

MEDLINE/PubMed nel XXI secolo

Nicola De Bellis
Carmela Palazzi

Biblioteca medica centralizzata
Università degli studi di Modena e Reggio Emilia
debellis.nicola@unimore.it
palazzi.carmela@unimore.it

Dal database bibliografico all'archivio integrato

1. Premessa e scopo del lavoro

Tra la seconda metà degli anni Novanta e la fine del 2002 la prestigiosa banca dati biomedica MEDLINE, prodotta dalla National Library of Medicine (NLM) di Bethesda, divenuta pubblica e gratuitamente accessibile sul World Wide Web nel 1997,¹ ha conosciuto un'evoluzione interna che ne ha modificato tanto la struttura quanto le potenzialità d'impiego nella direzione di una progressiva integrazione tra unità bibliografiche, letteratura full-text accessibile online e risorse aggiuntive per ricercatori e bibliotecari. Di pari passo con la moltiplicazione di siti e modalità di accesso a MEDLINE (gratuite o a pagamento), PubMed, l'interfaccia web ufficiale, ha potenziato il numero e la qualità dei propri servizi, consegnando ai medici un valido strumento di supporto alle valutazioni diagnostiche e alle decisioni terapeutiche nel quadro della Evidence Based Medicine e prospettando alle biblioteche biomediche l'opportunità di non restare utenti passivi di un semplice archivio di titoli, ma di riformulare le tipologie di accesso alle proprie collezioni digitali e le politiche di *document delivery* grazie ad alcune valide opzioni di personalizzazione.

Il presente articolo ripercorre le principali fasi della riorganizzazione di MEDLINE e delle banche dati del sistema MEDLARS in ambiente web e illustra le più importanti e innovative funzionalità di PubMed, soffermandosi sui due aspetti complementari della ricerca bibliografica e del recupero dei documenti. Nella prima parte, dopo una rassegna delle trasformazioni subite dal sistema MEDLARS, sarà esaminato criticamente il principale strumento di calibrazione delle ricerche bibliografiche in MEDLINE: il Medical Subject Headings (paragrafi 2.1, 2.2). Nella seconda si descriverà una delle più interessanti prerogative di MEDLINE/PubMed nel XXI secolo: la capacità di "accogliere", organizzare e preservare dati di natura, formato e provenienza differente in funzione della loro generale accessibilità. Da questa caratteristica deriva l'integrazione dei dati bibliografici con gli articoli full-text depositati in PubMed Central (paragrafo 3.1), ma anche la possibilità, per biblioteche e centri di documentazione di tutto il

mondo, di riconfigurare i link ipertestuali dei record bibliografici (LinkOut) in rapporto ai bisogni della propria utenza (paragrafo 3.2).

2. Un po' di storia: MEDLARS, MEDLINE, PubMed, il National Library of Medicine Gateway

MEDLINE (Medical Literature, Analysis, and Retrieval System Online) è l'archivio bibliografico online di letteratura biomedica della National Library of Medicine. Derivato dai repertori cartacei *Index medicus*, *Index to dental literature* e *International nursing index*, è disponibile gratuitamente su Internet attraverso il motore di ricerca PubMed del National Center for Biotechnology Information (NCBI) della NLM.² MEDLINE, che si avvale di aggiornamenti quotidiani, dal 1966 indicizza articoli pubblicati in oltre 4.600 periodici biomedici editi negli Stati Uniti e in altri 70 paesi. A ogni citazione, oltre alle tradizionali unità logico-informative (autore, titolo dell'articolo ecc.), sono associati un certo numero di descrittori di soggetto del Medical Subject Headings, il thesaurus ufficiale della National Library of Medicine, e un insieme di campi utili per una selezione mirata del materiale: la tipologia di pubblicazione (articolo di rivista, rassegna, sperimentazione clinica, meta-analisi, test randomizzato ecc.); le sostanze chimiche menzionate e la loro azione farmacologica; i *check tags*, cioè i dati relativi alla metodologia sperimentale (esperimenti su uomini, animali, in vitro, in gravidanza), al tipo di studio condotto (*case-report* o studio comparativo), al sesso e all'età dei soggetti studiati, alla natura e provenienza dei finanziamenti.

MEDLINE è stata concepita all'interno di MEDLARS (Medical Literature Analysis and Retrieval System), un sistema computerizzato d'indicizzazione, archiviazione e ricerca della letteratura scientifica che sin dal 1960 ha permesso a un pubblico prevalentemente composto da ricercatori e documentalisti di accedere a un vasto insieme di banche dati di argomento medico e biologico. MEDLARS è anche il termine con il quale la NLM era solita designare l'insieme dei suoi

¹ *Free web-based access to NLM databases*, "NLM Technical Bulletin", 296 (1997), May-June; URL: <http://www.nlm.nih.gov/pubs/techbull/mj97/mj97_web.html>. Ultima visita 27/12/2002.

² URL: <<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>>. Ultima visita 27/12/2002.

database bibliografici in linea, quindi, oltre a MEDLINE: AIDSLINE, HISTLINE, HSTAR, BIOETHICSLINE, SPACELINE, SDILINE, TOXLINE, CANCERLINE, POPLINE, OLDMEDLINE. Il sistema MEDLARS è operativo dal 1961. Inizialmente si limitava a elaborare complesse strategie di ricerca in modalità batch, senza interazione diretta con il computer, e solo dieci anni dopo, nel 1970-71, grazie a un software avanzato di retrieval (ELHILL), è divenuto stabilmente raggiungibile online, con la possibilità di accedere in tempo reale e in modo interattivo al database bibliografico degli articoli di riviste (MEDLINE) da terminali remoti. Nel marzo 1986, sulla scorta del successo dei microcomputer e del pressoché contemporaneo perfezionamento della tecnologia dei cd-rom, la NLM ha distribuito Grateful Med, un software e un'interfaccia di ricerca tagliati su misura per coloro che volevano interrogare gli archivi MEDLARS dal proprio personal computer. Nei primi anni Novanta, Grateful Med è stato affiancato da un progetto del NCBI inteso a far convergere in un'unica interfaccia di ricerca un subset di MEDLINE e le banche dati fattuali di genetica e biologia molecolare, compresa la mappa del genoma umano. È nato così il primo nucleo di PubMed, il quale sin dalle origini ha sviluppato e mantenuto una vocazione extra bibliografica che ne avrebbe condizionato l'evoluzione successiva, accentuandosi dopo la disponibilità gratuita via Internet dei due servizi, PubMed e Internet Grateful Med, nel 1997. Nello specifico, l'aggiunta al database di collegamenti ipertestuali ai siti degli editori in grado di fornire il testo completo degli articoli citati si è rivelata una mossa strategica, non priva di risvolti economici oltre che etico-scientifici:

si regala il prodotto che prima si vendeva e lo si dota di una serie di servizi a valore aggiunto, anticipando bisogni dell'utenza e vendendo servizi personalizzati legati al prodotto gratuito.³

Il resto è storia recente: i database bibliografici di Internet Grateful Med sono stati progressivamente smantellati e i record ridistribuiti, secondo una precisa strategia editoriale, nei server web della National Library of Medicine. Tra il 2000 e il 2002 è giunto a compimento il progetto di convertire e riversare una parte di questi file speciali, segnatamente le citazioni di articoli di periodici di AIDSLINE, HISTLINE, SPACELINE e BIOETHICSLINE, in MEDLINE, di

cui sono divenuti semplici sottoclassi (*subsets*) isolabili in fase di ricerca mediante appositi filtri.⁴ Con la crescita e la diffusione su larga scala del web e la conseguente ristrutturazione della rete informativa della NLM, i dati MEDLARS rimasti fuori da MEDLINE sono confluiti in forme diverse nella complessa architettura delle pagine web della NLM e delle altre sezioni del National Institutes of Health.⁵ Archivi e servizi sono andati incontro a un processo di crescente integrazione, rendendo sempre più sfumati i contorni che fino a qualche tempo fa separavano, sia da un punto di vista teorico che tecnico-applicativo, un database bibliografico da un catalogo di biblioteca o da un servizio organizzato di accesso ai documenti. La contemporanea crescita della domanda d'informazione biomedica da parte di un pubblico non specializzato, documentata dalle statistiche d'uso di MEDLINE,⁶ ha condotto la NLM a potenziare la propria rete informativa in due direzioni: da un lato, un insieme di risorse per il paziente e l'utente generico (MEDLINEplus, ClinicalTrials.gov, Tox Town); dall'altro, un'interfaccia semplificata per l'accesso contemporaneo a molteplici sistemi di recupero dell'informazione di tipo bibliografico, catalogografico e fattuale: il National Library of Medicine Gateway.⁷ Nato nell'ottobre 2000, Gateway è un metamotope pensato per l'utente

who comes to the National Library of Medicine not knowing exactly what is here or how best to search for it.⁸

La sua caratteristica peculiare è quella d'interrogare simultaneamente più database, compreso MEDLINE, raggruppando i risultati per categorie di facile comprensione.

