

---

# BOLLETTINO

# UNIONE MATEMATICA ITALIANA

*Sezione A – La Matematica nella Società e nella Cultura*

---

RENATO SPIGLER

## Ricerca scientifica e Internet, oggi

*Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. 10-A—La Matematica nella Società e nella Cultura (2007), n.1, p. 137–153.*

Unione Matematica Italiana

[http://www.bdim.eu/item?id=BUMI\\_2007\\_8\\_10A\\_1\\_137\\_0](http://www.bdim.eu/item?id=BUMI_2007_8_10A_1_137_0)

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

---

*Articolo digitalizzato nel quadro del programma  
bdim (Biblioteca Digitale Italiana di Matematica)  
SIMAI & UMI*

<http://www.bdim.eu/>



## Ricerca scientifica e Internet, oggi

RENATO SPIGLER

### 1. – Generalità.

Nel 1945, Vannevar Bush, l'uomo che in qualità di Direttore dell'Office of Scientific Research and Development era responsabile del coordinamento delle attività di 6000 scienziati americani coinvolti nello sforzo bellico durante la seconda guerra mondiale, scrisse un celebre articolo intitolato «As we may think» [5], destinato ad essere molto influente. Bush fu uno di quelli che raccomandò la creazione dell'organizzazione che sarebbe diventata l'NSF, la National Science Foundation. Nel suo articolo Bush osservò che nel corso di molti anni, le invenzioni umane avevano esteso le capacità *fisiche* piuttosto che quelle *intellettuali* dell'uomo, ma nel momento in cui egli scriveva erano già disponibili strumenti capaci di consentire l'accesso alla conoscenza acquisita nel corso delle epoche passate, nonché di manipolarla.

Bush si rendeva conto della possibilità e della necessità di registrare tutte le informazioni raccolte dagli scienziati giorno dopo giorno, ma, date le tecnologie del suo tempo, egli pensò di ricorrere massicciamente ai microfilm per immagazzinare i dati. In ogni caso, l'idea fondamentale era quella di una macchina che potesse aiutare le attività cognitive dell'uomo, la cosiddetta «memex» (v. sotto).

Alcuni, quali Doug Engelbart [13], l'uomo che, tra le altre cose, ha inventato il mouse dei computer, e Ted Nelson [27], trassero grande ispirazione dall'articolo di Bush. Engelbart mise a punto, insieme ai suoi colleghi William K. English e John F. Rulifson la prima implementazione al mondo di quello che sarebbe stato chiamato «ipertesto». Ciò fu fatto per favorire la collaborazione tra gruppi di lavoro geograficamente distanti, al fine di aumentare le capacità umane tra-

mite l'uso di computer. Per la verità qui andrebbe ricordato anche il bibliografo belga Paul Otlet, insieme a Henri La Fontaine, per aver introdotto l'idea di ipertesto e quella stessa di Internet (anche se in un'era pre-computer), nel 1895 [28]. A Bruxelles venne utilizzato un intero palazzo, denominato «Mundaneum», con la funzione di biblioteca mondiale, rappresentante una specie di Internet di articoli e libri, dato che la conoscenza poteva essere ricostruita da ciascun individuo per mezzo di una quantità enorme di documenti schedati. Il Mundaneum offriva un accesso automatizzato ai documenti e consentiva di mettere in relazione certi documenti con altri. Di fatto, Paul Otlet può essere considerato il fondatore di quella che si potrebbe chiamare la scienza della documentazione.

Sin dal 1930, Bush aveva introdotto l'idea di «memex», un concetto che è stato considerato come pioniere del Web, o World Wide Web. Memex era un dispositivo basato su microfilm, essenzialmente una scrivania ove un individuo poteva immagazzinare tutti i suoi libri, note, appunti, registrazioni. Questo dispositivo si prestava a ritrovare in modo rapido e flessibile tutte le informazioni immagazzinate. Si tratta di un'idea storicamente importante perché ha ispirato lo sviluppo delle tecnologie ipertesto. Una versione abbreviata di «As we may think», pubblicata da *Life* nel Novembre 1945, ispirò infatti Doug Engelbart e Ted Nelson, i quali – come accennato sopra –, svilupparono (indipendentemente l'uno dall'altro) l'idea degli ipertesti [35].

