

L'information technology per la gestione di dati sanitari: il progetto EPIweb

Information technology for the management of health care data: the EPIweb project

Pierpaolo Vittorini, Stefano Necozone, Ferdinando di Orio

Dipartimento di medicina interna e sanità pubblica, Università degli studi dell'Aquila

Corrispondenza: Pierpaolo Vittorini, piazzale Salvatore Tommasi 1, frazione Coppito, 67010 L'Aquila; e-mail: pierpaolo.vittorini@cc.univaq.it

Riassunto

Negli Stati Uniti, il Center for Disease Control and Prevention sta lavorando nella direzione di aumentare la permeabilità delle tecnologie informatiche per una migliore e più efficiente gestione dei dati sanitari. Con riferimento a tali risultati, l'articolo propone una discussione riguardante un sistema informativo basato su web - EPIweb - che permette di condurre attività epidemiologiche utilizzando il web come veicolo di selezione dei centri per il *data entry*, di raccolta ed

elaborazione dati, nonché di pubblicazione e discussione dei risultati. Tale sistema si propone per essere di semplice utilizzo, totalmente programmabile e particolarmente adatto per la gestione di studi multicentrici. L'articolo mostra le caratteristiche del sistema EPIweb, propone un esempio di utilizzo tipico, e termina con una discussione riguardante sia i vantaggi sia i possibili miglioramenti ed estensioni.

(*Epidemiol Prev* 2005; 29(5-6): 284-87)

Parole chiave: web, gestione dati, sistemi informativi, epidemiologia

Abstract

In the US, the Center for Disease Control and Prevention has produced has increased the permeability of the computer science technologies, in order to achieve a better and more efficient management of health care data. In this context, the present paper proposes a discussion regarding a web-based information system, called EPIweb. This system allows researchers to select the centers for the data entry, collect and elaborate health care data,

produce technical reports and discuss results. Such a system aims to be easy-to-use, totally configurable and particularly suitable for the management of multicenter studies. The paper shows the EPIweb features, proposes a sample system run, and concludes with a discussion regarding both the advantages and the possible improvements and extensions.

(*Epidemiol Prev* 2005; 29(5-6): 284-87)

Key words: web, data management, information systems, epidemiology

Introduzione

Per la ricerca epidemiologica e per le attività di valutazione e di *decision-making* in campo sanitario, sono oggi indispensabili strumenti informatici¹, statistici e di ricerca operativa² che siano in grado di gestire efficacemente grandi quantità di dati.³ L'uso di sistemi di gestione e di memorizzazione di informazioni sanitarie è un primo passo per garantire alta qualità ai dati,⁴ ed eventualmente permettere l'adozione di sofisticati meccanismi di investigazione degli stessi.⁵ Epi Info⁶ è un software sviluppato dal Center for Disease Control and Prevention (CDC)⁷ in grado di offrire un supporto importante nella fase di analisi dei dati, nonché nella successiva fase di creazione dei report. Tuttavia, accade con frequenza che i tempi necessari affinché i vari questionari arrivino in sede, vengano catalogati e quindi inseriti sul calcolatore siano molto lunghi, e che alcuni dati, scritti su carta, risultino di difficile comprensione o addirittura palesemente non corretti.

Tali problematiche, per le quali Epi Info non offre specifici controlli, diventano cruciali e di fondamentale importanza in altri scenari in cui risulti centrale una «[...] tempestiva divulgazione dei risultati ad adeguati decisori [...]»,⁸ oppure qualora ci si voglia concentrare sulla qualità dei dati, nel quale caso è indispensabile cercare di prevenire ogni situazione che

possa portare a dati illeggibili, male interpretabili o manifestamente errati.

Sulla base di tali considerazioni, per far fronte alle quali è indispensabile adottare un approccio integrato e totalmente informatizzato, il CDC ha sviluppato la PHIN,⁹ ovvero un'infrastruttura di rete formata da una serie di componenti software che permettono di eseguire tramite internet la fase di interscambio di dati. Nel *framework* di riferimento inerente la PHIN, l'elaborazione dei dati stessi viene però demandata ai nodi periferici.

