

Indagine comparativa tra database bibliografici

MEDLINE, EMBASE e PASCAL Biomed a confronto

di Antonella De Robbio

Con il presente studio si sono comparati i tre maggiori database bibliografici di ambito biomedico, MEDLINE, EMBASE e PASCAL Biomed, per la rilevazione delle coperture dei periodici indicizzati a livello di singolo database e a livello di sovrapposizioni tra più database (in particolare MEDLINE ed EMBASE), anche in riferimento ai paesi di pubblicazione dei periodici.

L'obiettivo principale dell'indagine è quello di fornire al ricercatore medico o al clinico, ma anche al bibliotecario che deve effettuare delle valutazioni per l'acquisizione di banche dati del settore, uno strumento informativo che possa essere di supporto per orientarlo sulle scelte.

Per la rilevazione dei dati si sono utilizzati più database:

- quelli sottoposti a valutazione, a carattere biomedico, e precisamente i tre citati sopra, più Current Contents (come riferimento, ma non come oggetto di indagine);
- quelli utilizzati quale strumento informativo di indagine, due database di periodici, uno generale Ulrich's e l'altro specialistico SERLINE.

Ai fini del recupero delle informazioni si sono utilizzate le tecniche di information retrieval sulle banche dati di periodici, che sono servite da strumenti contenitori di quei periodici a carattere biomedico su cui si è poggiata l'indagine.

Per il controllo e comparazione dei dati sono state anche utilizzate le liste dei periodici indicizzati dalle singole basi di dati recuperate dal web, in particolare le liste dei periodici delle sezioni di Current Contents sul sito statunitense dell'ISI (Institute for Scientific Information) e le liste relative a PASCAL Biomed sul sito francese dell'INIST (Institut de l'information scientifique et technique) del CNRS (Centre national de la recherche scientifique).

Le sei banche dati utilizzate per il presente lavoro si riferiscono tutte alle versioni su cd-rom per il Sistema di Database Networking ERL (Electronic Reference Library) della Silver Platter, con software di recupero dell'informazione SPIRS, interrogate sia con connessione via client su workstation per ambiente operativo Windows (WinSpirs) sia in accesso via Rete attraverso interfaccia WebSPIRS.

Si è provveduto anche ad effettuare un'indagine sulle ulteriori versioni esistenti per ogni database bibliografico di riferimento, con collegamenti a versioni su siti remoti disponibili gratuitamente, MEDLINE Plus attraverso il servizio PubMed della NLM (National Library of Medicine) di Bethesda, o attraverso richiesta di password provvisorie per trial, per quelle banche dati ad accesso a pagamento.

Lo scopo era quello di effettuare una valutazione in merito alle seguenti condizioni:

- analisi sulla struttura dei dati/record in relazione alla presenza di campi comuni;
- presenza o meno di abstract, e in che lingua;
- presenza o meno di tesauro o schema classificatorio ai fini del recupero dell'informazione;
- comparazione dei risultati ottenuti per alcune strategie di ricerca sui differenti database scientifici;
- crescita relativa agli ultimi dieci anni dei periodici di ambito biomedico;
- numero dei periodici correnti indicizzati in relazione ai periodici biomedici totali e in relazione ai periodici esistenti in generale;
- numero dei periodici correnti indicizzati da ciascun database e loro suddivisione per paese di pubblicazione;
- comparazione relativa alla copertura dei periodici relativamente a sovrapposizioni tra database e tra aree geografiche.

SCHEDE INFORMATIVE SULLE BASI DI DATI UTILIZZATE PER LA PRESENTE RICERCA

Le basi di dati differiscono strutturalmente in riferimento a più fattori:

- versione della stessa banca dati su differente supporto (on-line tradizionale, interfaccia web, cd-rom, per piattaforma database networking); differenti interfacce e differenti linguaggi per l'interrogazione;
- esistenza di più versioni/edizioni della banca dati anche per stessa tipologia di supporto (parti di database seleziona-

te in base a periodi o ambiti specializzati);

– versioni con aggiornamenti differenti (cd-rom con aggiornamenti semestrali, trimestrali, mensili).

Le basi di dati differiscono anche per diversità di produttore e possono avere:

– struttura differente dei record (differenze nei campi: per numero o per nome/etichetta di campo);

– riferimenti temporali differenti (copertura del database parziale o totale);

– interfaccia e linguaggio di interrogazione.

Date le differenze di cui sopra è evidente come il risultato di una ricerca possa cambiare per lo stesso database in riferimento a prodotti differenti.

L'analisi di comparazione condotta per questo lavoro si è basata sulle sei banche dati disponibili su cd-rom per il sistema di database networking ERL, ma si è riferita anche, come specificato sopra, ad altre versioni ai fini di indagine e di supporto.

Current Contents

Current Contents è il database multidisciplinare prodotto dall'ISI (Institute for Scientific Information) di Filadelfia, per la letteratura scientifica internazionale corrente. La sua forma elettronica primaria è su dischetto floppy, uno alla settimana per ciascuna delle sette sezioni, come per la forma cartacea. Esiste anche su cd-rom, in varie versioni differenti per i sistemi di database networking (OVID e ERL), con aggiornamenti sempre settimanali. Recentemente esiste anche la versione on-line disponibile in rete con interfaccia di accesso web presso il sito <<http://www.isinet.com>>.

Ognuna delle sette sezioni di CC copre un gruppo di discipline correlate, allo scopo di provvedere ad una copertura il più possibile esaustiva. Le sezioni si riferiscono allo spoglio di più di 8.700 riviste scientifiche internazionali. I dati bibliografici sono comprensivi di abstract ad eccezione della sezione Scienze sociali e comportamentali.

Le sezioni sono quindi le seguenti:

– Life Sciences (oltre 1.600 periodici selezionati);

– Agriculture, Biology and Environmental Sciences (oltre 1.100 periodici selezionati);

– Physical, Chemical and Earth Sciences (oltre 1.100 periodici selezionati);

– Clinical Medicine (oltre 1.000 periodici selezionati);

– Engineering, Technology and applied sciences (oltre 1.200 periodici selezionati);

– Social and Behavioral Sciences (oltre 1.500 periodici selezionati);

– Arts & Humanities (oltre 1.200 periodici selezionati).

Il repertorio è estremamente rapido nel mettere a disposizione citazioni e abstract: da controlli effettuati si può dire che l'articolo uscito sulla rivista appare circa 15 giorni dopo su Current Contents (talvolta pari al tempo necessario per l'arrivo via posta della rivista).

Gli aggiornamenti nei sistemi di database networking avvengono via FTP direttamente on-line.

Per l'ambito biomedico quindi le due sezioni di riferimento sono Life Sciences (oltre 1.600 periodici selezionati) e Clinical Medicine (oltre 1.000 periodici selezionati) con sovrapposizione di circa il 20% tra le due sezioni relativamente ai periodici coperti.

posizione di circa il 20% tra le due sezioni relativamente ai periodici coperti.

Nelle altre sezioni vi possono essere periodici con articoli di interesse biomedico (bioingegneria, biofisica, modelli matematici applicati alla medicina ...) che non rientrano nelle due sezioni di settore biomedico.

Nella versione on-line o per database networking, comprensiva di tutte e sette le sezioni, è possibile quindi effettuare ricerche trasversali per materie correlate, recuperando articoli di altri settori ma sull'argomento di interesse: in questo modo si raggiungono informazioni "borderline", per ricerche applicative che altrimenti non verrebbero recuperate dai database specifici, ottenendo una visione più ampia rispetto all'argomento trattato e soprattutto una pluralità d'offerta per "punti di vista differenti".

MEDLINE

È il database a carattere biomedico della Biblioteca nazionale medica degli Stati Uniti (US NLM) di Bethesda, uno dei più importanti per il settore, probabilmente il più noto per la numerosità delle versioni differenti disponibili anche gratuitamente.

Equivalenti elettronici dei repertori cartacei *Index Medicus* (bibliografia mensile che copre oltre 3.000 periodici) *Index to Dental Lit*, *International Nursing Index*, MEDLINE copre la letteratura biomedica internazionale dal 1966 ad oggi.