La diffusione di Gateway ha contribuito ulteriormente a ridisegnare i lineamenti del sistema informativo della biblioteca statunitense, compensando la chiusura pressoché contemporanea, nel 2001, di Internet Grateful Med.⁹ Gateway, la cui nuova versione è stata resa pubblica nel settembre 2002, permette infatti di consultare, da un'unica interfaccia:

- MEDLINE per le citazioni di articoli dal 1966 ad oggi;
- OLDMEDLINE per le citazioni di articoli comparse sul *Cumulated index medicus* (dal 1960 al 1965) e sul *Current list of medical literature* (dal 1957 al 1959);

³ ANTONELLA DE ROBBIO, *Medline free su web: i servizi PubMed e Internet Grateful Med della National Library of Medicine*, "Bollettino AIB", 4 (1997), p. 481.

⁴ *The reorganization of National Library of Medicine (NLM) bibliographic databases*; URL: <http://www.nlm.nih.gov/bsd/conversion/conversion_status.html>.

⁵ Sulla storia della NLM vedi: EVE-MARIE LACROIX – ROBERT MAHNERT, *The US National Library of Medicine in the 21st century: expanding collections, nontraditional formats, new audiences*, "Health Information and Libraries Journal", 19 (2002), p. 126-132; NORMAN H. HOROWITZ, *The National Library of Medicine*, "Neurosurgery", 51 (2002), p. 1304-1314.

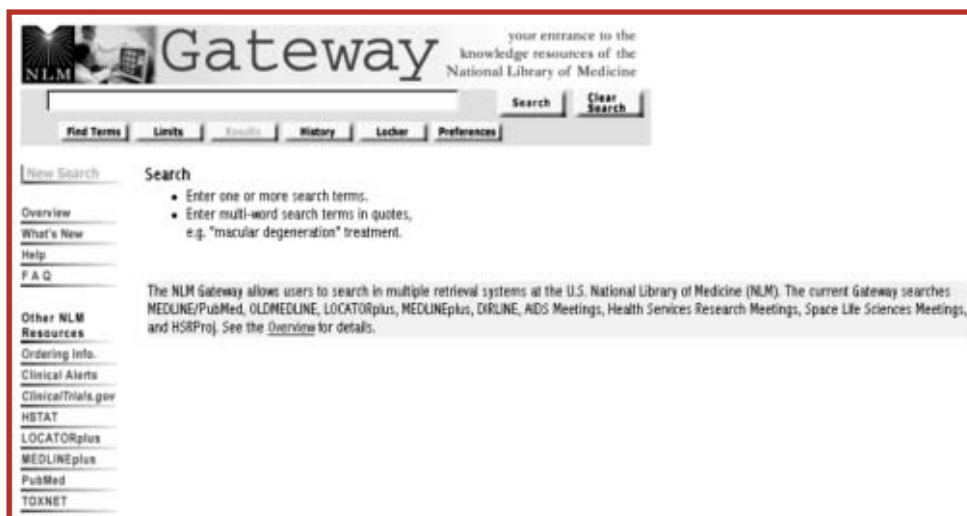
⁶ *Results of National Library of Medicine web site user survey*; URL: <<http://www.nlm.nih.gov/pubs/nlmnews/Online>>. Ultima visita 27/12/2002.

⁷ URL: <<http://gateway.nlm.nih.gov>>. Ultima visita 27/12/2002.

⁸ URL: <<http://www.nlm.nih.gov/pubs/factsheets/gateway.html>>.

⁹ *Internet Grateful Med to be retired; reminder of NLM Gateway availability*, "NLM Technical Bulletin", 318 (2001), January-February; URL: <http://www.nlm.nih.gov/pubs/techbull/jf01/jf01_igm_phaseout.html>. Ultima visita 27/12/2002.

Fig. 1



– gli abstract di convegni nei settori HIV/AIDS, Health Services Research e Space Life Science;

– HSRProj, il database delle citazioni relative a progetti di ricerca *in progress* nel settore delle tecnologie sanitarie e dello sviluppo di linee guida per la pratica clinica;

– MEDLINEplus, il portale di informazioni e servizi per il paziente, comprensivo di una sezione generale di *health topics*, una sezione sui farmaci e un'enciclopedia medica multimediale;

– ClinicalTrials.gov, il database di sperimentazioni controllate (*clinical trials*) sponsorizzate principalmente dal NIH ad uso del paziente;

– DIRLINE, la directory di organizzazioni operanti nel settore della sanità;

– LOCATORplus, il catalogo online di libri, periodici, audiovisivi e software della National Library of Medicine. Per OLDMEDLINE, gli abstract di convegni e HSRProj, Gateway è l'unica interfaccia di ricerca disponibile, mentre DIRLINE è interrogabile anche autonomamente,¹⁰ al pari di LOCATORplus.¹¹ Nel 1999 LOCATORplus ha inglobato i tre database catalografici noti in precedenza come CATFILE, CATFILEplus e SERFILE.¹²

strumenti, software utilizzati nella ricerca sanitaria e nelle scienze sociali.¹³

La letteratura tossicologica, al pari di quella demografica e oncologica, è ora indicizzata in archivi autonomi, dotati di un proprio sito web. Dalla sua creazione nel 1972, il database bibliografico TOXLINE (TOXicology Literature onLINE) si è costantemente evoluto nella struttura e nei metodi di connessione fino alla disponibilità gratuita su Internet Grateful Med (1998) e all'inclusione, assieme alle banche dati fattuali di chimica e tossicologia, nella rete informativa TOXNET, liberamente disponibile sul web dal 1999.¹⁴ Il nucleo di TOXLINE è diventato a sua volta un *subset* di MEDLINE/PubMed, mentre la NLM ha rafforzato la componente tossicologica di MEDLINE indicizzando 40 riviste in precedenza spogliate da TOXLINE.

POPLINE, l'archivio di abstract di convegni di argomento demografico, non è stato convertito nel Gateway. In compenso, una vasta mole di citazioni di letteratura demografica è stata riversata in MEDLINE/PubMed (articoli di riviste) e in LOCATORplus (monografie),¹⁵ mentre ha visto la luce un Population Information Program (PIP), a cura del John Hopkins University Center for Communication

¹⁰ URL: <<http://dirline.nlm.nih.gov/>>. Ultima visita 27/12/2002.

¹¹ URL: <<http://locatorplus.gov/>>. Ultima visita 27/12/2002.

¹² Il database CATFILE è la fusione di AVLINE (catalogo di audiovisivi e software biomedici) e CATLINE (catalogo di monografie pubblicate dal XV secolo ad oggi). CATFILEplus cumula i record di CATFILE con quelli di monografie e capitoli di miscellanee in precedenza indicizzati da HISTLINE, HealthSTAR, SPACELINE, BIOETHICSLINE e POPLINE. Il database SERFILE, catalogo dei periodici dal 1665 ad oggi, può considerarsi il successore in formato MARC del vecchio SERLINE.

¹³ URL: HSTAT <<http://hstat.nlm.nih.gov/hq/Hquest/screen/HquestHome/s/55433>>; HSRR <http://www.nlm.nih.gov/nichsr/hstrr_search/>. Ultima visita 02/01/2003.