Il sistema EPIweb: obiettivi

Il sistema EPIweb¹⁰ nasce per collocarsi a cavallo delle due proposte del Center for Disease Control and Prevention, sotto forma di un sistema informativo basato su web progettato per offrire capacità elaborative comparabili con quelle di Epi Info, per operare in internet come nel caso della PHIN, per offrire ai ricercatori la possibilità di gestire tramite interfacce facili da utilizzare per gli utenti tutte le fasi di:

- selezione dei centri coinvolti
- creazione dei questionari
- raccolta ed elaborazione dei dati
- pubblicazione dei risultati sotto forma di *technical report* o di forum di discussione

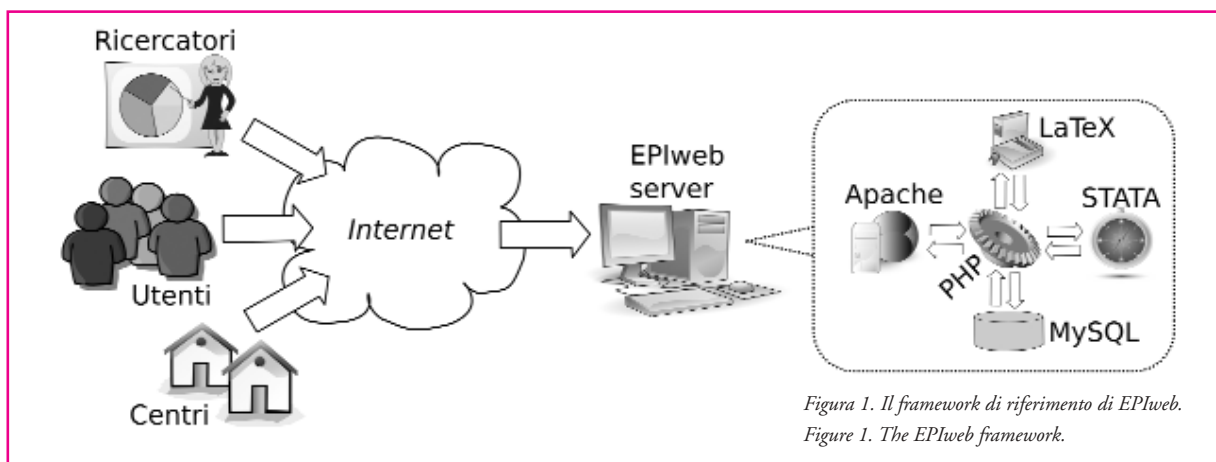


Figura 1. Il framework di riferimento di EPIweb.
Figure 1. The EPIweb framework.

e per assicurare un controllo sui dati coerenti con i dettami dell'ingegneria del software.¹¹

L'approccio totalmente programmabile offerto dal sistema EPIweb, e la sua capacità di gestire esclusivamente tramite internet tutte le fasi precedentemente elencate, offre uno strumento versatile e rilevante in particolare in un contesto di gestione di studi multicentrici. Strumenti di simile concezione sono presenti nella letteratura scientifica, ma generalmente si tratta di sistemi informativi sviluppati per un particolare studio multicentrico.^{12,13} La loro conseguente specificità è un indubbio vantaggio dell'applicativo in termini di adeguatezza alle esigenze dello studio, ma è anche uno svantaggio in termini di costi, in quanto l'attività di sviluppo dell'applicativo può assorbire una parte a volte rilevante del finanziamento, per poi lasciare uno strumento non riutilizzabile in altri studi. In un contesto di sorveglianza epidemiologica, infine, il web è stato spesso adottato vista la sua capacità di trasmettere in tempo reale i dati inerenti l'attività stessa.¹⁴ Per esempio, il sistema ESSENCE¹⁵ usa un sistema di trasferimento *file* crittografato per lo scambio di dati fra i vari nodi, mentre il Rapid Syndrome Validation Project¹⁶ usa una combinazione del web e della tecnologia Java per implementare un sistema che fornisca degli *early warnings* in relazione a minacce per la salute pubblica. In entrambi i casi, comunque, i sistemi menzionati sono ritagliati per monitorare ben specifici ambiti di interesse.