Contiene oltre 9 milioni di citazioni bibliografiche con abstract, relativi alla letteratura biomedica mondiale nei settori ricerca, clinica, oncologia, amministrazione, odontoiatria e scienze infermieristiche, con incremento di oltre 400.000 notizie ogni anno.

Seleziona articoli tratti dallo spoglio di oltre 4.000 riviste biomediche internazionali pubblicate in 70 paesi, più capitoli e articoli da monografie con disponibilità di author abstract per circa il 70% delle citazioni a partire dal 1975.

Il 74% del contenuto è in lingua inglese.

È possibile ricercare nel database oltre che con termini tratti dal linguaggio naturale a testo libero, anche attraverso termini indicizzati, sinonimi, e intestazioni di soggetto tratte dal Thesaurus della US National Library of Medicine MeSH (Medical Subject Headings) con possibilità di esplosione dei termini di ricerca. Il MeSH è composto da oltre 18.000 intestazioni di soggetto di ambito biomedico utilizzate in MEDLARS (insieme di oltre 40 banche dati della NLM).

Buona parte delle informazioni si riferiscono ad articoli di medicina clinica e quindi il database assume quel carattere di rilevanza non solo per la ricerca medica in generale, risultando un valido strumento informativo per i differenti ambiti medici, soprattutto per l'immediatezza con cui gli articoli di periodici vengono indicizzati nel database.

Molte sono le versioni disponibili:

– **On-line tradizionale con connessione su host computer (gruppo MEDLARS);**

– **On-line con interfaccia web**

Dal giugno 1997 la NLM offre MEDLINE gratuitamente sul web, attraverso il servizio PubMed con aggiornamento settimanale dei dati e con inclusione delle informazioni recenti, non ancora indicizzate, relative ai dati PreMedline <<http://> ➤

www.ncbi.nlm.nih.gov/PubMed/>. L'interrogazione di MEDLINE da PubMed prevede tre livelli: "basic search" in linguaggio naturale (che consente l'uso del carattere * per il troncamento), "advanced search" con ricerca per campi e con operatori booleani e infine quello per medici e bibliotecari "clinical queries", che permette le strategie di ricerca più mirate. È possibile anche accedere dal servizio IGM Internet Grateful Med sempre offerto dalla NLM, che consente l'accesso gratuito a decine di banche dati del gruppo MEDLARS. Attualmente sono 350 i periodici full-text raggiungibili da PubMed.

Esistono anche alcune interfacce differenti da PubMed offerte da case farmaceutiche, oppure interfacce gateway ai servizi di accesso NLM. Una ottima interfaccia, gratuita e in lingua italiana che consente l'interrogazione di PubMed, è quella messa a disposizione dalla Dematel (Servizi informativi telematici per la medicina e la ricerca biomedica) <<http://www.dematel.it/cgi-bin/med/med.pl>> con caratteristiche piuttosto diverse dalla versione originale, tutte orientate verso una consultazione facilitata, ma anche con possibilità di ricerche trasversali di utilizzo clinico.

Per una rassegna su web vedere: <<http://www.math.unipd.it/~adr/database/medlars.htm>>.

- Su cd-rom

Molte sono le versioni su cd-rom disponibili (per OVID, per ERL), e compatibili per l'installazione all'interno dei sistemi di database networking. Le diversità si riconducono a differenze sull'aggiornamento periodicità, ai riferimenti temporali per fette di database, a selezioni di parti di database, o, come nel caso di MEDLINE Advanced, di record muniti di campi aggiuntivi e funzionalità trasversali per le ricerche attraverso TAG.

- **Medline Clinical Collection KR-OnDisc** - versione Dialog OnDisc: una selezione di MEDLINE con particolare riferimento alla medicina clinica, sono indicizzati oltre 150 periodici in lingua inglese specializzati nelle discipline cliniche.

- **Versioni per OVID e versioni per ERL** (Express, Professional, Advanced...).

EMBASE (Excerpta Medica)

Base dati bibliografica specializzata in letteratura medica e farmacologica, versione elettronica del repertorio *Excerpta Medica*.

Comprende oltre 7 milioni di registrazioni bibliografiche con abstract per l'80% dei record, tratte da oltre 4.000 periodici internazionali, dal 1974 in avanti, con un incremento di oltre 425.000 nuove registrazioni ogni anno.

I dati sono immessi 15 giorni dopo l'arrivo del fascicolo cartaceo.

Embase è prodotta dalla Elsevier Science BV, suddivisa in una ventina di sezioni (disponibili eventualmente separatamente) e copre la letteratura internazionale biomedica. La copertura si riferisce in particolare alla farmacologia, e alle discipline mediche: oncologia, cardiologia, pediatria, gastroenterologia, angiologia, neurologia, psicologia, scienze biologiche, e comprende anche argomenti quali: medicina tropicale, sanità pubblica, ingegneria biomedica, medicina alternativa.

EMBASE offre la possibilità di utilizzare per la ricerca gli EMTAGS, identificatori che indicano aspetti generali applicabili teoricamente a tutti i descrittori che sono a loro volta suddivisi in categorie e sono codificati in campi "limit".

EMTREE Thesaurus è l'albero classificatorio che offre 37.000 termini medici e relativi a sostanze e farmaci.

Esistono versioni su cd-rom per il database networking, OVID e ERL, a partire dal 1980 (disponibili anche le singole sezioni).

EMBASE è disponibile anche on-line (DataStar, Dialog, DIMDI, Lexis/Nexis, Ovid Online, STN).

Sul sito HealthGate <<http://www.healthgate.com>> è possibile accedere alla banca dati o alle sezioni, con recupero delle informazioni a pagamento (\$ 2.00 a informazione recuperata a campi completi).

PASCAL Biomed

PASCAL è la base dati francese prodotta da INIST (Institut de l'information scientifique et technique) del CNRS (Centre national de la recherche scientifique), corrisponde alla bibliografia internazionale ex Bulletin Signaletique, e per i documenti a partire dal 1973 la banca dati fornisce anche il titolo in inglese.

Contiene oltre 2.600.000 referenze bibliografiche provenienti da oltre 6.000 periodici europei ed internazionali, con particolare riferimento alla letteratura scientifica francese, nei campi: scienza dell'informazione, tecnologia, fisica, chimica, medicina, biologia, scienze applicate. La ripartizione della base dati è più o meno la seguente: 74% lingua inglese, 10% lingua francese, 7% lingua russa, 5% lingua tedesca, 4% altre lingue.

PASCAL Biomed è la sezione medica del database PASCAL con registrazioni bibliografiche di articoli di periodici, tesi di dottorato, atti di convegni, rapporti, monografie e brevetti di argomento biomedico, ricavate da PASCAL.

Fra le discipline considerate biologia applicata, omeopatia, medicina, botanica, farmacologia, tossicologia, biotecnologia, psicologia, ecc.

PASCAL Biomed contiene anche 175.000 notizie provenienti dalla Banque de Donnés Santé Publique.

Peculiarità di questa base dati è la possibilità di recuperare i record dai campi descrittori in tre lingue europee: inglese, francese, spagnolo, per oltre 80.000 termini. Molti degli abstract sono in lingua francese.

PASCAL Biomed copre quindi 6.000 periodici, di cui 3.200 *cover-to-cover*, per gli altri periodici effettua selezione in relazione agli articoli biomedici. PASCAL esiste in edizione KR-OnDisc, Dialog On-Disc Site Enhanced.

È disponibile la versione cd-rom PASCAL Biomed per piattaforma database networking ERL a partire dal 1996.

Ulrich's

Ulrich's International Periodicals Directory, è prodotto da R.R. Bowker, divisione di Reed Elsevier Inc., e include oltre 220.000 dati di periodici regolarmente pubblicati in oltre 200 paesi. Sono compresi anche oltre 50.000 periodici cessati dal

1979. Il database offre informazioni utili ai fini dell'acquisizione dei periodici con campi relativi agli editori (circa 90.000).

SERLINE

Repertorio contenente informazioni bibliografiche su quasi 90.000 seriali a carattere biomedico. Il database include tutti i seriali e i congressi numerati posseduti nella collezione della Biblioteca medica nazionale di Bethesda. Alcuni periodici non posseduti, ma indicizzati comunque da MEDLINE, risultano presenti.

Esiste la versione *free* sul sito della NLM (IGM) raggiungibile via telnet.