¹⁴ URL: <<http://toxnet.nlm.nih.gov/>>. Ultima visita 27/12/2002. I database fattuali accessibili in TOXNET sono HSDB (Hazardous Substances Data Bank), IRIS (Integrated Risk Information System), CCRIS (Chemical Carcinogenesis Research Information System), GENETOX, TRI (Toxics Chemical Release Inventory), ChemIDplus, HSDB Structures e NCI-3D. Sulle vicende di TOXLINE vedi: *Next generation TOXLINE*, "NLM Technical Bulletin", 318 (2001), January-February; URL: <http://www.nlm.nih.gov/pubs/techbull/jf01/jf01_toxline.html>. Ultima visita 02/01/2003.

¹⁵ *Population-related citations added to MEDLINE/PubMed*, "NLM Technical Bulletin", 328 (2002), September-October; URL: <http://www.nlm.nih.gov/pubs/techbull/so02/so02_popline.html>. Ultima visita 02/01/2003.

Programs (JHU/CCP), che ha prodotto un omonimo ma ben più ambizioso POPLINE database.¹⁶

Nel novembre 1997 il database di letteratura oncologica CANCERLIT (in precedenza CANCERLINE) del National Cancer Institute (NCI), contenente citazioni e abstract di documenti pubblicati a partire dagli anni Sessanta su oltre 4.000 fonti di diversa estrazione (riviste biomediche, proceeding, monografie, rapporti tecnici, tesi di dottorato), è divenuto consultabile gratuitamente sul sito web del NCI assieme a PDQ (Physician Data Query), un archivio "operativo" composto da varie categorie di materiali.¹⁷

Sul versante tecnologico i cambiamenti sono stati altrettanto radicali.

Il vecchio software di gestione di MEDLARS, ELHILL, ha smesso di funzionare il 30 settembre 1999,¹⁸ sostituito da Entrez, il software che permette la consultazione incrociata di PubMed e delle basi di dati correlate del NCBI. Entrez, distribuito per la prima volta in versione sperimentale su cd-rom nel 1991, è al tempo stesso un sistema d'indicizzazione e recupero dell'informazione di tipo testuale, una collezione di dati da fonti eterogenee e un principio di organizzazione della letteratura scientifica. Il sistema si basa su due presupposti: che l'informazione biomedica può essere organizzata in "nodi", corrispondenti ai diversi database bibliografici (PubMed) e fattuali (database di genetica e biologia molecolare), ciascuno ottimizzato per la tipologia di dati ospitati, e che per poter confluire in un "nodo" l'informazione di un certo tipo deve essere "oggettivata" (in una pubblicazione o in un report univocamente identificabili) e al tempo stesso strutturalmente "aperta", predisposta cioè a "legarsi" con informazioni dello stesso o di altri nodi. Ad esempio, dato il record di una sequenza proteica, il sistema deve permettere di risalire, tramite un link ipertestuale, alla sequenza nucleotidica che la codifica; quest'ultima, a sua volta, deve potersi linkare alla citazione dell'articolo che la descrive, dall'abstract della citazione deve essere possibile contestualizzare i concetti chiave mediante un link con i libri online in cui sono discussi, e così via. Alimentando le connessioni tra i record di nodi differen-

ti, Entrez ambisce a diventare un sistema per la scoperta di nuova informazione, un *discovery system* anziché un semplice *retrieval system*:

By making links between selected nodes [...] Entrez is designed to infer relationships between different data that may suggest future experiments or assist interpretation of the available information, although it may come from different sources.¹⁹

Senza entrare nel merito di questa pretesa, che veicola un preciso modello di organizzazione e crescita della conoscenza scientifica,²⁰ la scelta di connettere fonti eterogenee all'interno di sistemi tecnologicamente e linguisticamente uniformi è fondamentale per il futuro della comunicazione scientifica:

gli scienziati hanno cominciato a prendere in seria considerazione l'eventualità che le tecnologie digitali possano essere applicate alla realizzazione di strumenti di ricerca con caratteristiche non immaginabili in quelli tradizionali, con l'obiettivo di creare una rete di connessioni tra documenti e risorse referenziali, di integrare le conoscenze e di ricostruire, ove necessario almeno in modo virtuale, i nessi fra entità altrimenti separate.²¹

2.1 Oltre MEDLINE: PubMed

PubMed non è solo una delle possibili interfacce di MEDLINE, ma un archivio in parte diverso e più ampio oltre che un punto di snodo per risorse e servizi correlati al nucleo bibliografico del database. Sotto il profilo dei contenuti, PubMed si contraddistingue per la tipologia delle fonti e la rapidità di aggiornamento: comprende infatti, oltre ai record targati MEDLINE, tre gruppi di citazioni prive d'indicizzazione completa, in particolare prive di parole chiave del Medical Subject Headings, quindi ricercabili solo attraverso i comuni canali di recupero dell'informazione

¹⁶ URL: <<http://db.jhuccp.org/popinform/index.stm>>. Ultima visita 27/12/2002. Il POPLINE della John Hopkins comprende citazioni di monografie, capitoli di miscellanee, articoli di periodici e quotidiani, letteratura grigia (rapporti tecnici, norme e provvedimenti legislativi, relazioni di convegni, tesi e dissertazioni).

¹⁷ CANCERLIT e PDQ si raggiungono a partire dall'URL <<http://cancernet.nci.nih.gov/>>. Ultima visita 27/12/2002. PDQ contiene: schede peer-reviewed con indicazioni sintetiche in merito a terapie, screening, prevenzione, genetica, supporto al paziente oncologico; un registro di sperimentazioni controllate (circa 1.800 aperte e 1.2000 chiuse) in corso o effettuate in tutto il mondo; una directory di medici e professionisti che forniscono consulenza genetica e un elenco di organizzazioni operanti nel settore della cura del cancro.

¹⁸ *NLM discontinues direct access to ELHILL and TOXNET command/menu systems on september 30, 1999*, "NLM Technical Bulletin", 307 (1999), March-April; URL: <http://www.nlm.nih.gov/pubs/techbull/ma99/ma99_elhill.html>. Ultima visita 27/12/2002.

¹⁹ JIM OSTELL, *The Entrez search and retrieval system*, in *The NCBI handbook*, Bethesda, NCBI, National Library of Medicine, 2002, p. 14.4; URL: <<http://www.ncbi.nlm.nih.gov/books/data/handbook/pdfs/ch14d1.pdf>>. Ultima visita 27/12/2002.

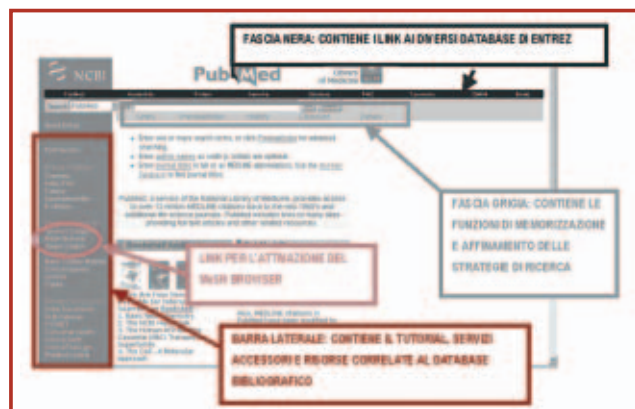
²⁰ Il richiamo alla *serendipity* per esaltare le possibili implicazioni scientifiche delle connessioni tra i database di Entrez ha un background consolidato nella storia e sociologia della scienza. Vedi: ROBERT K. MERTON – ELINOR BARBER, *Viaggi e avventure della serendipity. Saggio di semantica sociologica e sociologia della scienza*, Bologna, Il Mulino, 2002.

²¹ PAUL GABRIELE WESTON, *Dal controllo bibliografico alle reti documentarie: il catalogo elettronico nella prospettiva dell'interoperabilità fra sistemi eterogenei*, "Biblioteche oggi", 20 (2002), 7, p. 44.