La necessità di raccogliere, conservare, ed elaborare una grandissima quantità di informazioni, addirittura – in linea di principio – tutto quello che uno scienziato può osservare e misurare quotidianamente nel suo laboratorio, fu una delle motivazioni che portarono a sviluppare Internet, in vista della quantità di dati raccolti fino a un certo momento, i quali crescono in modo esponenziale. Un progetto proposto recentemente da Robert Mayo presso la Compaq e denominato FACTOID, è nello spirito di memex, ma basato sulle tecnologie disponibili al giorno d'oggi [15]. In ogni caso, uno dei problemi attuali, come già lo era ai tempi di Bush, è quello di effettuare una compressione dell'enorme massa di dati, una questione che Bush stesso riteneva importante.

È un fenomeno relativamente nuovo, risalente forse alla metà del

secolo scorso, che i dati scientifici e le nuove teorie sono stati prodotti in tale quantità che risulta difficile per un singolo studioso conoscerli tutti. L'inevitabile tendenza verso la specializzazione rende impossibile che un singolo esperto, per quanto colto e dotato di talento (e di memoria), sappia (e ricordi) tutto, pur limitatamente al suo campo. È chiaro d'altra parte che registrare tutta l'informazione (significativa) è essenziale.

Noi abbiamo bisogno non solo di registrare l'informazione, ma anche di conservarla, elaborarla (cioè modificarla, ritrovarla, stamparla, copiarla), fare ricerche su di essa, e aggiornarla. Tutti questi sono aspetti senza i quali oggi si potrebbe fare ben poco – o molto di meno di quel che si fa – nel campo della ricerca scientifica, sia sperimentale che teorica. Le scienze economiche, sociali e umanistiche sono interessate a questo non meno delle scienze naturali e della matematica.

Il concetto stesso di «informazione» è molto ampio, anche rispetto alla sua natura: non si tratta solo di testi e di formule, ma anche di grafici (a colori, stereoscopici), di disegni e di immagini fotografiche, di suoni (discorsi e musica), di animazioni e di film.

Bush concepì la possibilità che un futuro scienziato potesse muoversi per il suo laboratorio o «sul campo», e registrasse ogni cosa che valesse la pena di essere registrata, ad esempio nella forma di immagini fotografiche [5]. Quest'idea è stata riproposta da Robert N. Mayo, che ha immaginato lo stesso scenario, arricchito però dai nuovi e più efficienti strumenti offerti dalle nuove tecnologie [15].

## 2. – Cosa vorrebbe un autore dal Web?

Cosa vorrebbe, cosa si aspetta un autore dal Web, da Internet, riguardo all'informazione scientifica? Uno scienziato vorrebbe poter accedere a testi completi, presi da un'ampia raccolta di periodici. Un accesso, possibilmente gratuito e *rapido* a tale materiale è importante come pure lo è la possibilità di «scaricarlo» quando necessario. È anche essenziale disporre di sistemi di ricerca sì rapidi ma soprattutto intelligenti. Ed è anche importante disporre di archivi che elenchino i volumi, i libri, come pure di collegamenti [link] ad un gran numero di biblioteche.