ANALISI COMPARATIVA DELLA STRUTTURA DEI DATABASE

Diamo un'occhiata d'insieme comparativa ad alcune caratteristiche peculiari a ciascuno dei database esaminati, così come risultano nella tabella in fondo alla pagina.

Il record di EMBASE è decisamente più strutturato, 41 campi, rispetto a quello di MEDLINE con 33 campi e, nella versione Advanced, i campi aggiuntivi rispetto alle altre versioni sono 7.

EMBASE comprende vari campi relativi alle sostanze e ai farmaci, nome della sostanza, ditta produttrice, classificazione della sostanza, descrittore della sostanza, sequenza molecolare...

PASCAL Biomed è ancora più strutturato, con ben 42 campi per record, ma il maggior numero lo si riscontra in Current Contents che conta ben 45 campi per record. Relativamente a PASCAL l'alto numero di campi si può ricondurre alla triplicazione dei campi DE (descrittori) e identificatori, riportati in tre lingue diverse.

Tra EMBASE e MEDLINE vi sono 15 campi in comune, con stessa etichetta e stesso significato relativamente al contenuto del campo; la sovrapposizione con PASCAL Biomed e gli altri due database è di 8 campi.

È da tenere in considerazione che vi sono per tutti e tre i

database campi etichettati in modo differente ma che si riferiscono a stessi concetti. Per esempio i campi relativi all'indicizzazione del documento, sono sempre campi di interesse semantico o classificatorio, ma si riferiscono ai differenti sistemi di indicizzazione utilizzati o agli schemi di riferimento propri di ogni singolo database.

Altro esempio, il campo relativo al numero standard, per alcuni database IS per altri ISSN, per un altro ancora IS comprende sia ISSN che ISBN.

In sette casi abbiamo invece stessa etichetta di campo per significati diversi e in questi casi sarà necessario prestare attenzione in fase di ricerca comune su più database.

Per esempio il campo etichettato EM per EMBASE si riferisce ai concetti generali di indicizzazione degli EMTAGS, mentre per Current Contents è il campo e-mail address dell'autore.

I campi limite invece sono quei campi che permettono di limitare e di restringere i risultati di una ricerca effettuata su campi su cui è possibile utilizzare operatori matematici o gli operatori di campo. EMBASE ha 13 campi limite, MEDLINE 12 e PASCAL Biomed 9, con una sovrapposizione mista per cinque o sei campi. Nella tabella che occupa le pagine 60, 61 e 62, redatta appunto allo scopo di individuare aree comuni ai fini delle strategie di ricerca, vi è l'elenco dei campi complessivi con segnati in neretto i campi limite.

Greenhalgh,¹ nel suo articolo apparso nel 1997 sul "British medical Journal", evidenzia brevemente alcune importanti differenze tra MEDLINE ed EMBASE rilevate da ricerche effettuate sui due database bibliografici, relativamente alla letteratura medica nel periodo luglio-agosto 1996.

La metodologia utilizzata è stata la seguente: ricerca in entrambi i database utilizzando le strategie di ricerca da tesoro o intestazioni di soggetto o per parola libera. Le ricerche condotte prima su un database, poi sul secondo, hanno portato al recupero di articoli selezionati sulla base di rilevanze trovate o nel campo titolo, o nell'abstract o attraverso i campi indicizzati o relativi ai termini descrittivi di soggetto. Le query poste sono state 32, di cui 11 ritenute non rilevanti. Sulle rimanenti 21 si è condotta la comparazione che ha portato a stessi risultati tra i due database per sole quattro domande poste. EMBASE ha recuperato articoli unici per 14 delle domande poste, mentre MEDLINE ha riportato ➤

DATABASE	N. totali reg. bibliografiche	Incremento annuo	Presenza di abstract	Campi record (limit)	Tesoro	Periodici coperti
MEDLINE Advanced	9 milioni (dal 1966)	Oltre 400.000	68% (dal 1975)	33 (12) TAG	MesH 18.000 intestazioni	4.592
EMBASE	7 milioni (dal 1974)	Oltre 400.000	80% EMTAG	41 (13)	EMTREE 37.000 termini: medici e sostanze	4.168
PASCAL Biomed Tratta da PASCAL	2.600.000 (dal 1996 su cd-rom)	Oltre 400.000	In inglese In francese	42 (9)	Ricerca su campi descrittivi, non da tesoro 80.000 termini in francese, inglese e spagnolo	6.885 (di cui <i>cover-to-cover</i> oltre 3.500)

¹ T. GREENHALGH, *How to read a paper: the Medline database*, "British Medical Journal", 19 July 1997, 315, p. 180-183.

Tabella comparativa dei campi record dei quattro database MEDLINE, EMBASE, Current Contents e PASCAL Biomed

AB	Abstract	MED	EMB	CC	PASC Biom
AD	Indirizzo dell'autore	MED	EMB	CC ***	
AF	Ente a cui l'autore afferisce				PASC Biom
AGE	TAG età	MED			
AI	Indicatore presenza di abstract	MED		CC	PASC Biom
AN	Numero identificativo del documento	MED	EMB	CC	PASC Biom
AU	Autore (CA e PA)	MED	EMB	CC	PASC Biom
AUZ	Insieme dei campi autore			CC ***	
BA	Autore del volume			CC	
BL	Livello bibliografico				PASC Biom
BT	Titolo del libro		EMB	CC	
CA	Ente Autore			CC	PASC Biom
CC	Sezione del Current Contents			CC *	
CD	Data e luogo del congresso (per i record dal 1990) CODICE	EMB			PASC Biom
CF	Nome del congresso (dal 1990)	EMB			PASC Biom
CM	Commento	MED			
CN	Numero del contratto o della sovvenzione	MED		CC	
CO	Codice identificativo della rivista		EMB		
CP	Paese di pubblicazione	MED	EMB		PASC Biom
CR	Informazioni sul copyright			CC	PASC Biom
CT	Titolo del congresso (dal 1990)		EMB		
DA	Data di ingresso	MED			
DE	Descrittori (Medical Index Term) (DEE, DEF, DES)		EMB		PASC Biom
DEE	Descrittori MIT in inglese				PASC Biom
DEF	Descrittori MIT in francese				PASC Biom
DES	Descrittori MIT in spagnolo				PASC Biom
DG	Data Academic Degree				PASC Biom
DM	Nome della sostanza prodotta		EMB		
DN	Nome della marca del prodotto		EMB		
DT	Tipologia del documento		EMB	CC	
DR	Descrittore della sostanza		EMB		
EC	Classificazione EMCLAS		EMB		
ED	Editore/i		EMB	CC	
EL	Classificazione EMCLAS della sostanza		EMB		
EM	EMTAGS, concetti generali di indicizzazione Indirizzo e-mail (posta elettronica)		EMB	CC	
EML	Indirizzo e-mail (posta elettronica)		EMB		
GA	Numero identificativo dell'articolo			CC	
GS	Simbolo del gene		MED		
IB	International Standard Book Number			CC	PASC Biom
ID	Identificatori (IDE, IDF, IDS)				PASC Biom

IDE	Identificatori inglese				PASC Biom
IDF	Identificatori francese				PASC Biom
IDS	Identificatori spagnolo				PASC Biom
IN	Istituzione ricerca			CC	
IS	International Standard Serial Number ISBN/ISSN		EMB	CC	PASC Biom
ISSN	International Standard Serial Number	MED			
JC	Categoria del periodico	MED			
JI	Fascicolo della rivista (<i>table of contents</i>)			CC	
JN	Abbreviazione del titolo della rivista Nome del periodico	MED		CC **	PASC Biom
JS	Categoria/soggetto del periodico			CC	
KA	Chiavi assegnate dall'autore			CC	
KP	Chiavi Plus assegnate dall'ISI			CC	
KW	Insieme delle chiavi KA e KP			CC ***	
LA	Lingua originale del documento	MED	EMB	CC	PASC Biom
LC	Numero Library of Congress			CC	
LOC	Localizzazione del documento primario				PASC Biom
LS	Lingua/e del sommario		EMB		PASC Biom
LT	Tipo di letteratura				PASC Biom
MA	Numero di abstract della conferenza			CC	PASC Biom
MESH	Termini del vocabolario controllato o titolo degli argomenti	MED			
MIME	Termini del vocabolario controllato o titolo degli argomenti minori	MED			
MJME	Termini del vocabolario controllato o titolo degli argomenti maggiori	MED			
MN	Ditta produttrice del farmaco o sostanza		EMB		
MS	Sequenza molecolare		EMB		
MT	Tipo di contenuto del documento (natura o tipologia)				PASC Biom
NM	Nome della sostanza	MED			
NR	Numero degli articoli (<i>table of contents</i>) Numero di referenze			CC	PASC Biom
NT	Notes				PASC Biom
OR	Informazioni per l'ordine del volume			CC	
PA	Autore personale				PASC Biom
PB	Editore del libro		EMB		
PD	Descrizione fisica			CC	
PI	Identificatore del editore		EMB		
PS	Nome personale come soggetto	MED			
PT	Tipo di pubblicazione	MED	EMB	CC	PASC Biom
PU	Editore				PASC Biom
PY	Anno di pubblicazione (LT e MT)	MED	EMB	CC	PASC Biom
RAP	Request-A-Print			CC **	
RF	Numero dei riferimenti bibliografici	MED	EMB	CC	