(autore, titolo, parola libera). Il primo gruppo è dato dalle citazioni di articoli in corso d'indicizzazione, spesso i più recenti e *up to date* (Premedline o In process). Il secondo insieme è quello delle citazioni trasmesse direttamente dagli editori in formato XML (PubMed - as supplied by publisher): si tratta di una categoria composta di materiali, in gran parte destinati all'indicizzazione completa, che comprende sia le citazioni di articoli di riviste biomediche pubblicate prima che le riviste stesse entrassero nel circuito MEDLINE, sia le citazioni di articoli in formato elettronico *ahead of print* (o *epub*), cioè nella fase anteriore alla loro pubblicazione a stampa, sia infine le citazioni *out of scope* di riviste parzialmente indicizzate da MEDLINE (ad esempio le citazioni di articoli di astrofisica o di chimica pubblicati su periodici d'impostazione multidisciplinare come "Science", la cui sezione biomedica è normalmente indicizzata da MEDLINE). Fa parte di questo gruppo anche ciò che resta di un esperimento fallito, utile comunque a testimoniare la novità e popolarità delle *bibliographic linking features* di PubMed: PubRef, un complesso di circa 100.000 citazioni di articoli non biomedici, derivate principalmente dalla *Physical reviews series*, aggiunte al database su pressione della National Academy of Sciences e di alcuni editori nel 1999.²² Un terzo set di citazioni spurie si riferisce agli articoli di PubMed Central (PMC), costantemente arricchito da periodici biomedici (MEDLINE e non-MEDLINE) che sottoscrivono la politica del *free access*. L'home page di PubMed innesta sul modulo centrale della query box tre serie di âncore a servizi e banche dati della NLM: una barra laterale e due orizzontali (fascia nera e fascia grigia) (vedi figura 2).

La fascia nera contiene i link ai database di genetica e biologia molecolare e a un'interfaccia semplificata per la consultazione dell'archivio full-text di PubMed Central. La fascia grigia permette di attivare le funzioni avanzate di limitazione, tematica o cronologica, della ricerca (*limits*), di scorrimento degli indici e previsualizzazione dei risultati (*preview/index*), di recupero e raffinamento progressivo delle ricerche già effettuate (*history*), di memorizzazione temporanea dei record selezionati (*clipboard*), di verifica del modo in cui Entrez ha interpretato la strategia di ricerca (*details*). La barra laterale contiene, oltre alle âncore di supporto verso tutorial e *utilities*, una serie di risorse correlate (*related resources*) esterne al database, e un insieme di servizi accessori (*PubMed services*) che servono a potenziare la ricerca in varie direzioni: un database delle riviste indicizzate in MEDLINE (*journals database*), il MeSH Browser, un form per il recupero mirato di citazioni delle quali si possiedono già alcune informazioni (*single citation matcher*; *batch citation matcher*), un sistema preimpostato di filtri per le ricerche di tipo clinico e il recupero

Fig. 2



delle rassegne sistematiche (*clinical queries*)²³ e uno spazio riservato per la memorizzazione permanente delle strategie di ricerca e la personalizzazione dei link esterni ai record bibliografici (Cubby, LinkOut).

I risultati delle ricerche su PubMed possono essere memorizzati in un'area provvisoria (*clipboard*), quindi stampati e scaricati in vari formati. Partendo da un record che si reputa interessante è possibile trovarne altri di argomento affine mediante il link *related articles*, che lavora su set precalcolati di citazioni "pesate" da un algoritmo. La strategia di ricerca usata in una data occasione può essere memorizzata in modo permanente come URL nei bookmark del proprio browser, al pari di qualunque altra pagina web, ma la funzione più raffinata in tal senso è quella del *cubby* (in italiano "nicchia", "angolo riservato"), che rende possibile, previa registrazione e procedura di login, la memorizzazione permanente di una strategia di ricerca allo scopo di ripetere la stessa interrogazione periodicamente e recuperare, di volta in volta, i nuovi record inseriti nel database. Questa procedura rimpiazza di fatto SDILINE (Selective Dissemination of Information onLINE), il vecchio archivio MEDLARS che raccoglieva, su base mensile, le citazioni più recenti di MEDLINE, comprese quelle dei fascicoli mensili dell'*Index medicus* in corso di stampa, e rende PubMed per certi aspetti simile a un servizio di alerting e SDI. Come vedremo più avanti, una seconda, fondamentale funzione del Cubby è quella di fungere da area di lavoro per la personalizzazione dei link che rinviano dai record bibliografici ai fornitori, commerciali e non, di articoli full-text (LinkOut).

Perfezionando una tecnica già presente in Internet Grateful Med, PubMed va incontro all'utente inesperto e gli offre la possibilità di effettuare una ricerca a testo libero "intelligente". È sufficiente infatti digitare nella *query*

²² *PubRef to be removed from PubMed*, "NLM Technical Bulletin", 318 (2001), January-February; URL: <http://www.nlm.nih.gov/pubs/techbull/jf01/jf01_pubref.html>. Ultima visita 02/01/2003.

²³ Nel 2002 la NLM ha cominciato a indicizzare per MEDLINE il *Cochrane database of systematic reviews* (a partire dal fascicolo 2 del 2000) e ha sviluppato un filtro per la ricerca mirata delle rassegne sistematiche: *Cochrane database of systematic reviews included in Medline*, "NLM Technical Bulletin", 324 (2002), January-February; URL: <http://www.nlm.nih.gov/pubs/techbull/jf02/jf02_cochrane.html>. Ultima visita 02/01/2003.

box uno o più termini che rappresentano l'argomento dei documenti cercati per ottenere un risultato che, sebbene difetti in precisione, spesso ha una buona percentuale di richiamo. Questo accade perché i termini inseriti senza ulteriore specificazione subiscono un processo di "mappatura" automatica (*automatic term mapping*), vengono cioè confrontati, ed eventualmente associati in fase di recupero, con quattro tabelle di terminologia controllata predisposte dal sistema, principalmente voci del Medical Subject Headings e del metathesaurus che lo contiene: l'Unified Medical Language System. Ad esempio, una ricerca per "manic depression AND prozac" viene automaticamente tradotta in: "((bipolar disorder)[MeSH Terms] OR manic depression[Text Word]) AND ("fluoxetine"[MeSH Terms] OR prozac[Text Word])". In realtà, a fronte della polisemia e ambiguità del linguaggio naturale adoperato dai non specialisti per interrogare la banca dati, il processo di "mappatura" non garantisce risultati sempre e comunque corretti. Per essere incanalato automaticamente verso la forma "ammessa", infatti, un termine o una combinazione di termini deve comunque rientrare nella gamma di varianti lessicali previste e implementate dal sistema. Ecco allora che, ad esempio, "nose bleed" o "nosebleed" vengono automaticamente (e correttamente) reinterpretati come "epistaxis", mentre la forma equivalente "bloody nose" non riceve lo stesso trattamento, e un acronimo comunemente associato alla progettazione assistita dal calcolatore (CAD), il cui impiego in campo odontotecnico ha una solida tradizione, viene addirittura tradotto con "coronary arteriosclerosis". L'idea di interrogare MEDLINE/PubMed con la stessa leggerezza con cui s'interroga un motore di ricerca full-text, scrivendo un termine nella casella di ricerca e cliccando su "go", ha un certo fascino, ma non tiene conto della complessità strutturale del database, e il modo in cui un motore di ricerca traduce una *query* dell'utente in linguaggio controllato

has a fundamental limitation; namely, that it does not dialog with the user to resolve ambiguity or to offer alternatives.²⁴

L'alternativa nondimeno esiste ed è quella di imparare a colloquiare con la banca dati mediante le regole del linguaggio controllato che gli indicizzatori applicano quotidianamente alla designazione del contenuto semantico dei documenti: il Medical Subject Headings o MeSH.

