

PubMed

La nuova versione di un efficace strumento di ricerca bibliografica in ambito biomedico

di Lucio Lubiana

PubMed <<http://www.ncbi.nlm.nih.gov>> è stato progettato e realizzato, nel corso del 1999, dal National Center for Biotechnology Information (NCBI) alla National Library of Medicine (NLM) ed è localizzato presso il National Institute of Health (NIH). Venne sviluppato in collaborazione con alcune case editrici mediche come "strumento di ricerca e di accesso alle citazioni della letteratura e di legame al testo completo dei periodici dei siti web delle case editrici partecipanti".¹ Nel corso dei primi mesi del 2000 è stata creata una nuova versione di PubMed che si differenzia da quella precedente soprattutto per la modalità di ricerca e di legame fra i diversi moduli di ricerca. La nuova versione ha una interfaccia di ricerca che permette un legame ai campi di limitazione (anno, paese ecc.), agli indici, alla storia della ricerca, con la possibilità di selezione degli articoli trovati e di quelli vicini al soggetto di ricerca.

PubMed è un motore di ricerca che si fonda su archivi bibliografici di biomedicina (Medline, PreMedline), su quelli del NCBI (GenBank), ricercabili tramite il sistema Entrez,

e sulle riviste elettroniche, a testo completo, di alcune case editrici aderenti a questo progetto, che inviano elettronicamente alla NLM le loro citazioni prima o dopo la pubblicazione su carta o su web.

In questo modo si ha la possibilità di passare, dopo aver fatto la ricerca con PubMed, per le riviste che lo permettono, dalla citazione bibliografica al testo completo dell'articolo in formato elettronico.

PubMed permette attualmente l'accesso a più di 10 milioni di citazioni bibliografiche di Medline (dal 1966 ad oggi), PreMedline, HealthStar e le citazioni o il full-text forniti dalle case editrici partecipanti. Inoltre, fornisce l'accesso e un legame integrato con gli archivi della biologia molecolare inclusi nel sistema NCBI Entrez Retrieval System (Entrez Molecular Sequence Database System). Questi database contengono in un sistema integrato diversi archivi automatizzati fra cui le sequenze di DNA delle proteine (*3D protein structure data*), i *population study data sets* e il genoma completo.

La principale caratteristica di PubMed è che tutti i riferimenti bibliografici della ricerca biomedica ar-

chiviati possono essere consultati, tramite Internet, gratuitamente da ogni persona. Ciò per quanto riguarda la consultazione degli archivi di biomedicina e la ricerca della letteratura pubblicata nelle riviste partecipanti al progetto PubMed e quella distribuita, con il sistema del controllo della qualità (*peer review control*), tramite BioMed Central <<http://www.biomednet.com>>, che attualmente consente l'accesso elettronico libero agli indici e agli abstract di 150 riviste di biologia e di medicina.²

PubMed, denominato inizialmente E-biomed, comprendeva gli articoli di medicina di base, medicina clinica, bioingegneria, biologia, biologia molecolare e un archivio elettronico di preprint.

Nel modello di PubMed gli autori, e non le case editrici, mantengono i diritti di copyright, permettendo così la loro libera distribuzione. Le case editrici commerciali e le aziende farmaceutiche, invece, mantengono i diritti d'autore per le ricerche pubblicate e per quelle utilizzate per la produzione industriale di farmaci. Perciò alcune case editrici commerciali di medicina, all'inizio, manifestarono le loro perplessità sulla libera diffusione di informazioni bibliografiche biomediche non controllate, sostenendo che potevano avere degli effetti dannosi sulla salute dei cittadini.

Il National Institute of Health si mosse, al riguardo, con estrema cautela, avvertendo gli autori della pericolosità di sorpassare il meccanismo di controllo della qualità dei loro articoli resi pubblici. Decise, quindi, di assegnare a un gruppo di esperti il lavoro di controllo della qualità degli articoli (*peer review control*) e dei documenti inviati per essere inseriti nell'archivio di PubMed Central.

Anche quelli inviati a BioMed Central vengono sottoposti al controllo di qualità da parte di un gruppo di ricercatori, e quindi, se accolti, de-

positati in PubMed dove sono liberamente accessibili.