Il sistema EPIweb: caratteristiche tecniche

In relazione al *framework* di riferimento di EPIweb (figura 1), tutte le varie attività che compongono uno studio epidemiologico (i.e. selezione dei centri, creazione dei questionari, inserimento ed elaborazione dei dati, produzione dei report) vengono eseguite a distanza, dalle rispettive sedi, e tutte in formato esclusivamente digitale. Secondo EPIweb, i soggetti che possono partecipare a uno studio sono tre: i ricercatori che organizzano e gestiscono lo studio, i centri che effettuano il data entry, e gli utenti che leggono i report e di-

battono i risultati di uno studio. Il fulcro del sistema è l'EPIweb server, ovvero un server al quale sia i ricercatori, sia i centri, sia gli utenti possono accedere utilizzando un normale browser per internet. Per motivi di sicurezza e di privacy, è necessario che:

- ciascun partecipante all'indagine epidemiologica utilizzi una coppia login/password per accedere al sistema;
- le operazioni, quali per esempio la trasmissione dati, vengano eseguite su un canale crittografato;
- il server sia protetto da intrusioni tramite *firewall*.

Dal *callout* di figura 1, è inoltre possibile notare come l'architettura del sistema EPIweb sia estremamente modulare, in quanto si appoggia su STATA¹⁷ per le elaborazioni statistiche, sul DBMS MySQL¹⁸ per la gestione dei dati, sul sistema LaTeX¹⁹ per l'impaginazione dei report in formato PDF, sul noto web server Apache²⁰ per la gestione della comunicazione, e utilizza degli script PHP²¹ per l'implementazione delle varie funzionalità. Questa architettura comporta un notevole risparmio nei tempi di sviluppo dell'applicativo, una grande potenza in termini di elaborazioni statistiche, una buona efficienza di gestione dei dati e un'ottima qualità nei report prodotti.

Inoltre, la gestione dei dati avviene tenendo in considerazione alcune problematiche di qualità¹¹ inerenti la tempestività, la completezza dei valori, l'usabilità, l'interpretabilità, l'accuratezza e l'efficienza di memorizzazione.

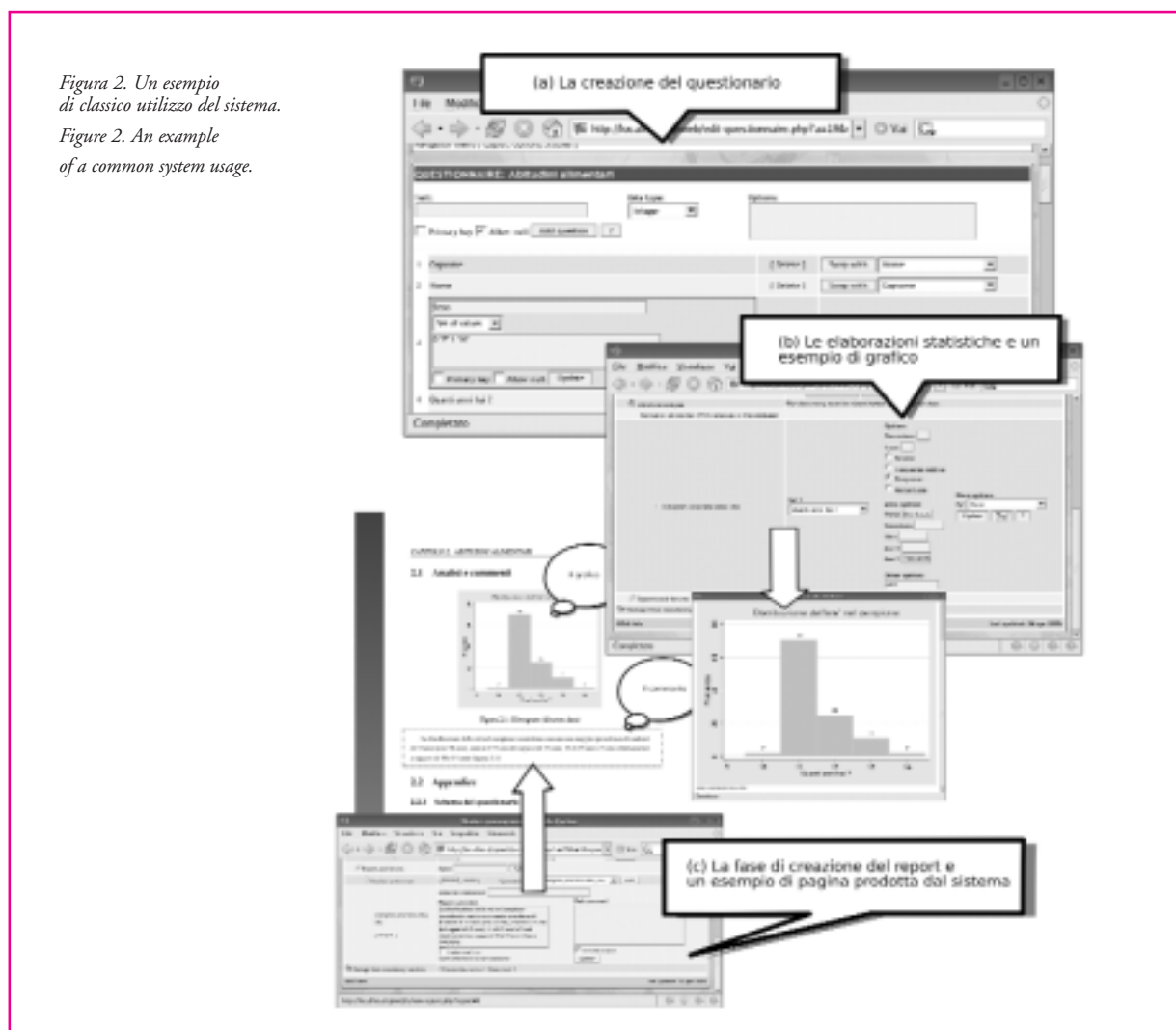
Un'ulteriore problematica ampiamente riconosciuta dalla comunità scientifica in relazione agli applicativi aperti alle rete internet è quella relativa all'interoperabilità, ovvero la capacità di un dato sistema di «dialogare» con sistemi differenti. Per tale motivo, EPIweb è stato progettato in modo che i suoi dati possano essere usati all'interno di altri software tramite un formato standard aperto.