ARGOMENTI

RL	Lingua del documento a cui si riferisce la recensione/revisione			CC	
RN	Numero assegnato alle sostanze chimiche o agli enzimi (CAS)	MED	EMB		
RP	Numero di report Autore cui si mandano richieste di estratti		EMB *	CC	
RPT	Informazioni sul report				PASC Biom
RT	Tipo di record			CC	
RW	Autore del documento a cui si riferisce la recensione/revisione			CC	
RY	Anno di pubblicazione del documento a cui si riferisce la recensione/revisione			CC	
SB	Subset	MED			
SC	Identificatore periodico	MED			
SE	Titolo della collezione			CC	
SH	SubHeading MESH	MED			
SI	Abbreviazione della banca dati in cui è registrata una sequenza molecolare Fonte dell'indicizzazione	MED			PASC Biom
SN	Nome dell'ente universitario o scuola				PASC Biom
SO	Fonte della citazione bibliografica	MED	EMB	CC	PASC Biom
SU	Soggetto		EMB ***		
TG	Checktags (speciali descrittori)	MED	EMB		
TI	Titolo Titolo (T1 e TT)	MED	EMB	CC	PASC Biom
T1	Titolo originale				PASC Biom
TIZ	Insieme dei campi titolo			CC ***	
TN	Nome depositato Numero ufficiale tesi		EMB		PASC Biom
TO	Titolo originale (non in inglese)	MED	EMB		
TS	Titoli		EMB ***		
TT	Titolo tradotto				PASC Biom
UD	Codice di aggiornamento	MED	EMB	CC	
XID	Numero del record in ADONIS (solo per il periodo 1991-93) (La base dati Adonis contiene dati bibliografici full-text)		EMB *		

* i campi non sono ricercabili, ma solo visualizzabili

** i campi non sono né ricercabili, né visualizzabili, servono unicamente per downloading o per document delivery

*** i campi sono unicamente ricercabili, ma non visualizzabili (set di campi ricercabili assieme)

risultati unici solo in due casi.

EMBASE ha "reagito positivamente" con risposte adeguate per 20 delle 21 questioni esaminate, mentre MEDLINE ha risposto in modo pertinente in 17 casi.

Nella ricerca che ho condotto, mi sono limitata a tre query che potessero in qualche modo categorizzare alcuni aspetti specifici e peculiari delle modalità di ricerca intrinseche in ciascuno dei due database e i risultati sono riassumibili in

più punti.

È necessario comunque considerare i tempi di aggiornamento dei record nel database, perché il non tenerne conto può inficiare i risultati nel recupero delle informazioni.

Se si pone una domanda su un determinato argomento in MEDLINE Advanced, laddove l'aggiornamento dei dati avviene settimanalmente via FTP e i dati PREMEDLINE sono presenti, è ovvio che si otterranno risultati più esaustivi rela-

tivamente ai periodi recenti, rispetto ad altre versioni dello stesso MEDLINE, oppure di EMBASE.

Current Contents in certi casi risulta essere ancora più tempestivo, del resto la sua caratteristica principe è quella dell'immediatezza, sebbene gli abstract non siano molto accurati e manchi del tutto di uno schema classificatorio o un tesoro che permetta di orientare le ricerche.

EMBASE è comunque immediato nel mettere a disposizione le informazioni, che risultano essere già inserite solo quindi pochi giorni dopo l'arrivo del fascicolo cartaceo, e anche l'abstract risulta essere presente in percentuale maggiore rispetto a MEDLINE.

Anche l'organizzazione dei periodi temporali/fette di database all'interno dei cd-rom va tenuta in considerazione, altrimenti si dovrà procedere con ulteriori raffinamenti per campi limite su periodi determinati.

Le notizie bibliografiche di EMBASE sono indicizzate in modo più accurato rispetto a MEDLINE, questo è emerso appunto dai "test" effettuati sui due database dove si sono poste le query in modo differente a seconda dell'argomento.

Gli argomenti/query posti si sono riferiti a:

- tecnica Lasik (laser in situ keratomileusis) applicata alla miopia di grado alto;
- utilizzo di nebulizzatori per la somministrazione di farmaci per la terapia dell'asma;
- Sarcoma di Kaposi come patologia secondaria dell'AIDS.

Questa indagine è stata condotta non tanto per cercare di evidenziare articoli comuni recuperati, ma per valutare le tecniche messe a disposizione dai database, gli strumenti offerti in relazione a differenti esigenze informative e le potenzialità complessive, intrinseche ed estrinseche, di ciascun database.

Una valutazione in termini di comparazione/sovrapposizione degli articoli recuperati non sarebbe stata corretta, per due motivi. Il primo di ordine metodologico, in quanto le tecniche per il recupero sono state differenti in ragione delle differenti strutture della rete semantica offerte da ciascun database, l'altro di ordine numerico, in quanto il campione era limitato e comunque non rappresentativo dell'intero contenuto dei database.

L'analisi ha comunque portato a dei risultati di valutazione complessiva.

Si è notato che EMBASE è più preciso rispetto all'assegnazione di termini classificatori, presenti in Emtree, la struttura gerarchica, con presenza di termini anche per specifiche "tecniche" di recente acquisizione, nomi di sostanze o farmaci, per un totale di 37.000 termini descrittivi.

Probabilmente lo schema MeSH di MEDLINE è un po' più lento nell'aggiornamento e immissione di nuove intestazioni di soggetto, in quanto la logica che sottende ai due strumenti è differente: lo schema di MEDLINE è più simile ad uno schema classificatorio, come meccanismo di "crescita e aggiornamento", Emtree è più una derivazione del lavoro di classificazione nei campi EC e EL di termini e sostanze e

dei campi descrittivi DE (Medical Index Term).

Soprattutto relativamente alle sostanze e ai farmaci, EMBASE è senza dubbio più ricco di campi, recuperando più informazioni rispetto a MEDLINE.

I concetti contenuti in EMTAG creano una meta-indicizzazione utile ai fini di recuperi trasversali, sebbene anche MEDLINE Advanced offra la possibilità di "ricercare trasversalmente" attraverso meta-concetti TAG.

Queste peculiarità permettono di reperire informazioni rilevanti per esempio per il campo della *Medicina basata sulle evidenze*, laddove si possono trasversalmente settorializzare tutti gli articoli solo di ambito clinico, oppure tutti quelli con TAG "human" e, nel caso di applicazioni terapeutiche, limitare a campi TAG "Age", per fasce di età.

Relativamente ai tre argomenti posti in query differenti sui singoli database, è emerso che EMBASE recupera informazioni non sufficientemente rappresentate in MEDLINE, ma soprattutto il recupero è più mirato in quanto l'indicizzazione è più efficace, soprattutto per quei settori di medicina clinica, terapia² o per gli ambiti chirurgici.

Analisi più approfondite e studi mirati rispetto alle tecniche di recupero dell'informazione da questi database, potrebbero risultare un utile strumento per stabilire "criteri metodologici sistematici" per l'applicazione dei nuovi modelli di riferimento nell'ambito della Medicina basata sulle evidenze.³

LA CRESCITA DELLA LETTERATURA BIOMEDICA

La banca dati Ulrich's International Periodicals Directory si riferisce a 228.446 periodici.