Soltanto i preprint di biomedicina inviati a PubMed non vengono controllati da questo gruppo di esperti per accertarne la validità scientifica. La diffusione libera su web dei preprint di medicina è stata anche intrapresa dal gruppo editoriale del "British Medical Journal" e dalla biblioteca dell'Università di Stanford che hanno creato Netprints, un preprint server di medicina e clinica medica <<http://www.clinmed.netprints.org>>, simile a quello dei fisici.³

Gran parte delle case editrici commerciali hanno ignorato PubMed ma, all'inizio del 2000, dodici case editrici (includendo Elsevier e Academic Press, due grosse case editrici di periodici di medicina) hanno annunciato l'intenzione di sviluppare dei legami elettronici fra le loro 3.000 riviste, permettendo ai loro clienti di navigare dalle riviste di una casa editrice a quelle delle altre case editrici aderenti al progetto.

Le case editrici scientifiche di società accademiche, invece, hanno accolto PubMed in modo non uniforme. Alcune riviste hanno subito aderito al progetto, mentre la maggior parte di esse non lo ha ancora fatto. Fra quelle che hanno aderito ci sono il "New England journal of medicine", il "British journal of pharmacology", "Glia" e altre che stanno inviando al server i nuovi articoli elettronici proposti per la pubblicazione. Il "Proceedings of the National Academy of Sciences" (PNAS) ha deciso, a partire dal gennaio 2000, di donare il testo completo dei suoi articoli elettronici a PubMed.

L'American Society of Cell Biology, nel settembre del 1999, ha deciso di donare gratuitamente, a partire dal 2000,

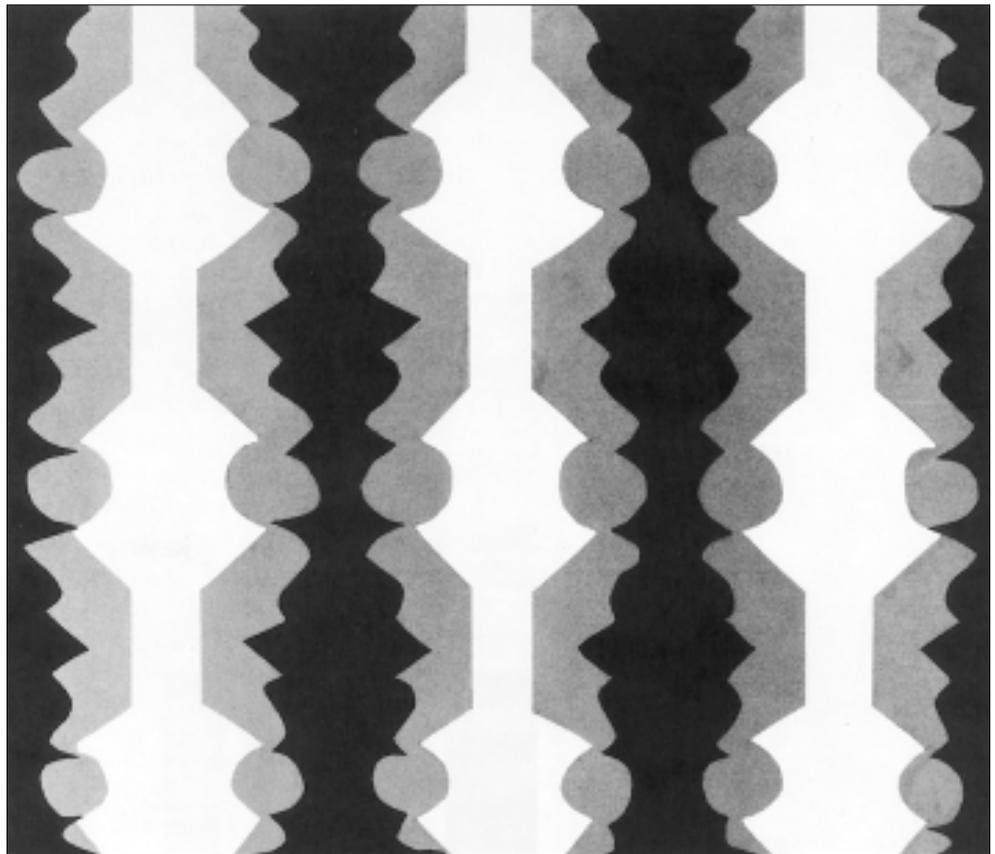
a PubMed le sue pubblicazioni, fra cui *Molecular biology of the cell*, a condizione che tutte le altre citazioni inserite nel database fossero state controllate (*peer-reviewed*), al contrario di quanto inizialmente aveva pensato il NIH.⁴