L'accesso ad un gran numero di testi di articoli completi, pubblicati su di una grande varietà di periodici, è importante al giorno d'oggi anche per la Matematica, dato che alcuni dei migliori matematici applicati pubblicano i loro lavori (anche) su riviste di fisica e/o orientate verso le altre scienze o l'ingegneria. Se infatti il matematico intende contribuire allo sviluppo di una data scienza applicata, a risolverne i problemi specifici fornendo risultati utili ad essa, è necessario che il suo contributo sia comunicato ai potenziali utenti sui *loro* giornali, piuttosto che su periodici matematici tradizionali. Questo naturalmente richiede uno sforzo aggiuntivo da parte di questi autori, perché ogni disciplina (e quindi ogni periodico specializzato che la tratta) ha un suo linguaggio e un suo stile. Inutile dire che vi sono regole anche molto diverse da quelle a cui i matematici sono abituati quando sottopongono i loro lavori ai loro giornali. In ogni caso, chi opera in aree di ricerca in rapida evoluzione e/o a carattere interdisciplinare ha bisogno di accedere tempestivamente ad un grande numero di riviste. Questa necessità rappresenta un problema anche per via delle risorse finanziarie disponibili, spesso limitate, dato che gli Editori commerciali intendono trarre profitto dall'impresa scientifica (dei costi comunque ci sono). Ne segue che i vantaggi offerti dall'attraente prospettiva rappresentata dai cosiddetti «Open Archives» dovrebbe essere considerata molto seriamente [31], ma soprattutto quando questi consentono non soltanto di acquisire dei metadati, cioè i dati di riferimento, ma quando permettono un effettivo libero accesso alla letteratura, posta nei cosiddetti Open Access repositories.

Servono dei motori di ricerca (dei «browser») efficienti, cioè rapidi e intelligenti. Il loro valore è inestimabile. Si pensi ad esempio a Google. Google è un sistema «intelligente» ed è in grado di effettuare ricerche sui quasi dieci di miliardi di pagine esistenti attualmente sul Web. Alcuni sistemi di ricerca adottati da compagnie telefoniche, come pure quello della ISI Web of Knowledge (quella dell'Impact Factor), persino nel caso di basi di dati di molte associazioni scientifiche nonché di molte biblioteche scientifiche, non sembra lo sia. E tuttavia è stato stimato che persino Google è in grado di localizzare non più del 30% di ciò che si cerca.

Riguardo alla rapidità di accesso, connessione, e «download» di

materiale, parte del problema – almeno in certi Paesi – è dovuto al modo di operare, alla bassa qualità ed efficienza delle compagnie telefoniche. Non sempre (o non ancora) vengono utilizzate connessioni in fibra ottica, che permettono di trasmettere a larga banda, con grande velocità. Pure la qualità dei server non è sempre ai livelli più elevati.

Un'altra cosa che gli autori vorrebbero vedere realizzata nel mondo dell'editoria scientifica è il poter seguire periodicamente lo stato dei loro lavori, dopo che sono stati inviati per la pubblicazione. Ogni periodico dovrebbe predisporre immediatamente una pagina Web, organizzata in modo efficiente, per ogni singolo articolo ad esso sottoposto (non tutti i periodici lo fanno). L'efficienza si riferisce alla cura con cui tale pagina dovrebbe essere frequentemente aggiornata, diciamo ogni una o due settimane (alcuni non aggiornano affatto tale pagina). Alcuni periodici, anche facenti capo ad associazioni scientifiche di primo piano, e – si badi bene – anche in casi in cui i lavori sono di interesse applicativo concreto e quindi di interesse immediato, non corrispondono a standard adeguati. Il modello da imitare qui è quello offerto dall'American Physical Society (APS) [2].

### **3. – La valutazione della ricerca scientifica.**

La necessità di suddividere risorse spesso limitate (in termini di posti di lavoro e di finanziamenti) e di selezionare comunque i migliori studiosi di ogni disciplina richiede di valutare la qualità del prodotto delle indagini svolte dai vari ricercatori.

#### *3.1 – Ancora sull'Impact Factor.*

Nella comunità scientifica si è discusso molte volte sull'opportunità o meno di utilizzare il cosiddetto «fattore d'impatto» (Impact Factor, o brevemente IF); si veda ad esempio [16, 17], dove il ruolo dell'IF nella valutazione della qualità della ricerca scientifica, soprattutto nel campo della Matematica è stato ampiamente analizzato. Più di recente, a testimonianza dell'importanza della questione del valore o meno dell'IF nella valutazione delle ricerche matematiche, è intervenuto John Ewing, executive director dell'American Mathematical Society