2.2 Medical Subject Headings e Unified Medical Language System

MEDLINE è una banca dati unica nel suo genere perché, sin dagli albori, ha consegnato al ricercatore uno strumento fondamentale per interpretarne il funzionamento e orientarsi nel labirinto della letteratura biomedica. Il MeSH è un dizionario di terminologia controllata usato per indicizzare gli articoli di MEDLINE e gli altri documenti acquisiti dalla NLM, ma che negli ultimi anni ha allargato la propria sfera d'applicazione, fungendo da modello per la revisione di soggettari locali e investendo la catalogazione delle risorse Internet all'interno di motori di ricerca specializzati come HON, OMNI, CliniWeb, l'indice delle patologie del Karolinska Institutet Library.²⁵ Le voci (*headings*) del MeSH che identificano i singoli concetti sono presentate sia in ordine alfabetico che in una complessa struttura gerarchica ramificata dal generale al particolare (*MeSH tree structures*). In sintonia con la tradizione delle *rules* cutteriane, gli indicizzatori assegnano a ogni citazione i soggetti (con relative suddivisioni o *subheadings*) più specifici adatti a rappresentarne il contenuto semantico. Ciò garantisce un'alta percentuale di precisione delle ricerche, bilanciata tuttavia dalla possibilità, offerta dai software di information retrieval, di scegliere se "esplodere" automaticamente la ricerca per una data voce recuperando tutte le citazioni indicizzate con le voci gerarchicamente inferiori a quella iniziale.

Il MeSH aspira a una standardizzazione della terminologia medica che costituisce, al pari dell'interoperabilità di protocolli e sistemi informatici, un presupposto essenziale dello sviluppo di reti informative efficienti. La sua importanza risalta più chiaramente se si considera che non è un dizionario isolato, ma solo una delle fonti terminologiche del Unified Medical Language System (UMLS), un progetto più ampio presentato come

a long term NLM research and development effort designed to facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources.²⁶

In sostanza, l'UMLS dovrebbe garantire in prospettiva lo sviluppo di software capaci di effettuare "mappature" o traduzioni sempre più precise del linguaggio naturale in linguaggio controllato, incrementando la precisione dell'information retrieval. Il nucleo del progetto è un metathesaurus che raccoglie, raggruppa per concetti e integra

²⁴ MARGARET H. COLETTI – HOWARD L. BLEICH, *Medical Subject Headings used to search the biomedical literature*, "Journal of the American Medical Informatics Association", 8 (2001), 4, p. 322; vedi anche: LORA B. VAULT, *Variations in Medical Subject Headings (MeSH) mapping: from the natural language of patron terms to the controlled vocabulary of mapped lists*, "Journal of the Medical Library Association", 90 (2002), 2, p. 174-178.

²⁵ URL dei siti menzionati sono rispettivamente: <<http://omni.ac.uk/>>, <<http://www.hon.ch/>>, <<http://www.ohsu.edu/clinweb/>>, <<http://www.mic.ki.se/Diseases/index.html>>. Ultima visita 27/12/2002. Un'esperienza italiana di rielaborazione del soggettario locale alla luce del MeSH è quella della biblioteca dell'Istituto superiore di sanità: MARIA ALESSANDRA FALCONNE – PAOLA FERRARI, *Le nuove scelte di catalogazione semantica della biblioteca dell'Istituto superiore di sanità*, "Bollettino AIB", 42 (2002), 3, p. 333-336.

²⁶ NATIONAL LIBRARY OF MEDICINE, *UMLS knowledge sources*, 13th ed., U.S. Department of Health and Human Services, National Institutes of Health, National Library of Medicine, 2002, p. 10.

Fig. 3

informazioni semantiche su una miriade di termini biomedici estrapolati da thesauri, nomenclature e sistemi di classificazione, generali o specializzati, creati e mantenuti da organizzazioni diverse a livello internazionale. Fanno parte dell'UMLS, ad esempio, l'International Classification of Disease (ICD), la Systematic Nomenclature of Medicine (SNOMED), il Library of Congress Subject Headings (LCSH) e le diverse edizioni del MeSH tradotto nelle principali lingue europee, tra cui l'italiano. Nel metathesaurus tutte le parole e le frasi che hanno lo stesso significato formano un concetto distinto, una classe di sinonimi. I concetti sono correlati tra loro e organizzati per categorie, dal generale al particolare, all'interno di una complessa rete semantica (*semantic network*), mentre un apposito lessico specializzato in lingua inglese (*SPECIALIST lexicon*) definisce le varianti lessicali per una parte della terminologia biomedica predisponendo le informazioni sintattiche, morfologiche e ortografiche necessarie allo sviluppo di sistemi di "mappatura" automatica del linguaggio naturale (Natural Language Processing System o NLP).²⁷

L'uso del thesaurus nella ricerca bibliografica consente di ridurre al minimo l'incidenza dei sinonimi e al tempo stesso di evitare il recupero di articoli che contengono il termine cercato senza che questo rappresenti un tema rilevante dello studio. A tale proposito la NLM mette a disposizione dell'utente due versioni online del Medical Subject Headings: la prima, più sintetica, è raggiungibile direttamente dall'home page di PubMed ed è integrata nel modulo di ricerca bibliografica; la seconda, più estesa e con molteplici funzionalità di recupero e scorrimento dei termini, ma indipendente da PubMed, si raggiunge dal sito ufficiale della NLM.²⁸ Un confronto tra le

due versioni rivela differenze interessanti non tanto sotto l'aspetto delle procedure e delle modalità d'interrogazione, ma soprattutto per quello che concerne le potenzialità di ricerca e recupero dei metadati (i descrittori di soggetto) necessari per una ricerca efficace dei dati, cioè delle citazioni di letteratura biomedica targate MEDLINE. Infatti nella versione indipendente, che può essere consultata in una seconda finestra del browser durante una sessione di PubMed, è possibile sia effettuare ricerche per radice o frammento di parola su tutte le partizioni del dizionario, compreso l'indice dei sinonimi o *entry terms*, sia leggere per esteso il contenuto dei diversi campi del record di ciascun MeSH, comprese le note di catalogazione, sia, infine, limitare l'interrogazione a campi specifici, ad esempio i descrittori principali (*main headings*), i qualificatori (*qualifiers* o *subheadings*), i concetti relativi a sostanze chimiche (*supplementary concepts*).

Il MeSH è un dizionario in continua evoluzione: nuove voci vengono introdotte, altre sono abbandonate o sostituite, la rete di *cross references* si espande quotidianamente per aiutare i ricercatori a orientarsi nella congerie di sinonimi e gli informatici a preparare software per la "mappatura" del linguaggio naturale. Alla luce di tale evoluzione, che riflette le trasformazioni del linguaggio e delle pratiche scientifiche sedimentate nella letteratura specializzata, la possibilità di leggere per intero le note di catalogazione e la storia dei descrittori (*annotations*, *online notes*, *history notes*), assente dal MeSH Browser di PubMed, può risultare decisiva. Nel caso delle note storiche, la cosa è tanto più evidente se si pensa che quasi mai il lavoro pionieristico in un dato settore e su un dato argomento è indicizzato con il MeSH specifico per quell'argomento. In una ricerca sulla scintigrafia miocardica, ad esempio, una volta accertato che il MeSH per la tecnica scintigrafica è "radionuclide imaging", solo leggendo attentamente il record in corrispondenza del campo "Annotation" troviamo un'indicazione decisiva per il perfezionamento della strategia: laddove si tratti della scintigrafia miocardica, la combinazione di voci da preferire è "heart/radionuclide". L'*history note* inoltre ci dice che per recuperare le citazioni anteriori al 1977, anno d'introduzione del MeSH "radionuclide imaging", la voce da usare è "radioisotope scanning". Analogamente, solo dall'annotazione storica apprendiamo che per trovare articoli sul trapianto polmonare anteriori al 1990 dobbiamo usare la combinazione MeSH/subheading "lung/transplantation" anziché il MeSH "lung transplantation".

Il MeSH online, in ultima analisi, colma le lacune più evidenti del MeSH Browser di PubMed e dell'*automatic term mapping* in quanto, come sostengono due tra i suoi ideatori,

allows you to find descriptors of interest without assuming knowledge of MeSH's often complex vocabulary structure and rules.²⁹

²⁷ Una concisa introduzione al progetto dell'UMLS si trova in: STUART J. NELSON – TAMMY POWELL – BETSY L. HUMPHREYS, *The Unified Medical Language System (UMLS) project*, URL: <<http://www.nlm.nih.gov/mesh/umlsforelis.html>>. Ultima visita 27/12/2002.

²⁸ URL: <<http://www.nlm.nih.gov/mesh/2002/MBrowser.html>>. Ultima visita 27/12/2002.