Uno scenario di utilizzo tipico

Per accedere al sistema EPIweb, è necessario disporre di alcune credenziali di accesso rilasciate dall'amministratore di

Figura 2. Un esempio di classico utilizzo del sistema.

Figure 2. An example of a common system usage.



sistema; con tali credenziali, ci si collega al sito in cui è attivo il sistema. Quindi, si aggiungono i centri che saranno chiamati a eseguire il data entry.

In figura 2(a) si mostra la fase successiva di sviluppo del questionario: come è possibile notare, la parte superiore dell'interfaccia rende possibile l'inserimento di nuove domande, la parte inferiore permette di modificarle. In questa fase si devono introdurre i controlli tesi per esempio a limitare la presenza di dati mancanti (usando l'opzione «Allow null»), oppure a prevenire l'inserimento di dati errati (usando il formato «Set of values» e introducendo i valori ammissibili con relativa codifica). In figura è mostrato il caso particolare della variabile «Sesso», introdotta per essere un insieme di valori la cui codifica sia $\{0=F, 1=M\}$, e per la quale non sia possibile accettare un dato mancante.

Al termine dello sviluppo del questionario, si abilita il data entry e si comunicano ai centri le credenziali necessarie per accedere al sistema. Al centro che accederà al sistema, verrà

reso disponibile un *form* per immettere i dati creati automaticamente dal sistema sulla base del questionario sviluppato in precedenza.

Si eseguono quindi le elaborazioni statistiche e/o i grafici ritenuti necessari per l'analisi (in figura 2(b) è mostrato un esempio di grafico). Tra tutte le analisi possibili, si menzionano varie tipologie di grafici e un buon numero di analisi statistiche sia descrittive sia inferenziali.

Al termine delle elaborazioni, il ricercatore può occuparsi della costruzione del report. I report sono composti aggiungendo commenti alle elaborazioni statistiche prodotte precedentemente: il sistema provvederà automaticamente ad assemblare il report sotto forma di documento PDF. In figura 2(c) sono mostrate l'interfaccia usata per costruire il report e una pagina dello stesso così come prodotta dal sistema.

Vantaggi, limiti e possibili estensioni del sistema

Lo strumento permette di condurre attività epidemiologiche

cogliendo alcune importanti problematiche relative alla necessità di dotarsi di sistemi informativi versatili e programmabili, nonché alla possibilità di disporre di dati sanitari preventivamente controllati secondo alcuni parametri scelti dal ricercatore. Inoltre, l'uso di internet rende il processo di raccolta dati più veloce, economico e sicuro, in quanto i questionari sono inviati in tempo reale, eliminando il cartaceo e utilizzando canali protetti da crittografia.