Anche la rivista "Annals of internal medicine" ha deciso di inviare i suoi articoli elettronici a PubMed.⁵ L'Università di Stanford ha sviluppato, invece, il progetto High wire press (<<http://highwire.stanford.edu>>) e ha creato un archivio elettronico della letteratura scientifica liberamente accessibile che attualmente permette l'accesso elettronico a 137.000 articoli scientifici di cui 22.000 provenienti dal "British medical journal" (dal 1994 al 2000), 26.000 articoli pubblicati nelle dieci riviste dell'American Society for Microbiology (anni 1995-1999), 5.300 articoli del "Journal of biological chemistry" (per il

1999), l'accesso al PNAS (1990-1995) e alle riviste dell'American Physiological Society.⁶

E-biosci: il progetto europeo

Il progetto PubMed è stato attentamente seguito dalle istituzioni scientifiche europee che non sono riuscite a creare, in anticipo su PubMed, uno strumento unitario di recupero della letteratura scientifica su web. Perciò l'Organizzazione europea di biologia molecolare (in seguito EMBO-European Molecular Biology Organization) ha tentato, nell'estate del 1999, di accordarsi con il National Institute of Health (USA) per creare un comune e libero sito Internet della letteratura scientifica. L'accordo non è stato trovato e l'EMBO ha deciso di creare un proprio archivio europeo di letteratura scientifica deno- ➤



minato E-Biosci, ancora in fase di progettazione. Questo progetto è portato avanti dalla European Molecular Biology Laboratory di Cambridge (UK). Verrà inoltre creato, all'interno di E-Biosci, anche un archivio elettronico delle tesi di PhD europee.

Lo scopo di E-Biosci è quello di evitare il monopolio degli Stati Uniti nel campo dell'archiviazione e della libera diffusione della letteratura scientifica.

La principale differenza fra PubMed e E-Biosci è costituita dal contenuto dei due database. In entrambi i sistemi gran parte dei documenti accettati è controllata (*peer-reviewed*) e proviene dalle esistenti riviste di medicina.

I documenti non controllati (*preprint, non-peer-reviewed*) vengono trattati diversamente dai due sistemi. PubMed li accetta mentre E-Biosci accoglie solamente quelli che sono stati sottoposti al controllo della loro qualità.

Perciò l'EMBO sta pianificando la costruzione di un sistema centrale di controllo della qualità della documentazione scientifica amministrato da riconosciute società scientifiche e da quelle aderenti a EMBO. Lo stesso principio è stato accolto anche dall'Asia-Pacific International Molecular Biology Network.

PubMed, invece, accogliendo anche la letteratura non controllata, si fonda su un sistema di controllo degli articoli decentralizzato e con una centralizzata infrastruttura informatica capace di archiviare e rendere accessibile, in modo sicuro e stabile, la documentazione su web.

Un altro problema aperto è l'accesso alle riviste elettroniche. PubMed lo consente per quelle case editrici che permettono l'accesso libero alle loro riviste mentre a quelle che non lo permettono PubMed dà la possibilità di collegarsi elettronicamente ai loro siti web. E-Biosci, da quanto se ne sa, invece, prevederà un accesso libero al suo archivio bi-

bliografico e uno a pagamento alle riviste elettroniche legate a questo archivio.⁷

Descrizione di PubMed

PubMed è la base dati bibliografica che fa riferimento agli archivi della biologia molecolare (Entrez Molecular Sequence Database System – Entrez) e a Medline e permette un accesso alternativo, rispetto a quello dei tradizionali periodici, alla ricerca bibliografica.

