tali tabelle porta all'ottimizzazione della quantità di tabelle di lavoro prodotte, mentre la definizione di una gerarchia di analisi porta a un'organizzazione più razionale, dando così una maggior chiarezza e affidabilità alle uscite di elaborazione.

Il programma trova applicazione negli studi osservazionali (studi di coorte) e negli studi clinici sperimentali (trial clinici), dove la gestione di grandi quantità di dati e la loro analisi rappresentano punti essenziali per la buona riuscita della ricerca. In questi casi è essenziale poter tenere traccia delle sotto-analisi statistiche, le quali, previa interpretazione da parte del ricercatore, possono condurre a ulteriori livelli di analisi. Per soddisfare queste necessità, il programma aiuta a tenere la storia di tutto il percorso d'analisi, distinguendo le varie fasi per cronologia e per autore di interrogazione. ■

Riferimenti bibliografici

1. Sellitri C, Galati F per il Gruppo di lavoro per il Progetto Piramide. Il sito web Piramide. Dalla citazione al testo completo. *Not Ist Super Sanità* 2004;17(11):8-10.
2. Appelgren E, Luzi P. Convegno. *L'integrazione di archivi elettronici per l'epidemiologia e la sanità pubblica: finalità e metodi*. Riassunti. Roma, 17-18 maggio 2007. Roma: Istituto Superiore di Sanità. (ISTISAN Congressi 07/C2). 61 p.