In Ulrich's si trovano anche i record di 52.137 periodici cessati e 29.576 periodici di cui non si conosce lo stato (se attivo o cessato).

I periodici attivi sono 146.733, alla data del marzo 1999, ma solo 122.430 record nello stato dichiarato "active", sono muniti di campi "data di pubblicazione" e perciò l'analisi sulla crescita incrementale si è potuta condurre solo su questa parte di periodici attivi e non su tutto l'insieme dei 146.733.

Da un'analisi effettuata risultano 10.277 i periodici di ambito strettamente medico, e quasi 5.000 quelli biomedici. Il repertorio infatti è a carattere generale e risulta non esaustivo per l'ambito disciplinare biomedico in esame.

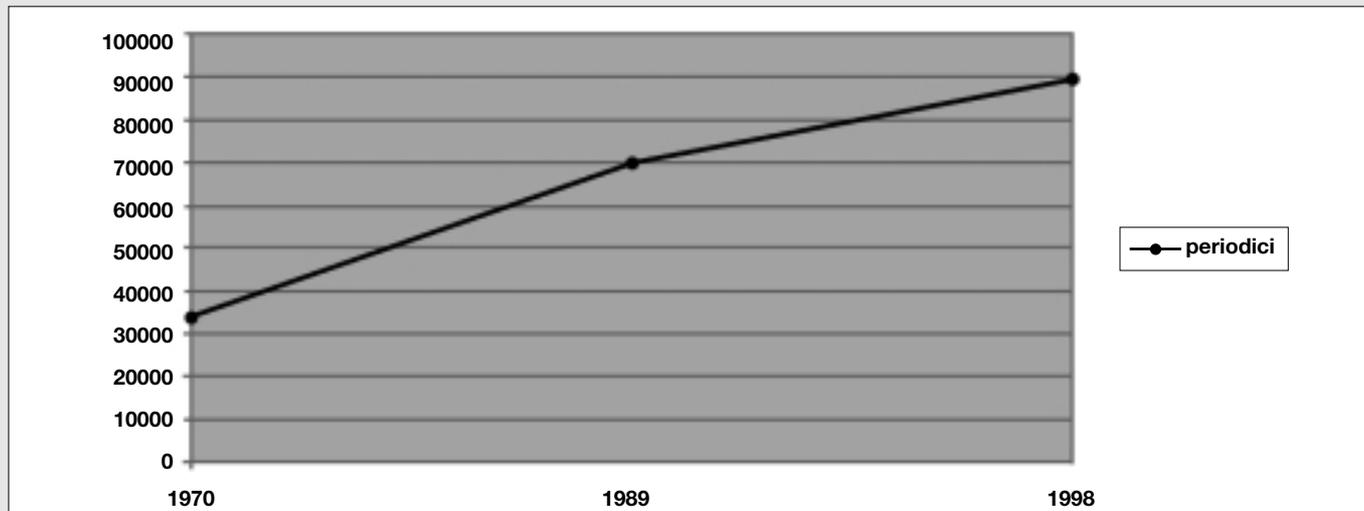
È stato quindi utilizzato come punto di riferimento ai fini del calcolo sulla crescita generale dei periodici, nel corso degli ultimi decenni, al fine di ottenere un dato più generale di contesto: 15.000 periodici biomedici contro i quasi 90.000 elencati da SERLINE, mi hanno fatto propendere per SERLINE come strumento per le indagini più approfondite.

Se andiamo infatti a vedere il movimento periodici all'interno dello strumento Ulrich's vi sono alcuni dati di estremo interesse e che comunque si allineano con le rilevazioni effettuate su SERLINE per quanto riguarda le percentuali (non i dati numerici).

² M.J. BARILLOT - B. SARRUT - C.G. DOREAU, *Evaluation of drug interaction citation in nine on-line bibliographic databases*, "Ann. Pharmacother", 31 (1997), p. 45-49.

³ M. GARDNER, *Information retrieval for patient care*, "British medical Journal", 1997, 314, p. 950-953; A. CARTABELLOTTA, per il Gruppo italiano di medicina basata sulle evidenze (GIMBE), *Verso un'assistenza sanitaria basata sulle evidenze scientifiche: strumenti, competenze, ostacoli*, "Recenti Pro Med", 1997, 88, p. 435-438.

Crescita periodici SERLINE



Se andiamo ad analizzare la presenza di nuovi periodici nel corso degli anni e dividiamo in blocchi alcuni periodi temporali, possiamo infatti notare che il numero dei periodici in generale, ad oggi, si è triplicato rispetto al 1970.

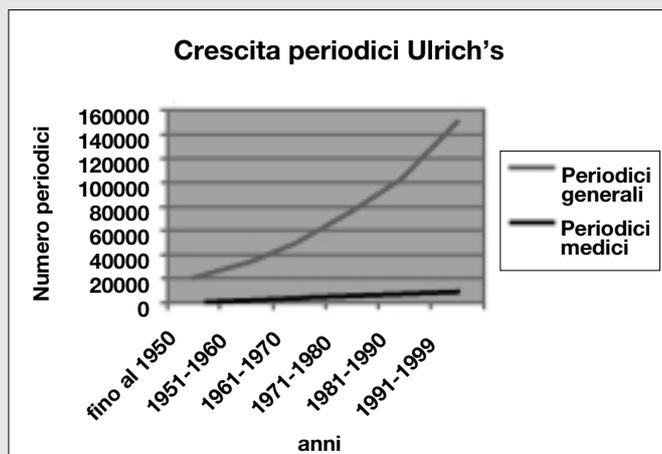
Ulrich's

Per gli anni dal 1900 al 1950 si contano 17.384 periodici, di cui 1.990 solo medici.

A questi vanno aggiunti i 3.884 relativi agli anni 1700-1800 che tuttora sono dichiarati aperti e i 75 che sono datati 1600 come data di inizio di periodico tuttora corrente.

11.699 nuovi ingressi negli anni tra il 1951 e il 1960, mentre dal 1961 al 1970 l'incremento è pari a 17.289, di cui 1.072 per i periodici medici. Dal 1971 al 1980, anni della crescita abnorme della letteratura scientifica, i periodici crescono di ben 25.829 testate, di cui 1.683 per la medicina, e dal 1981 al 1990, vi è un'ulteriore lievitazione di 28.694 unità, con 2.460 per l'area medica.

Crescita periodici Ulrich's



Gli ultimi otto anni degli anni Novanta, hanno registrato una crescita di 17.461 titoli di cui 1.750 medici, dato che sebbene inferiore ai venti anni precedenti, si riferisce alla fine del 1998: mancano ancora tutto il 1999 e probabilmente molti dei dati relativi al 1998 non sono ancora stati registrati all'interno del repertorio. Per il 1999 comunque Ulrich's annuncia ben 112 periodici nuovi di prossima pubblicazione, di cui 29 solo per l'ambito strettamente medico.

In termini di crescita per i periodici in generale (non tanto per quelli biomedici che questo repertorio non rappresenta in modo esaustivo), si possono tracciare alcune linee fondamentali.

Per i periodici relativi agli anni Sessanta la presenza è di 33.000 periodici. Se partiamo dal 1970 abbiamo 50.000 testate presenti nella banca dati con un incremento pari al doppio a fine del 1990, dove si arriva a 104.812 periodici. Il numero prevedibilmente verrà triplicato per la fine del 1999.

SERLINE

Per la specificità del settore mi sono comunque riferita, per un'indagine più approfondita, alla banca dati SERLINE della National Library Medicine di Bethesda, che raccoglie tutti i periodici della biblioteca nazionale statunitense e che produce anche la base dati biografica medica MEDLINE.

SERLINE contiene 89.501 periodici di ambito biomedico, indicizzati sia da MEDLINE che da altri database di settore, oppure non indicizzati da nessun repertorio.

Molti dei record si riferiscono a periodici "spenti" o assorbiti da altri o che hanno cambiato titolo, e questi sono 60.506. I periodici attivi correnti di biomedicina sono 28.995 al dicembre 1998.