In dettaglio il sistema Entrez (sistema di ricerca dei database di NCBI) comprende le seguenti informazioni:

- sequenze di DNA depositate presso la GenBank dell'European Molecular Biology Laboratory (EMBL) e presso la DNA Data Bank of Japan (DDBJ);
 - sequenze di proteine registrate presso il Swiss-Prot (Protein Identification Resources-PIR, Protein Data Bank-PDB);
 - dati sul genoma e sulla mappatura di cromosomi;
 - strutture proteiche a 3D derivate dalla PDB e incorporate nella MMDB (Molecular Modeling Data Base);
 - PubMed, base dati bibliografica formata da Medline e PreMedline.
- Tra PubMed e le precedenti è stato creato un legame elettronico in grado di associare la ricerca di citazioni bibliografiche con quella sulle sequenze molecolari e viceversa. Una interfaccia grafica permette di visualizzare i genomi completi e cromosomi mentre programmi grafici consentono di visualizzare, in tempo reale, le strutture proteiche tridimensionali con relative rotazioni.

Caratteristiche di PubMed

PubMed consente l'accesso a citazioni bibliografiche provenienti da:

- a) PreMedline (PREliminary Med-

line) – è l'archivio bibliografico dove vengono quotidianamente inserite in via preliminare le citazioni presenti nelle riviste analizzate dal sistema prima che la NLM le sottoponga al processo di indicizzazione e *controllo di qualità*. Ad ogni record presente nel PreMedline viene dato un numero di identificazione (PMID, PubMed unique Identifier) che si aggiunge a quello assegnato dal Medline (UI, Unique Identifier);

b) Medline (dal 1966 a oggi);

c) periodici trasmessi elettronicamente dagli editori. Alcune citazioni reperibili nel PubMed possono non presentare il numero di identificazione Medline (Medline UI). Ciò si verifica quando le citazioni vengono trasmesse elettronicamente dall'editore direttamente a PubMed, che solo in un secondo momento le invia alla NLM per essere inserite in PreMedline.

Alcune citazioni bibliografiche inviate elettronicamente a PubMed e provenienti da periodici indicizzati selettivamente non vengono inserite in Medline in quanto esulano dagli argomenti trattati dall'archivio Medline.

PubMed, inoltre, offre i seguenti servizi di ricerca:

- “Journal browser” (permette di consultare le riviste partecipanti a PubMed con le loro abbreviazioni del Medline, il codice ISSN e un loro legame alla pagina web della loro casa editrice);
- “MeSH Browser” (per consultare i descrittori di soggetto del Medline);
- “Matcher” (per trovare le citazioni bibliografiche incomplete di articoli di rivista tramite i seguenti campi: nome della rivista o sua abbreviazione in base al Medline, volume, fascicolo, prima pagina, cognome dell'autore);
- “Single citation” (utilizzata per cercare una citazione di un articolo di rivista con i seguenti campi: titolo del periodico, data, volume, fa-

scicolo, prima pagina, cognome e iniziali del nome dell'autore);

– “Clinical queries” (questo strumento di ricerca è indirizzato ai clinici che intendono ricercare, in base alle diverse categorie terapia-specificità-diagnosi-sensibilità, i documenti riguardanti le diverse malattie);

– “Old PubMed” (collegamento con la maschera di ricerca della precedente versione di PubMed).

In collaborazione con le case editrici, il NCBI sta adattando i libri per il loro inserimento su web, legandoli con PubMed. Il primo libro disponibile, a questo riguardo, è *Molecular biology of the cell* di Alberts (et al.).

Al sito web principale di PubMed vengono, inoltre, offerte le seguenti ulteriori risorse:

“Orders documents”, “Grateful Med”, “Consumer Health” e “Critical alerts”.

I documenti possono essere ordinati con “Loansome Doc” e solo dopo averli selezionati e trasferiti in “Clipboard” possono essere ordinati cliccando il bottone “order” (solo per i clienti registrati con “Loansome Doc”).

Tutti questi servizi sono integrati fra loro e possono essere selezionati, in ogni momento, da qualsiasi punto di PubMed.

Inoltre, nella pagina principale è stato creato un legame fra la nuova versione di PubMed e quella vecchia (“Old PubMed”).

Ricerca in PubMed

a) Ricerca di base in PubMed

L'archivio a cui ci si collega automaticamente è Medline.

Il ricercatore ha tuttavia la possibilità di cambiare file selezionando una delle seguenti basi dati previste dal sistema:

- Medline
- DNA Sequences
- Protein Sequences
- Nucleotide links
- Structure links

Genomes

PopSet (set di sequenze di DNA raccolti per analizzare l'evoluzione della popolazione).