²⁹ *New MeSH browser available on the web*, "NLM Technical Bulletin", 307 (1999), March-April; URL: <http://www.nlm.nih.gov/pubs/techbull/ma99/ma99_meshbrowser.html>. Ultima visita 27/12/2002.

3. Dalla citazione bibliografica ai documenti: il LinkOut

Entrez è una rete di banche dati progettata per esaltare le connessioni interne ed esterne tra diverse unità informative.³⁰ Un record di uno dei database di Entrez può essere linkato a un altro record del sistema oppure a risorse esterne al NCBI (LinkOut). Un tipico esempio di link interno è quello tra la citazione di un articolo e le citazioni degli articoli affini per argomento oppure il link, di recente introduzione, tra la citazione di un articolo e quelle dei commenti che ne analizzano il significato in termini di Evidence Based Medicine.³¹ I link esterni invece provengono da editori, aggregatori, biblioteche, centri di sequenziamento del DNA, siti web per lo specialista e per il paziente. Gli URL delle risorse esterne sono trasmessi e costantemente aggiornati dalle persone o dalle organizzazioni che le hanno create o ne detengono il possesso e che ne stabiliscono le modalità di accesso. La separazione tra i record di dati (ad esempio gli abstract di PubMed) e le coordinate dei link esterni, memorizzate in un apposito database di Entrez,

enables both the external link providers and NCBI to manage linking in a flexible manner [...] linking information can be updated as frequently as necessary.³²

Tra le possibili risorse esterne, una delle più importanti per le biblioteche è costituita dai siti degli editori e aggregatori di riviste elettroniche che, in modo gratuito o sulla base di precise restrizioni contrattuali, mettono a disposizione degli utenti il full text degli articoli citati in PubMed. In questa particolare accezione, il LinkOut consente di superare il livello puramente bibliografico della ricerca e di trovare uno sbocco naturale in quello, ben più concreto e rilevante per l'utente, della visualizzazione immediata del full text, creando un punto d'intersezione importante tra il concetto tradizionale di "bibliografia", una lista di titoli elencati secondo vari criteri, e quello, più orientato alla prassi del recupero dei documenti, di "catalogo".

3.1 Il LinkOut al servizio della comunità scientifica: PubMed Central

Il tradizionale paradigma della comunicazione scientifica in campo biomedico, la catena commerciale autore-editore-utente/acquirente, è entrato in crisi profonda alla fine degli anni Novanta.³³ L'aumento vertiginoso dei costi degli abbonamenti a fronte dei budget inadeguati di biblioteche ed enti di ricerca, la necessità di garantire una diffusione rapida e puntuale dei risultati delle ricerche (e la possibilità di farlo grazie a Internet), la "decostruzione" del tradizionale concetto di "periodico" alla luce di documenti digitali dalla fisionomia nuova e cangiante, di versioni web delle riviste diverse dalle versioni a stampa (vedi "Pediatrics", "FASEB Journal", la formula *electronic-long/paper-short* di BMJ), di articoli pubblicati *ahead of print* e addirittura *instead of print*,³⁴ sono solo alcuni dei fattori che hanno indotto il cambiamento. Un'editoria *web-based* è ormai comunemente accettata, per molti addirittura inevitabile, ma il peso organizzativo e finanziario del modello commerciale non tende a diminuire. Una delle vie di fuga da questa situazione imbarazzante è stata individuata nell'allestimento di archivi di letteratura biomedica universalmente e liberamente accessibili. Unanime il consenso sugli obiettivi, perlomeno tra scienziati e bibliotecari, diverse sono state invece le formule proposte per raggiungerli e i soggetti entrati in campo: editori che sponsorizzano server di preprint sul modello dell'arXiv di Los Alamos (NetPrint, sponsorizzato da BMJ e HighWire Press); organizzazioni intese a favorire lo sviluppo di soluzioni decentrate di autoarchiviazione della letteratura scientifica unitamente a strumenti di recupero dei metadati necessari a garantirne l'accesso universale (OAI); editori impegnati a ospitare sui propri siti riviste *peer reviewed* e a pubblicare riviste elettroniche a prezzi contenuti (BioMed Central, HighWire Press); organizzazioni di scienziati e biblioteche che stringono coalizioni con editori (PloS, SPARC) o chiedono il supporto finanziario della comunità scientifica e accademica (BOAD) allo scopo di fondare nuove riviste gratuite e incoraggiare la libera circolazione della documentazione scientifica.³⁵ In questa cornice estremamente

³⁰ KENT A. SMITH – ED SEDUEIRA, *Linking at the US National Library of Medicine*, "Learned Publishing", 14 (2001), p. 23-28; BARRY A. PALEVITZ, *Linking up with LinkOut*, "The Scientist", 16 (2002), 6, p. 27.

³¹ *Linking MEDLINE citations to Evidence-Based Medicine assessments and summaries*, "NLM Technical Bulletin", 326 (2002), May-June; URL: <http://www.nlm.nih.gov/pubs/techbull/mj02/mj02_ebm.html>. Ultima visita 02/01/2003.

³² KATHY KWAN, *LinkOut: linking to external resources from Entrez databases*, in *The NCBI handbook*, Bethesda, NCBI, National Library of Medicine, 2002, p. 16.3; URL: <<http://www.ncbi.nlm.nih.gov/books/data/handbook/pdfs/ch16d1.pdf>>. Ultima visita 27/12/2002.

³³ EUGENIO PELIZZARI, *Crisi dei periodici e modelli emergenti nella comunicazione scientifica: uno spazio d'azione per le biblioteche*, "Biblioteche oggi", 20 (2002), 9, p. 46-56; BARRY P. MARKOVITZ, *Biomedicine's electronic publishing paradigm shift*, "Journal of the American Medical Informatics Association", 7 (2000), 3, p. 222-229; ANTONELLA DE ROBBIO, *Evoluzione e rivoluzione dei periodici elettronici*, "Bibliotime", 3 (2000), 1; URL: <<http://www.spbo.unibo.it/bibliotime/num-iii-1/derobbio.htm>>. Ultima visita 02/01/2003.; IDEM, *Open Archive. Per una comunicazione scientifica 'free online'*, "Bibliotime", 5 (2002), 2; URL: <<http://www.spbo.unibo.it/bibliotime/num-v-2/derobbio.htm>>. Ultima visita 01/01/2003.

³⁴ TONY DELAMOTHE, *Is that it? How online articles have changed over the past five years*, "BMJ", 325 (2002), p. 1475-1478.

³⁵ The Public Library of Science (PloS), URL: <<http://www.publiblibraryofscience.org/>>; Scholarly Publishing and Academic Resources Coalition (SPARC), URL: <<http://www.arl.org/sparc/home/index.asp?page=0>>; Budapest Open Access Initiative (BOAI), URL: <<http://www.soros.org/openaccess>>; Open Archive Initiative (OAI), URL: <<http://www.openarchives.org/>>; Netprints, URL: <<http://clinmed.netprints.org/home.dtl>>; BioMed Central, URL: <<http://www.biomedcentral.com/>>; HighWire, URL: <<http://highwire.org/>>. Ultima visita 02/01/2003.

fluida si colloca l'iniziativa del NIH di un archivio "istituzionale" di letteratura periodica integrato con il database bibliografico PubMed.