Di questi 6.162 sono indicizzati da MEDLINE e da EMBASE. Relativamente alla crescita si può dire che dei circa 33.500 periodici di SERLINE presenti negli anni Settanta (MEDLINE inizia nel 1966), alla fine del 1989 il numero era più che raddoppiato, superando i 70.000 periodici, per arrivare

alla fine del 1998 a 89.501 periodici. In questo modo si arriverà alla fine del secolo, inizio del nuovo millennio, a quota 100.000 periodici biomedici, triplicando il dato di partenza.

L'INDICIZZAZIONE DEI PERIODICI BIOMEDICI EFFETTUATA DA MEDLINE, EMBASE E PASCAL BIOMED

Il numero dei periodici correnti indicizzati in relazione ai periodici biomedici totali e in relazione ai periodici esistenti in generale è stato valutato in questa ricerca da più punti di vista.

Come è stato annunciato sopra, i periodici biomedici correnti presenti in SERLINE sono, alla data del 31 dicembre 1998, 28.995 su 89.501 totali. La parte non corrente, di 60.506 testate, si riferisce a periodici cessati, cambi di titolo, fusioni, assorbimenti...

Dei 28.995 correnti, presenti in SERLINE, non tutti sono indicizzati da qualche repertorio o banca dati.

Il gruppo di banche dati di MEDLARS ne indicizza oltre 8.800, di cui 4.592 coperti dalla sola MEDLINE. Poi vi sono quelle coperte da EMBASE o da altre banche dati di settori non medici.

SERLINE infatti enumera anche tutte quelle riviste di ambiti correlati o affini quali biologia, zoologia, psicologia, agraria, applicazioni scientifiche di ambito biomedico, chimica...

Numerosi sono infatti i periodici indicizzati da grossi database quali Biological Abstract, CAB, Agricola, Chemical Abstract, e altri presenti in SERLINE.

La presente ricerca vuole mettere in luce in particolare le sovrapposizioni tra i due database medici per eccellenza, MEDLINE e EMBASE, quale punto di partenza per ulteriori analisi in comparazione con altri database del settore.

Quest'ultimo dato si riferisce ai periodici coperti dal database PASCAL Biomed, non compreso in SERLINE. Non è quindi stata calcolata l'eventuale area di sovrapposizione tra

questo database e gli altri due, anche se questa indagine potrebbe essere interessante da sviluppare in un secondo momento.

Si è scelto di procedere affiancando anche il database PASCAL Biomed in quanto si riteneva comunque utile un'analisi sulla copertura per area geografica dei periodici indicizzati, al fine di stabilire il "carattere geografico" del database in rapporto con gli altri due.

In ogni caso per l'area di pubblicazione italiana si è provveduto ad effettuare comunque l'analisi per la copertura periodici, in sovrapposizione con gli altri due database, come verrà spiegato in dettaglio più avanti.

NUMERO DI RIVISTE INDICIZZATE DAI TRE DATABASE BIBLIOGRAFICI MEDLINE, EMBASE, PASCAL BIOMED, SUDDIVISI PER PAESE EUROPEO

La tabella riportata nella prima colonna di pagina 66 riassume in dettaglio la copertura periodici effettuata dai tre database, senza sovrapposizioni.

Il calcolo delle aree di sovrapposizione è stato effettuato per MEDLINE e EMBASE per l'intero blocco europeo, con delle analisi singole anche per i paesi quali Francia, Italia, Gran Bretagna, Germania.

Se andiamo a vedere i singoli paesi sembra quasi che le riviste, dalle evidenze numeriche, siano più o meno le stesse, ma nella realtà delle singole basi di dati non è proprio così, in quanto per fare un esempio, il dato dell'Austria che riporta 21 riviste coperte sia da MEDLINE che da EMBASE farebbe presupporre le stesse riviste come oggetto di copertura e invece abbiamo una sovrapposizione di sole 11 riviste. Così pure per il Belgio che richiama una copertura comune di 15 riviste su 28 coperte da MEDLINE e 21 da EMBASE.

È importante però segnalare, da un punto di vista metodologico, che si è preso in considerazione il campo "paese di pubblicazione" in quanto era l'unico riferimento che potesse ricondurre, in qualche modo, ad un'area geografica più o meno definita. Infatti non sarebbe stato lo stesso procedere con recupero di dati dal campo "lingua", perché è noto come quasi tutti i paesi, compreso l'Italia pubblicano riviste italiane in lingua inglese.

È anche vero che molte riviste stanno subendo un processo di "internazionalizzazione", con comitati editoriali misti, scritte in lingua inglese e pubblicate magari in Oriente, e dunque risulta sempre più arduo stabilire la nazionalità di un periodico.

Per alcuni paesi il dato "luogo di pubblicazione" fotografa la situazione reale, per altri si avvicina abbastanza alla realtà, in altri ancora invece è un dato falsato.

In altri paesi quali l'Olanda, dove hanno sede alcuni dei colossi editoriali europei, quando si evidenzia un numero elevato di periodici in quell'area, siamo di fronte a produzioni intellettuali anche (e/o in prevalenza) di altri paesi.

Questo accade anche per la Svizzera ove si pubblicano un gran numero di riviste di altri paesi o a carattere internazionale.

Vi sono due linee da considerare circa le pubblicazioni quale prodotto della ricerca di ogni singolo paese: ➤

Analisi comparativa copertura periodici

Riviste correnti indicizzate da MEDLINE	4.592
di cui di INFERMIERISTICA	384
di cui di ODONTOIATRIA	501
di cui di ONCOLOGIA	189
Riviste correnti indicizzate da EMBASE	4.168
di cui di INFERMIERISTICA (anche in MEDLINE)	29
di cui di ODONTOIATRIA (anche in MEDLINE)	26
di cui di ONCOLOGIA (anche in MEDLINE)	180
Riviste indicizzate in totale dai due database MEDLINE e EMBASE	6.162
Riviste indicizzate in comune tra i due database (sovrapposizione)	2.598
Riviste correnti indicizzate da PASCAL Biomed	6.885
Totale riviste correnti di ambito biomedico	28.995

8.800 riviste sono indicizzate solo dalle quaranta banche dati del Progetto MEDLARS

Copertura periodici per l'intero blocco europeo effettuata sui database in esame

Paese	MEDLINE	EMBASE	PASCAL Biomed
Austria	21	21	17
Belgio	28	21	120
Bosnia Erzegovina	2	1	0
Bulgaria	9	17	4
Croazia	11	22	4
Danimarca	76	44	70
Finlandia	14	5	13
Francia	138	159	1.791
Germania	278	297	513
Gran Bretagna	675	697	997
Grecia	12	11	2
Italia	139	175	103
Iugoslavia	8	4	1
Liechtestein	0	0	1
Lussemburgo	1	1	7
Macedonia	1	0	0
Monaco	1	0	1
Norvegia	34	29	18
Olanda	168	208	359
Polonia	59	36	14
Portogallo	7	10	5
Repubblica Ceca	40	31	16
Romania	14	11	4
Russia	72	41	25
Slovacchia	3	9	1
Slovenia	2	3	0
Spagna	53	86	53
Svezia	25	17	52
Svizzera	165	144	260
Turchia	8	36	0
Ucraina	11	5	2
Ungheria	12	12	0
Totale	2.087	2.153	4.453

– i lavori scientifici relativi alla produzione intellettuale dei ricercatori di un paese possono venire pubblicati sia all'interno del paese dell'autore (su riviste nazionali) oppure su riviste straniere o internazionali, si tratta di percentuali variabili da paese a paese;

– quanti dei lavori pubblicati dalle riviste nazionali si riferiscono effettivamente a produzione intellettuale di quel paese.

Per esempio, per quanto riguarda il nostro paese, ciò che viene prodotto a livello intellettuale in ambito scientifico,

viene pubblicato su riviste straniere o internazionali, oppure su riviste nazionali.

La produzione editoriale di riviste in Italia, registra per la maggioranza articoli di autori italiani, sebbene negli ultimi anni i comitati editoriali di riviste nazionali stanno allargando cercando di recuperare articoli di autori stranieri.

In ogni caso le riviste pubblicate in Italia per la maggior parte dei casi si possono etichettare e come provenienti da "paese di pubblicazione=Italia" e dunque in relazione all'indagine su MEDLINE ed EMBASE il risultato può ritenersi attendibile, non in termini assoluti (quante riviste italiane biomediche esistono e sono pubblicate in Italia), ma in termini relativi (in relazione a quanto indicizzato da questi database).