La ricerca si fa digitando, nello spazio apposito, uno o più termini che devono esprimere un concetto, oppure riferendosi al cognome e nome dell'autore di una pubblicazione o al titolo di una rivista.

Questi diversi campi possono essere selezionati prima della digitazione tramite la funzione “limited by” (tutti i campi, tipo di pubblicazione, età, lingue, entez date, umano o animale, genere).

Le altre opzioni, tra cui effettuare la scelta, si riferiscono al numero di documenti visualizzati per pagina e al modo di visualizzarli (quelli stabiliti per default sono venti).

Per quanto riguarda la *ricerca per argomento*, i termini da ricercare possono provenire dal *linguaggio naturale* o dal thesaurus della National Library of Medicine (NLM). Se si digita un termine non seguito

da un apposito qualificatore, il PubMed lo controlla nei seguenti indici:

– MeSH translation table (comprende descrittori MeSH, subheading, sinonimi, nome di sostanze chimiche);

– Journals translation table (titoli di riviste indicizzate);

– Phrase liste (considera un migliaio di frasi derivate dal MeSH, dai titoli, dai riassunti delle citazioni e dai nomi delle sostanze chimiche);

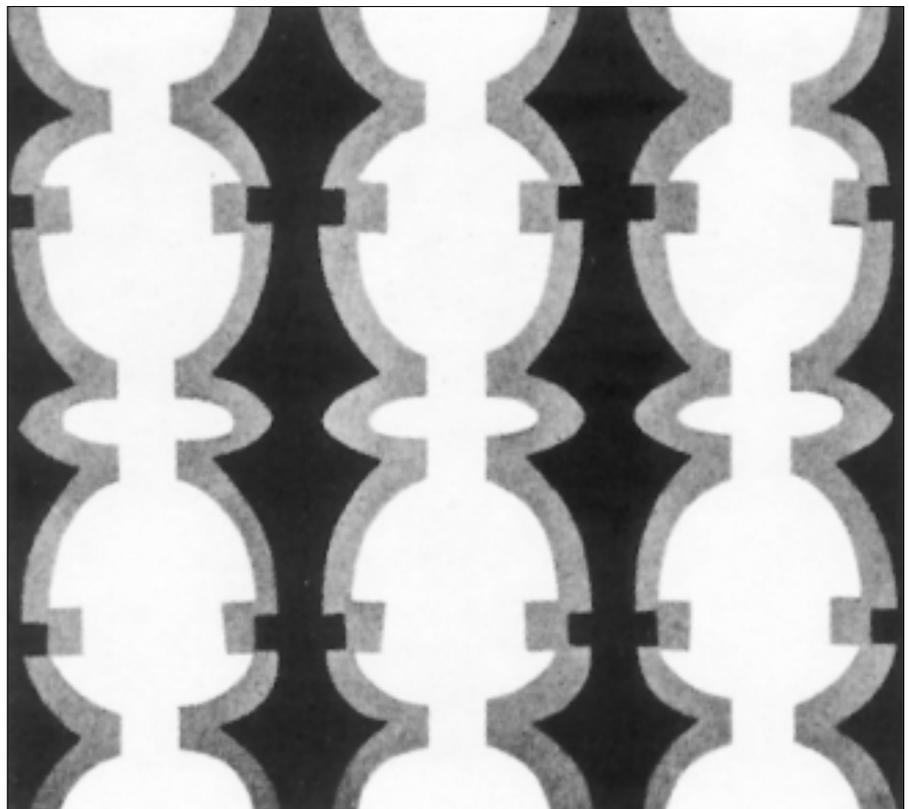
– Author index (contiene tutti i nomi degli autori di pubblicazioni di PubMed).

Se il termine non è trovato in uno di questi indici, PubMed lo cercherà in tutti i campi del record.

È inoltre possibile navigare da un argomento selezionato agli altri vicini.

b) Ricerca avanzata in PubMed

Permette all'utente di ricercare l'archivio nelle seguenti modalità: ➤



- specificando i campi di ricerca;
- visualizzando i termini alfabeticamente adiacenti a quello inserito;
- costruendo una propria strategia di ricerca aggiungendo o togliendo i termini e modificando le relazioni tra di essi con gli operatori logici booleani (AND, NOR, OR).