PubMed Central,³⁶ nato da un'idea di Harold Varmus e lanciato ufficialmente nel febbraio 2000, è un archivio digitale di letteratura biomedica sviluppato e mantenuto dal NCBI con il preciso scopo di preservare e rendere gratuitamente accessibili, nel più breve lasso di tempo possibile dopo la loro pubblicazione, i risultati della ricerca sperimentale, già filtrati e validati dai comitati di revisione delle riviste di provenienza. Gli articoli di PubMed Central sono tutti rigorosamente linkati ai record di PubMed, ma a differenza di BioMed Central, le cui riviste sono comunque raggiungibili a partire da PubMed, PubMed Central non è un editore commerciale, non chiede soldi agli autori e, sebbene inizialmente fosse previsto un canale di archiviazione dei preprint, non pubblica articoli esenti da *peer review*. Alla base dell'iniziativa c'è lo stesso criterio ispiratore di PubMed e di GenBank, cioè la convinzione che un archivio omogeneo di dati estratti da fonti diverse e riversati in un contenitore organizzato secondo standard condivisi possa fungere da stimolo per "intentional and serendipitous discoveries".³⁷ Gli editori che decidono di aderire al progetto, per ora limitato alla letteratura di lingua inglese, devono garantire che le riviste depositate soddisfino alcuni requisiti di "qualità", quali l'inclusione in uno dei repertori internazionali d'indicizzazione (MEDLINE, Biosis, EMBASE, PsycINFO, Agricola, Science Citation Index) o la presenza nei rispettivi comitati editoriali di almeno tre scienziati cooptati in progetti di ricerca finanziati da grosse organizzazioni americane o estere. I full text vengono trasmessi a PubMed Central, assieme alle immagini ad alta risoluzione (file TIFF o PostScript), sotto forma di file SGML o XML e convertiti in uno specifico DTD, basato sui più recenti standard XML, che definisce in maniera univoca i tag per ciascun'unità informativa (autore, ente di appartenenza, testo ecc.). Ciò semplifica ulteriormente sia le operazioni di ricerca sul testo completo dell'articolo che la connessione con altri segmenti informativi (record bibliografici, genomi, strutture macromolecolari, libri online). Inizialmente era previsto che i full text, consegnati al NCBI appena pubblicati o dopo un periodo variabile da sei mesi a un anno dalla pubblicazione, venissero caricati sul server di PubMed Central, il cui staff avrebbe provveduto a formattarli in maniera appropriata e a collegarli con le citazioni bibliografiche di PubMed. L'ostilità degli editori, i

quali temevano (e temono tuttora) di perdere una quota consistente degli introiti legati ai diritti economici sui lavori pubblicati, ha condotto i responsabili del progetto, nel marzo 2001, a rendere più flessibile la procedura concedendo che la visualizzazione dei full text avvenga sul sito degli editori stessi, alle condizioni da loro stabilite, e che il copyright rimanga all'editore o all'autore secondo quanto specificato (la gratuità dev'essere comunque garantita, ma il limite di tempo non è rigido).³⁸ Malgrado le restrizioni temporali sul *free access* e sul deposito in PMC, gli editori sono comunque tenuti a trasmettere una copia dei lavori al NCBI, in modo da agevolare quelle procedure di "normalizzazione" dei formati e "interconnessione" dei dati di diversa natura e provenienza che abbiamo visto caratterizzare complessivamente il sistema Entrez e che restano essenziali per garantire tanto la preservazione e l'accessibilità nel tempo delle pubblicazioni (significativa la scelta di SGML e XML), quanto l'"esplosione" della loro potenzialità informativa derivante dalla contiguità con materiali già presenti nello stesso o in altri archivi NCBI.

3.2 Il LinkOut al servizio della biblioteca: la personalizzazione del database

Sebbene iniziative di archiviazione "aperta" delle fonti della ricerca e della pratica scientifica stiano faticosamente affiorando anche al di fuori dei tradizionali settori della matematica e della fisica, la maggior parte della letteratura periodica corrente nel settore biomedico resta saldamente nelle mani degli editori commerciali. Dato un archivio di citazioni bibliografiche, il LinkOut permette comunque di tracciare la via più breve per recuperare in tempo reale i documenti cui le citazioni rinviano. In tempo reale significa due cose: recupero immediato per gli articoli fulltext che sono già disponibili online sia gratuitamente, su PubMed Central o direttamente sul sito dell'editore, sia a pagamento per le riviste di cui si possiede (a titolo privato o istituzionale) un contratto di licenza o per le quali si accettano i costi di un pay-per-view; recupero dilazionato in poche ore o in alcuni giorni per gli articoli di riviste possedute da biblioteche con le quali si intende attivare un rapporto di document delivery.

Dal punto di vista dell'utente-ricercatore, l'utilità del LinkOut è data dalla possibilità di personalizzare la visualizzazione dei link ai siti di tutti i potenziali fornitori di ar-

³⁶ URL: <<http://www.pubmedcentral.nih.gov>>. Ultima visita 27/12/2002. Alla fine di dicembre 2002 sono presenti in PMC 48 riviste indipendenti (più 7 annunciate come *forthcoming*) e l'intero subset delle riviste BMC di BioMed Central (57 titoli). Sul dibattito che ha accompagnato l'iniziativa di Varmus vedi: MICHAEL B. EISEN – PATRICK O. BROWS – HAROLD E. VARMUS, *Public-access group supports PubMed Central*, "Nature", 419 (2002), 6903, p. 111; RICHARD J. ROBERTS, *PubMed Central: the GenBank of the published literature*, "PNAS", 98 (2001), 2, p. 381-382; TONY DELAMOTHE, *PubMed Central: creating an Aladdin's cave of ideas*, "BMJ", 322 (2001), 6, p. 1-2; DEBORAH GESENSWAY, *PubMed Central: government as publisher?*, "Annals of internal medicine", 133 (2000), p. 841-844; ADDEANE S. CAELLEIGH, *PubMed Central and the new publishing landscape: shifts and tradeoffs*, "Academic Medicine", 75 (2000), 1, p. 4-10; MICHAEL DAY, *The scholarly journal in transition and the PubMed Central proposal*, "Ariadne", 21 (1999); URL: <<http://www.ariadne.ac.uk/issue21/pubmed>>. Ultima visita 01/01/2003.

³⁷ *PubMed Central overview*; URL: <<http://www.pubmedcentral.nih.gov/about/intro.html>>. Ultima visita 27/12/2002.

³⁸ EDWIN SEQUEIRA – JOHANNA MCENTYRE – DAVID LIPMAN, *PubMed Central decentralized*, "Nature", 410 (2001), p. 740.

ticoli full-text che hanno segnalato il proprio posseduto al NCBI. La modifica si effettua all'interno del Cubby mediante la configurazione delle *LinkOut preferences* e ha effetto solo e unicamente fintanto che si resta nello spazio riservato. Le possibilità offerte dal sistema sono due: aggiungere (Add Icon) o nascondere (Hide from LinkOut) un'icona che funge da ancora verso il sito di un fornitore. L'opzione Add Icon incide direttamente sulla pagina dell'abstract aumentando il numero delle icone di fornitori disponibili, ovvero dei canali esterni percorribili per il recupero dei full text. Nell'esempio di figura 4 si è scelto di visualizzare come provider il CISTI (Canada Institute for Scientific and Technical Information).

Dopo aver salvato le preferenze, lanciando una ricerca nel Cubby e visualizzando l'abstract delle citazioni si vedrà comparire, per una parte di esse, l'icona cliccabile del CISTI come potenziale link per la fornitura di full text.

Dal punto di vista della biblioteca, la modifica del LinkOut coinvolge direttamente la politica di gestione degli accessi alle collezioni locali di e-journal. Una biblioteca titolare di una licenza per l'accesso a periodici elettronici indicizzati da MEDLINE può far comparire, nei record bibliografici selezionati durante una ricerca, un link personalizzato al full-text degli articoli contenuti in tali periodici, a patto ovviamente che gli editori delle riviste linkate siano *LinkOut participants*, abbiano cioè già fornito al NCBI i dati necessari (gli URL) per accedere ai propri e-journal. Il link, sotto forma di un'icona cliccabile, sarà attivo sempre e soltanto per gli utenti che si collegano a PubMed dai pc dell'intranet della biblioteca (o del sistema bibliotecario) che effettua la registrazione e comparirà ogniqualvolta, al termine di una sessione di ricerca, si visualizzeranno gli abstract. Ovviamente il collegamento mantiene le restrizioni imposte all'accesso dal tipo di contratto stipulato con l'editore, pertanto, se l'accesso a una rivista è mediato dall'indirizzo IP, l'utente potrà leggere la rivista solo da computer con indirizzo di rete compreso nel range di IP della biblioteca; se invece si tratta di un accesso mediato da password, l'utente abilitato dovrà inserire la password corretta per una data rivista.³⁹

Questa procedura è stata implementata con successo dalla Biblioteca medica centralizzata di Modena nell'estate del 2002. La sua diffusione contribuirebbe, a nostro avviso, a creare un importante canale di condivisione e reciproca