Ho premesso l'esistenza di questi sconvolgimenti nei movimenti editoriali rispetto alle produzioni della ricerca e ai loro ambiti nativi, al fine di una lettura più critica della tabella qui a fianco riportata e che comunque, rispetto al dato europeo nel suo insieme, può risultare un valido risultato quale strumento di riferimento.

Il totale delle riviste coperte dai tre database va comunque visto in modo più dinamico, sia rispetto al totale degli altri Stati, sia rispetto alle sovrapposizioni tra database.

Nella seguente tabella si specifica in dettaglio l'area della Gran Bretagna, tenendo però conto che per esempio molti degli "Acta Scandinavica" o altre pubblicazioni di area nordica, vengono pubblicate in questo paese.

Solo blocco Gran Bretagna

Gran Bretagna e Irlanda del Nord	MEDLINE	EMBASE	PASCAL Biomed
Scozia	26	32	0
Galles	0	0	0
Irlanda con Irlanda Nord	48	45	14
Inghilterra	601	620	983
Totale	675	697	997

Nessuna rivista indicizzata da nessuno dei tre database in esame per i seguenti paesi europei:

- Albania
- Galles
- Islanda
- Lituania
- Malta
- Moldavia

Inoltre, come si evince dalla tabella riportata nella pagina a fronte, relativamente al contesto europeo (ma questo è valido anche a livello internazionale), vi sono grosse discrepanze numeriche di copertura tra i vari Paesi.

I motivi possono essere due:

- effettiva scarsa produzione intellettuale di quella zona geografico/politica;
- non rappresentatività nei database considerati della produzione scientifica di alcuni paesi.

Vi sono interessanti lavori a riguardo, in particolare per l'area europea, i due articoli sull'indicizzazione di periodici

Analisi dettagliata della copertura periodici relativamente a Gran Bretagna, Francia, Italia e Germania per i tre database MEDLINE, EMBASE, PASCAL Biomed (con valutazione comparativa tra MEDLINE e EMBASE)

Paese	MEDLINE (1)	EMBASE (2)	Somma MED+EMB (3)	Sovrapp. MED/EMB (4)	Solo MED senza EMB (5)	Solo EMB senza MED (6)	PASCAL Biomed (7)
Gran Bretagna e Iralanda Nord	675	697	897	475	200	222	997
Francia	138	159	210	87	51	72	1791
Italia*	139	175	233	81	58	94	103
Germania	278	297	397	178	100	119	513

* le riviste in lingua italiana indicizzate sono 189 (MEDLINE 111 e EMBASE 137)

Analisi dettagliata della copertura periodici a livello mondiale per i tre database MEDLINE, EMBASE, PASCAL Biomed (con valutazione comparativa tra MEDLINE e EMBASE)

Ambito geografico	MEDLINE (1)	EMBASE (2)	Somma MED+EMB (3)	Sovrapp. MED/EMB (4)	Solo MED senza EMB (5)	Solo EMB senza MED (6)	PASCAL Biomed (7)
Europa	2.087	2.153	2.927	1.313	774	840	4.453
Stati Uniti	1.790	1.413	2.200	1.003	787	410	1.777
Canada	68	44	86	26	42	18	103
Cina	53	30	73	10	43	20	18
Giappone	224	221	328	117	107	104	40
Australia	51	47	74	24	27	23	142
Altri Paesi (Americhe, Africa, Oriente)	319	260	474	105	214	155	352
Totale	4.592	4.168	6.162	2.598	1.994	1.570	**6.885

** Le 6.885 riviste coperte da PASCAL Biomed, recuperate dal sito web dell'INIST si riferiscono a oltre 3.500 biomediche indicizzate *cover-to-cover*, e per le restanti si tratta delle riviste coperte da PASCAL, indicizzate in modo selettivo da PASCAL Biomed.

1) riviste indicizzate dal database MEDLINE: è composto dalla somma di (4) riviste comuni in sovrapposizione e (5) riviste indicizzate solamente da MEDLINE

2) riviste indicizzate dal database EMBASE: è composto dalla somma di (4) riviste comuni in sovrapposizione e (6) riviste indicizzate solamente da EMBASE

3) è il risultato di (1) e (2) combinati in .OR.: sono tutte le riviste indicizzate da entrambi i database MEDLINE e EMBASE nei loro insiemi separati, ma messi assieme. È anche la somma di (4) (5) (6)

4) è il risultato di (1) e (2) combinati in .AND.: sono le riviste indicizzate in comune dai database MEDLINE e EMBASE, dunque le sovrapposizioni

5) riviste indicizzate solo dal database MEDLINE e non da EMBASE

6) riviste indicizzate solo dal database EMBASE e non da MEDLINE

7) riviste indicizzate da PASCAL BIOMED (non effettuate ricerche incrociate con gli altri due database)

della Croazia⁴ e della Bosnia-Herzegovina⁵ effettuata da MEDLINE e quello relativo alla situazione della Repubblica Slovacca⁶ nel quale si riferisce che MEDLINE rappresenta favorevolmente a livello quantitativo la produzione biomedica del paese (nel presente lavoro ho rilevato sole tre riviste coperte per la Slovacchia).

Altro dato interessante è quello relativo all'indicizzazione da parte dei tre database di periodici che ho inserito sotto la voce "Altri Paesi" e comprendente Africa, Oriente, Americhe...

È subito evidente il basso numero di periodici indicizzati da ciascuno dei database. ➤

⁴ E.J. HUTH, *MEDLINE indexes the Croatian Medical Journal* [editoriale], "Croatian Medical Journal", 39 (1998), 1, p. 1-2.

⁵ I. MASIC, *Medicinski Arhiv-MEDLINE-Internet* [editoriale], "Medicinski Arhiv", 51 (1997), 3-4, p. 75-76.

⁶ Z. SIMOVICOVA, *The representation of Slovak medical journals in MEDLINE*, "Bratislava Lekarske Listy", 99 (1998), 5, p. 219-221.

Se andiamo a vedere la letteratura a riguardo, si possono citare due interessanti lavori relativi alla produzione biomedica dell'America Latina, entrambi significativi per due aspetti differenti. Un articolo di Rosselli⁷ dell'Unità epidemiologica clinica dell'Università di Bogotà che in qualche modo denuncia una scarsa rappresentazione da parte di MEDLINE delle pubblicazioni biomediche della Colombia, l'altro di Bunout e Reyes⁸ che attraverso MEDLINE rilevano dati numerici sulla produzione intellettuale dei ricercatori medici cileni. I due articoli nell'insieme sono interessanti e rappresentativi al contempo. L'articolo colombiano esordisce dichiarando che L'America Latina genera una bassa proporzione di riferimenti quotati in MEDLINE, lo studio voleva quindi mettere in luce quanta produzione scientifica "from and about Colombia in Medline" era stata effettivamente prodotta nel periodo 1987-1996. Dall'analisi condotta è emerso che negli ultimi dieci anni i periodici dell'America Latina indicizzati da MEDLINE si sono progressivamente ridotti nel numero e negli articoli contenuti, fino a giungere ad una completa esclusione dei periodici colombiani a partire dal 1991. In questi dieci anni la produzione delle istituzioni di ricerca biomediche della Colombia pubblicata in periodici stranieri è stata pari a 531 articoli, di cui il 41% proveniente da istituzioni universitarie. L'articolo riferisce di come la produzione locale sia bassa rispetto a paesi contigui come il Venezuela o il Cile, nonostante gli sforzi governativi per aumentarne i livelli.

L'articolo cileno prende in esame invece la produzione intellettuale biomedica nel periodo gennaio-agosto 1997 utilizzando il database MEDLINE per il recupero di 390 paper di autori residenti in Cile, attraverso il campo "AD" (indirizzo dell'autore). È da considerare però, a livello metodologico, che MEDLINE recupera da AD solo l'indirizzo del primo autore, e che dunque potrebbero essere stati esclusi dall'indagine eventuali articoli effettuati con autori cileni in collaborazione ai principali non cileni. Dei 390 individuati in relazione ad aspetti di "human medicin" solo 188, pari al 48% sono stati pubblicati su riviste cilene, mentre il resto è stato pubblicato in riviste straniere. Degli articoli pubblicati all'estero, 30 sono stati considerati di alta rilevanza per la "medicina clinica" e nove di questi autori sono stati invitati a "ritradurre" (dall'inglese allo spagnolo) gli abstract dei loro lavori che sono stati allegati allo studio.