Il primo strumento per la ricerca avanzata è il “Journal browser”. Si usa questo modulo di ricerca se si desidera trovare un particolare periodico per includerlo, poi, nella ricerca o per trovare il suo sito web o la sua abbreviazione o il codice ISSN. Il secondo strumento è il “Citation matcher” che serve per trovare la citazione completa di un articolo di cui si sa solo l'autore, la rivista o altre citazioni. Il ricercatore può trovare una singola citazione completa usando anche la funzione “Single citation matcher”. Nel fondo della schermata principale del sito web c'è, inoltre, il “Consumer resources” dentro a “Related resources”. L'informazione qui contenuta è destinata ai professionisti e al pubblico generico: fornisce l'accesso a dizionari, database, pubblicazioni, novità editoriali in campo medico, biblioteche e documentazione sulle malattie.

Per limitare il PubMed a certi campi del record bibliografico si seleziona il tasto “Limits” nella pagina principale.

Le ricerche possono essere limitate ai seguenti campi: tipo di pubblicazione, lingua, età, uomo o animale, genere, data di pubblicazione, data di “entrez”.

I limiti di ricerca non possono essere inclusi nei record di Pre-Medline e in quelli inviati dalle case editrici.

Il “Preview/Index”, invece, serve per cercare, selezionando da un'apPOSITA lista, il campo su cui fare la ricerca, esaminare, quindi, il numero di citazioni trovate in un campo prescelto e combinarle, utilizzando gli operatori booleani (AND, NOT, OR), con altri campi.

Si usa per raffinare la ricerca con termini provenienti da specifici campi di ricerca e per vedere i risultati prima di mostrare le citazioni trovate.

La storia della ricerca eseguita può essere visualizzata cliccando su “History” nella pagina principale di PubMed. Queste informazioni vengono perdute dopo un'ora di inattività nella ricerca di PubMed.

“Clipboard” ha la stessa funzione di “History” ma, a differenza di quest'ultimo, permette di ricercare, ordinare, salvare o di collegarsi al sito web del periodico trovato. Questo strumento mantiene in memoria fino a 500 risultati di ricerca che vengono perduti dopo un'ora di inattività. Le citazioni trovate vengono visualizzate in ordine cronologico e quelle più recenti precedono quelle più vecchie.

Conclusioni

La filosofia su cui si fonda PubMed è quella della libera e volontaria partecipazione delle case editrici e delle istituzioni pubbliche alla diffusione della letteratura biomedica in formato elettronico, nella versione a testo completo e in quella di citazione bibliografica, fra i ricercatori, le biblioteche e il pubblico. Questo obiettivo viene perseguito dalla National Library of Medicine per superare i limiti fisici, geografici e di reddito (fra paesi ricchi e quelli poveri) della ricerca biomedica. In questo modo PubMed sta diventando lo strumento principale di ricerca bibliografica in biomedicina che permette il passare, dove ciò è possibile, dalla citazione al testo completo degli articoli.

Non si sa quale futuro avrà invece l'archivio scientifico elettronico europeo E-Biosci ma sicuramente dovrà perseguire obiettivi diversi da PubMed per poter giustificare la sua attività e i suoi pubblici finanziamenti.

Molti scienziati pensano che esso dovrà collaborare e non competere con PubMed. Altri, invece, lo considerano uno strumento di accesso alla letteratura biomedica alternativo a PubMed.

Diversi sono ancora i nodi da sciogliere nello sviluppo di PubMed e in quello di E-biosci, che non si sa ancora bene quali ambiti disciplinari coprirà e quale rapporto avrà con le case editrici.

Vi è poi il nodo rappresentato dall'individuazione del confine che separa la ricerca libera, accessibile a tutti, da quella protetta, riservata solo a particolari categorie di utenti e che non è praticabile tramite PubMed, ma solo attraverso i canali controllati dalle case farmaceutiche e dalle agenzie che gestiscono le “patenti”.

Il nodo del copyright rappresenta un altro elemento critico di frizione fra PubMed e le case editrici commerciali e fra queste e gli autori, che in PubMed mantengono i diritti, mentre nelle riviste delle case editrici commerciali essi vengono trasferiti agli editori o alle aziende (farmaceutiche) che hanno finanziato la ricerca.