L'adozione poi di un approccio che usi la tecnologia dei sistemi informativi basati su web, al posto di sistemi *stand-alone* (come Epi Info) o di applicazioni totalmente distribuite (come la PHIN), impone alcune considerazioni. Anzitutto, si riducono i costi connessi con il *deployment* di un applicativo, ovvero i costi connessi con l'esecuzione di tutte quelle attività che seguono la prima *release* (per esempio gli aggiornamenti, qualora applicati all'EPIweb server, diventerebbero immediatamente disponibili anche ai *client*). Inoltre, poiché i dati sono memorizzati sul server, quest'ultimo può essere configurato per essere protetto da virus, malfunzionamenti hardware e intrusioni esterne. Ancora, poiché le elaborazioni sono eseguite in maniera centralizzata, anche *client* dotati di bassa potenza di calcolo sono in grado di eseguire analisi complesse.

Negli studi pilota condotti finora sul sistema EPIweb, sono emersi alcuni possibili miglioramenti ed estensioni. In particolare, si ritiene necessario introdurre un ulteriore controllo relativamente alla consistenza interna del questionario, al fine di prevenire la possibilità che vengano date risposte a domande non richieste. Inoltre, si ritiene importante rendere più versatile il sistema di generazione di report che, nell'attuale versione, prevede un'organizzazione fissa dei vari capitoli.

Conflitti di interesse: nessuno

Bibliografia

- Eubank S, Guclu H, Kumar VS et al. Modeling disease outbreaks in realistic urban social networks. *Nature* 2004; 13: 429(6988): 180-84.
- Hauck K, Smith PC, Goddard M. The Economics of Priority Setting for Health Care: A Literature Review. *Health, Nutrition, and Population*. The World Bank, 2003.
- Capocaccia R, Gatta G, Roazzi P et al. The EURO-CARE-3 database: methodology of data collection, standardisation, quality control and statistical analysis. *Ann Oncol* 2003; 14: 14-27.
- Redman TC. *Data Quality for the Information Age*. Artech House, 1996.
- Bellazzi R, Zupan B (eds). Intelligent Data Analysis. *Methods Inf Med* (Special Issue) 2001; 40(5).
- CDC. *Center for Disease Control and Prevention*. <http://www.cdc.gov/>
- CDC. *Epi Info*. <http://www.cdc.gov/epiinfo/>
- Lober WB, Karras BT, Wagner MM et al. Roundtable on bioterrorism detection: information system-based surveillance. *J Am Med Assoc* 2002; 9: 104-15.
- CDC. *Public Health Information Network*. <http://www.cdc.gov/phn/>
- Vittorini P, Necozone S, di Orio F. Il web come strumento per la gestione di dati in sanità: il progetto EPIweb. *Journal of Preventive Medicine and Hygiene* 2004; 45(4): 337.
- Strong DM, Wang RY. Beyond accuracy: what data quality means to data consumer. *Journal of Management Information Systems*, 12(4), 1996.
- Sanson RL, Morris RS, Stern MW. EpiMAN-FMD: a decision support system for managing epidemics of vesicular disease. *Rev Sci Tech* 1999; 18(3): 593-605.
- Unützer J, Choi Y, Cook IA, Oishi S. Clinical Computing: A Web-Based Data Management System to Improve Care for Depression in a Multicenter Clinical Trial. *Psychiatr Serv* 2002; 53: 671-78.
- Bravata DM, McDonald KM, Smith WM et al. Systematic review: surveillance systems for early detection of bioterrorism-related diseases. *Ann Intern Med* 2004; 140(11): 910-22.
- Lombardo JS, Burkom H, Pavlin J. ESSENCE II and the framework for evaluating syndromic surveillance systems. *MMWR Morb Mortal Wkly Rep* 2004; 53 (Suppl): 159-65.
- Zelico A, Brillman J, Forslund DW, George JE, Zink S, Koenig S. *The Rapid Syndrome Validation Project (RSVP)*. Sandia National Laboratories, 2001.
- Stata Corporation. *Stata Base Reference Manual* (4 volumes). Stata Press, 2003.
- MySQL. *MySQL database server*. <http://www.mysql.com/>
- Mittelbach F, Goossens M, Rahtz S, Kopka H, Daly PW. *The LaTeX Companions, Revised Boxed Set: A Complete Guide and Reference for Preparing, Illustrating, and Publishing Technical Documents (2nd Edition)*. Addison-Wesley Professional, 2004.
- Laurie B, Laurie P. *Apache: The Definitive Guide* (3rd Edition). O'Reilly, 2002.
- Schlossnagle G. *Advanced PHP Programming*. Sams, 2004.