Fig. 4

trasparenza del patrimonio di riviste online delle biblioteche italiane, aggiungendo ai repertori tradizionali (ACNP e i molteplici cataloghi di riviste elettroniche allestiti dalle singole biblioteche o dai singoli poli bibliotecari) il peso scientifico e organizzativo di un database internazionale della portata di MEDLINE.⁴⁰

Per gli articoli che non sono disponibili immediatamente in formato elettronico è verosimile che l'utente singolo o la biblioteca si organizzi per richiederne una copia a un fornitore commerciale o a una biblioteca. Loansome Doc, introdotto per la prima volta nel 1991, può essere considerato l'estensione alle biblioteche non statunitensi di DOCLINE, il sistema automatizzato di prestito interbibliotecario e document delivery della NLM e del National Network of Libraries of Medicine (NN/LM). Attraverso Loansome Doc e previa registrazione gratuita, un utente privato o istituzionale può richiedere direttamente a una o più biblioteche prescelte (affiliate a DOCLINE) il full text degli articoli selezionati durante una sessione di PubMed. Il server della NLM preposto allo smistamento delle richieste è l'attore invisibile dell'intero processo limitandosi a predisporre le infrastrutture necessarie ad avviare e coordinare la comunicazione reciproca tra un utente "remoto" di una banca dati citazionale e una biblioteca altrettanto "remota". La ricerca bibliografica trova così la sua naturale conclusione innescando una procedura automatica di localizzazione e recupero dei documenti nella quale viene lasciata piena facoltà all'utente finale di impostare la variabile più importante dell'intero percorso: la scelta del fornitore. La Biblioteca medica centralizzata dell'Università di

³⁹ L'elenco delle biblioteche che partecipano al LinkOut (per l'Italia finora solo la Biblioteca medica centralizzata dell'Università di Modena) si trova all'URL: <http://www.ncbi.nlm.nih.gov/entrez/journals/active_libraries.html>. Ultima visita 01/01/2003.

⁴⁰ La partecipazione al LinkOut, ovvero la procedura di registrazione del posseduto elettronico di una biblioteca presso il NCBI si svolge in tre fasi. Il primo passo è quello di contattare via e-mail il NCBI manifestando la volontà della biblioteca di far figurare il proprio posseduto nel LinkOut. Dopo circa un mese, quest'ultima riceverà in posta elettronica un messaggio con username e password necessari per procedere nella registrazione. Il passo successivo consiste nel preparare due file da inviare al NCBI: il primo è una scheda della biblioteca (*identity file*), il secondo è un file che descrive il posseduto (*holdings file*) e identifica i provider delle riviste online cui la biblioteca è in grado di fornire l'accesso. L'ultima operazione da compiere è una modifica dell'indirizzo di connessione alla banca dati. Una volta che i LinkOut files sono stati elaborati e inviati al NCBI, la loro attivazione avviene infatti solo dopo aver cambiato l'URL con il quale la biblioteca si connette a PubMed. Il nuovo URL avrà la seguente struttura: <[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding="medlib"](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?holding=)>, dove "medlib" è l'username attribuito alla biblioteca per il LinkOut. Le istruzioni dettagliate per partecipare al LinkOut si trovano in: *LinkOut and library holdings*; URL: <<http://www.ncbi.nlm.nih.gov/entrez/linkout/doc/liblinkout.html>>. Ultima visita 01/01/2003.

Modena e Reggio Emilia ha verificato l'efficienza e la semplicità di funzionamento del Loansome Doc sperimentando come fornitore DOCLINE la Deutsche Zentral Bibliothek für Medizin di Colonia, una delle colonne portanti del sistema SUBITO.

Tanto PubMed Central che Loansome Doc, in ultima analisi, rivelano l'orientamento della NLM verso una soluzione "moderata" dei paradossi sollevati dalle dinamiche della *scholarly communication* in ambienti tecnologicamente evoluti, tuttavia le scelte tecniche votate all'interoperabilità dei sistemi, alla standardizzazione dei formati e dei protocolli si adattano, in prospettiva, anche a soluzioni completamente diverse. Pertanto, se è vero che la linea di confine è demarcata con chiarezza tra l'autoarchiviazione delle pubblicazioni al di fuori dei tradizionali meccanismi di peer review e la sedimentazione controllata di letteratura "certificata" in un server istituzionale finanziato e "controllato" dalla comunità scientifica ufficiale (PubMed Central), tra lo scambio paritario e autogestito di file in circuiti di *peer to peer networking* stile epigoni di Napster (da Gnutella a FastTrack) e un server targato NLM che smista le richieste dell'utente e le indirizza a un provider liberamente selezionato dall'utente stesso (Loansome Doc), è anche vero che fa parte della natura ambigua del "confine" la possibilità di contemplare i due scenari collocati sulle sponde opposte allo scopo di valutarne attentamente le differenze, i limiti, le potenzialità.

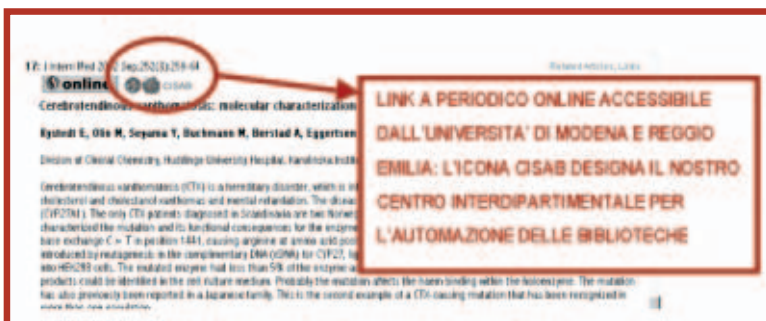
4. Conclusioni

L'accumulo di grandi quantità di dati e risultati di ricerche sperimentali è stata la linea di tendenza predominante nelle scienze biomediche durante l'ultimo decennio. Affinché tali informazioni possano produrre un approfondimento significativo delle conoscenze e delle pratiche terapeutiche occorre tuttavia disporre di valide reti informative che ne garantiscano la sintesi e il recupero mirato alla luce di nuove ipotesi scientifiche. La contemporanea esplosione delle reti telematiche e dei protocolli d'integrazione di linguaggi e applicazioni operanti su piattaforme differenti ha reso ancora più impellente il compito di organizzare e "legare" informazioni e servizi all'interno di archivi multifunzione, superando ripartizioni tradizionali come quella tra banche dati primarie e secondarie, fattuali e bibliografiche o, più in generale, tra archivi bibliografici e cataloghi. PubMed, la versione web ufficiale della banca dati MEDLINE, ha intrapreso questo cammino a partire dalla seconda metà degli anni

Fig. 5



Fig. 6



Novanta. Il lavoro di controllo terminologico sul linguaggio naturale nell'alveo dell'UMLS (*automatic term mapping*); la messa a punto e il perfezionamento di uno strumento essenziale per la formulazione di strategie di ricerca efficaci quale il MeSH Browser; l'adattabilità a esigenze particolari di recupero selettivo dell'informazione (Cubby); il link diretto con i full text per gli articoli depositati in PubMed Central e BioMed Central o gratuitamente disponibili presso i siti degli editori; la possibilità di personalizzare i link con fonti informative e servizi esterni al database, in particolare il link con i fornitori del full text degli articoli in formato elettronico (LinkOut); la possibilità di automatizzare il servizio di document delivery e di attivare le richieste di copie dei documenti dall'interno di PubMed verso fornitori commerciali o istituzionali prescelti dall'utente (Loansome Doc): tutte queste opportunità, che costituiscono il punto di forza dei sistemi commerciali di archiviazione e distribuzione dell'informazione scientifica,⁴¹ integrano in modo decisivo il substrato bibliografico del database e permettono a ricercatori e bibliotecari di tutto il mondo di trasformare una base dati citazionale in uno strumento potente di organizzazione e recupero dell'informazione medica al servizio della pratica clinica e della ricerca primaria.

⁴¹ Vedi l'orientamento di Elsevier verso la crescente integrazione di database bibliografici e fattuali attraverso il portale BioMedNet e Embase.com; URL: <http://www.e-products.elsevier.com/biomedical/index.html>. Ultima visita 01/01/2003.