Un altro problema ancora aperto è costituito dal complesso rapporto fra PubMed e la case editrici commerciali di letteratura biomedica, la cui filosofia economica è in netta contrapposizione con quella di PubMed. Infatti, alcune (Elsevier e Academic Press) stanno già progettando nuovi sistemi elettronici e nuove modalità economiche per regolare l'accesso alle proprie riviste elettroniche, cercando così di contenere il danno economico e di prestigio che la diffusione di PubMed potrebbe arrecare loro.

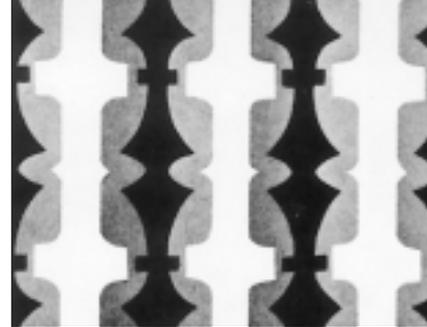
Non è chiaro, infine, il limite che separa, nel progetto PubMed, la ricerca medica di base da quella applicata (finanziata dalle case farmaceutiche private) e se quest'ultima verrà liberamente diffusa da PubMed. ■

Note

¹ Citazione da *Introduction to PubMed*, in *PubMed Overview* <<http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html>> (sito visitato il 03.03.2000). Cfr. J.W. WILSON, *PubMed: a winner on the web for the National Library of Medicine*, "Online", 21 (1997), 6, p. 60-61; G.R. NOTESS, *Free MEDLINEs on the web*, "Database", 21 (1998), 3, p. 71-74; D. HUTCHINSON, *Medline for health professionals: how to search PubMed on the Internet*, New Wind Publishing, 1998; F.E. BLOMM, *Juste a minute, please*, "Science", 285, 1999, p. 197; A.S. RELMAN, *The NIH E-Biomed proposal: potential threat to the evaluation and orderly dissemination of new clinical studies: an editorial*, "New England journal of medicine", 340, 1999, p.1828-29; M. WADMAN, *Critics query financing of E-Biomed*, "Nature", 400, 1999, p. 200.

² Per una trattazione esauriente della prima versione di PubMed si veda: A. DRACOS – M. DELLA SETA – R. CAMMARANO,

PubMed: guida pratica alla consultazione del Medline su Internet, Roma, Di Renzo, 1999; A. DE ROBBIO, *MEDLINE free su web: i servizi PubMed e Internet Grateful Med della National Library of Medicine*, "Bollettino AIB", 37 (1997), 4, p. 481-494; E. SANTORO, *Internet in medicina. Guida all'uso e applicazioni pratiche*, Roma, Il pensiero scientifico, 1998; L. BIANCIARDI, *Alcune considerazioni sulle versioni gratuite della base dati Medline*, "Bollettino AIB", 37 (1997), 1, p. 47-51. ³ D. BUTLER, *Publishing group offers peer review on PubMed central*, "Nature", 402, 11 November 1999, p. 110; ID., *Biomed Central boosted by editorial board*, "Nature", 405, 25 May 2000, 6785, p. 383; *Netprints: the next phase in the evolution of biomedical publishing*, "British Medical Journal", 319, 11 December 1999, p. 1515-1516 <<http://www.bmj.com>>; B.O. MARKOVITZ, *Biomedicine's electronic publishing paradigm shift: copyright policy and PubMed Central*, "Journal of the American Medical Informatics Association", 7 (2000), 3, p. 222-229.



tion", 7 (2000), 3, p. 222-229.

⁴ E. MARSHALL, *PNAS to join PubMed Central: on condition*, "Science", 286, 22 October 1999, p. 655-656.

⁵ P. MAGUIRE, *Will NIH's new web site change medical publishing?*, "ACP-ASIM Observer", January 2000 <<http://www.acponline.org/journals/news/jan2000/newnih.htm>>.

⁶ Per una informazione più esauriente sul progetto HighWire si veda il sito <<http://www.stanford.edu/dept/news/pr/00/00030highwire.html>> (visitato il 05.03.2000) e D. BUTLER, *Biology back issues free as publishers walk HighWire*, "Nature", 404 (2000), 6774, p. 117.

⁷ D. BUTLER, *Europe strengthens its band in bioscience website talks...*, "Nature", 401, 20 September 1999, p.413; R. KOENIG, *Publishers discuss European E-Print site*, "Science", 287 (2000), 5453, p.563-564.