CONCLUSIONI

Da studi precedentemente effettuati nel 1993,⁹ e successivamente nel 1997,¹⁰ erano emersi dati diversi relativamente all'area di sovrapposizione che era stata calcolata rispettivamente attorno al 30% e 34%, ma probabilmente la discordanza di risultato può ricondursi a differenti fattori. Il primo articolo si basava sugli anni fino al 1993 e la com-

parazione si riferiva più che alla copertura dei periodici alle differenze tra le versioni on line. Il secondo articolo, che in sostanza riassume il lavoro di Greenhalgh, si è basato su una comparazione effettuata non tanto a partire dalla copertura periodici di ciascun database, ma da referenze recuperate sulla base di strategie di ricerca. Quest'ultima metodologia si è limitata comunque alla valutazione delle sovrapposizioni di quanto recuperato e non di quanto effettivamente presente nei database e per di più si è ricondotta ad alcuni "casi" che riflettono "solo una parte" del contenuto: 21 query poste e non tutte le condizioni possibili.

La sovrapposizione calcolata sulla base dei riferimenti bibliografici, per produrre risultati corretti, andrebbe calcolata sull'effettivo numero totale di record presenti per ciascun database, ma è evidente come questa metodologia richieda l'utilizzo di strumenti adeguati.

L'individuazione dei titoli dei periodici coperti da ciascun database può dare un'idea forse più vicina alla realtà di quali siano le effettive sovrapposizioni.

Il limite di questa scelta sta comunque nel fatto che non si è mai sicuri di "come" quel database indicizza quel determinato periodico, se in modo completo (*cover-to-cover*), tutti gli articoli contenuti nei fascicoli, oppure se procede in modo selettivo, con scelta degli articoli rilevanti.

La metodologia utilizzata per il presente lavoro è stata finalizzata all'individuazione di aree di sovrapposizione in termini percentuali e di interpretazione di alcune evidenze ai fini di ulteriori sviluppi per lavori comparativi con altri database. L'analisi dei dati disaggregati, per aree geografiche, può servire a sfatare alcuni luoghi comuni, laddove vengono indicati certi database come prevalentemente a "carattere europeo", quando invece si riferiscono ad un contesto geografico generale, non necessariamente sbilanciato verso l'Europa, oppure orientati verso un singolo paese dell'Europa che però non riguarda il nostro paese in particolare.

È evidente che la sovrapposizione per l'ambito europeo tra MEDLINE e EMBASE è piuttosto rilevante (44,8%) e che l'area scoperta, risulta ben equilibrata tra i due database (26,5% e 28,7%).

PAESE	Sovrapp. MED/EMB %	Solo MED senza EMB %	Solo EMB senza MED %
Gran Bretagna e Irlanda Nord	52,9	22,2	24,8
Francia	41,4	24,3	34,3
Italia	34,7	24,9	40,4
Germania	44,9	25,1	30,0

⁷ D. ROSELLI, *Latin American biomedical publication: the caso for Colombia in Medline*, "Medical Education", 32 (1998), 3, p. 274-277.
⁸ S. BUNOUT - H. REYES, *Biomedical papers written by Chilean authors published in international journals in 1997. A review of MEDLINE*, "Revista Medica de Chile", 126 (1998), 6, p. 677-678.
⁹ J.P. ROVERS - J.E. JANOSIK - P.F. SOUNEY, *Crossover comparison of drug information online database vendors: Dialog and MEDLARS*, "Ann. Pharmacoth", (1993), 27, p. 634-9.
¹⁰ D. WOODS - K. TREWHELLAR, *Medline and Embase complement each other in literature searches*, "Letters. British Medical Journal", 11 April 1998, 316, p. 1.166.

La sovrapposizione è ulteriormente accentuata per le pubblicazioni anglossassoni che arriva al 52,9%. Francia e Germania si mantengono all'interno dei valori della media europea, mentre l'Italia risulta coperta meglio da EMBASE (40,4%) rispetto al solo MEDLINE (24,9%).

Questi dati però fanno riflettere, non tanto in relazione all'area di sovrapposizione più o meno ampia, ma in relazione alla quantità di periodici coperti dai due database in relazione all'effettiva produzione intellettuale dei Paesi europei. E questo è un discorso che vale non solo per la letteratura biomedica, ma anche per la letteratura tecnico-scientifica in generale.

Per la presente ricerca ho voluto rilevare anche la copertura periodici del terzo database bibliografico, PASCAL Biomed, e questo terzo strumento mi ha permesso di comparare e valutare alcuni dati di rilevanza fondamentale.

Se vediamo i dati di PASCAL Biomed in relazione ai paesi coperti, salta subito all'occhio come la Francia sia il Paese meglio rappresentato rispetto agli altri, e questo è ovvio in quanto si tratta di un database francese, ex Bulletin Signaletique.

Ma se andiamo a vedere quanti sono i periodici coperti relativamente alla Francia da MEDLINE ed EMBASE (che viene dichiarata da alcuni distributori come database a carattere europeo) si notano solo 210 periodici per entrambi i database su una produzione presunta di oltre 1.791 titoli (coperti da PASCAL), nemmeno il 12%. È vero però che, come specificato precedentemente, le quasi 7.000 riviste coperte da PASCAL Biomed si riferiscono ad una indicizzazione *cover-to-cover* per circa la metà, in quanto le restanti sono coperte solo per articoli selezionati, possiamo presumere un dato percentuale del 24% di copertura per l'area francese.

Se trasponiamo ipoteticamente il modello francese anche all'Italia, che raggiunge una copertura di 233 periodici per entrambi i database e di solo 103 per il database europeo PASCAL Biomed, possiamo intuire quanto vasta possa essere la "zona scoperta" e quanto potrebbe essere interessante lavorare su questo fronte.

Quanti sono i periodici biomedici italiani non indicizzati dai tre database?

Si tratta di materiale intellettualmente di valore oppure non sono stati presi in considerazione appunto perché non rilevanti ai fini di una immissione in un database?

È certo che andrebbe effettuata un'analisi più seria e ap-

profondita sul "parco macchine periodici" di produzione nazionale (non solo per il settore biomedico).

In totale la sovrapposizione per la copertura dei periodici tra i due database MEDLINE e EMBASE raggiunge il 42,2%. Insieme coprono una parte rilevante di letteratura biomedica internazionale.

Ambito geografico	Sovrapp. MED/EMB %	Solo MED senza EMB %	Solo EMB senza MED %
Europa	44,8	26,5	28,7
Stati Uniti	45,5	35,8	18,7
Canada	30,2	48,9	20,9
Cina	13,6	59,0	27,4
Giappone	35,6	32,7	31,7
Australia	32,4	36,4	31,2
Altri PAESI (Americhe, Africa, Oriente)	22,1	45,2	32,7
Totale	42,2	32,3	25,5

Si può affermare quindi che EMBASE è complementare a MEDLINE e che PASCAL Biomed copre una buona parte della letteratura europea con particolare riguardo a Francia, ma anche Gran Bretagna e Germania, mentre la letteratura medica italiana non è sufficientemente rappresentata da nessuno dei database presi in esame.

Le potenzialità di EMBASE non risultano essere sfruttate appieno sia in quanto non sufficientemente conosciute dai ricercatori in generale sia perché è uno strumento non molto diffuso. La scarsa diffusione potrebbe essere una delle ragioni dell'alto costo di questo database, rispetto ad altri: più la base di acquisto si allarga e più i costi si abbassano. Una comparazione più accurata dei prodotti che presentano i database di settore biomedico, deve tener conto oltre che degli aspetti inerenti alla struttura del database, dei record, della presenza o meno di tesaurus, anche a parità di condizioni (stesso supporto, stessa interfaccia e stesso software di recupero dell'informazione), degli aspetti connessi alla copertura temporale. Le differenze che si possono desumere da una simile comparazione possono risultare significative anche per un più corretto utilizzo da parte dell'utenza, soprattutto specialistica. ■