(AMS) [14]. Ewing ricorda che l'IF è stato creato da Eugene Garfield verso la fine degli anni '50 come un modo di misurare il valore di un dato periodico, calcolandone il numero medio di citazioni per articolo, riportate in un dato arco di tempo. L'Institute for Scientific Information (ISI) pubblica il Science Citation Index, che cataloga varie migliaia di riviste scientifiche (ma molto meno della metà delle riviste matematiche esistenti). L'IF attuale della rivista  $X$  è calcolato come rapporto  $A/B$ , dove  $A$  è il numero di citazioni fatte negli articoli del database dell'ISI durante il 2006 agli articoli pubblicati in  $X$  nel biennio precedente, cioè nel 2004-2005, mentre  $B$  è il numero totale di articoli pubblicati in  $X$  nel 2004-2005. Rimandiamo all'articolo di Ewing per ulteriori dettagli, segnalando che l'AMS ha creato ora un suo database della letteratura matematica, che tiene conto di 325 riviste matematiche pubblicate a partire dal 2000. Si calcola così il Mathematical Citation Quotient (MCQ), che come l'IF è il numero medio di citazioni ad una rivista su di un dato arco di tempo, ma qui tale intervallo è esteso a cinque anni [24].

Più che togliere all'IF ogni utilità, Ewing osserva che se ne è fatto un uso errato, dato che – ammesso che il valore di un articolo si misuri davvero col numero di citazioni che se ne fanno – l'IF non misura affatto il valore di un dato articolo, e neppure il valore della rivista. Meno ancora quello di un dato ricercatore.

La grande mole di dati che l'ISI o lo stesso Web raccolgono potrebbe forse essere utilizzata elaborando questi dati in modo più accurato, più intelligente. Ad esempio, Sydney Redner e Pu Chen [7] suggeriscono di adottare l'algoritmo di Google, denominato PageRank, al posto di utilizzare l'IF dell'ISI. PageRank è in grado di attribuire un peso alle citazioni fatte. Dai primi esperimenti sembra che in tal modo si riesca a identificare i lavori che, se pur poco citati (quindi con un basso IF), sono invece molto importanti.

### 3.2 – *L'indice di Hirsch.*

Un contributo molto recente è apparso sui prestigiosi «Proceedings of the National Academy of Sciences» americana [21]. J.E. Hirsch propone di quantificare i risultati della ricerca scientifica di un dato

individuo mediante l'indice  $h$  pari al numero di articoli aventi almeno  $h$  citazioni ciascuno. Ovvero, uno studioso ha un indice di Hirsch pari ad  $h$  se  $h$  dei suoi  $N$  articoli hanno almeno  $h$  citazioni ciascuno, mentre i rimanenti  $N - h$  hanno ciascuno non più di  $h$  citazioni. Il calcolo di  $h$  si può fare facilmente ordinando le citazioni ottenibili da Web of Science (nella ISI Web of Knowledge), ordinando gli articoli per «times cited». Un tale indice è meno ingenuo di quello che potrebbe sembrare a prima vista. Si osservi che avere pochi articoli pur con un numero arbitrariamente grande di citazioni (eccellenza in pochi contributi ma bassa produttività), come pure averne moltissimi con poche citazioni ciascuno (grande produttività ma non grande valore di alcun singolo contributo), porta a valori modesti di  $h$ . Non c'è dubbio che un grafico, un istogramma, che riporti in ascissa il numero progressivo  $n$  degli articoli scritti e in ordinata il numero  $N_c$  di citazioni corrispondente darebbe maggiori informazioni sia dell'IF che dell'indice di Hirsch. Se nell'istogramma precedente si ordinano gli articoli per numero di citazioni decrescente, l'indice  $h$  si vede subito considerando la bisettrice del I quadrante del piano  $(n, N_c)$ , mentre l'area dell'istogramma (pari alla somma delle ordinate,  $\sum_n N_c(n)$ , fornisce il numero totale delle citazioni che un dato autore ha avuto in tutti i suoi articoli (ovviamente se presenti nel database dell'ISI).

Così, nell'ambito della Fisica Hirsch ha calcolato che E. Witten ha l'indice  $h = 110$ , il più alto di tutti, ossia Witten ha scritto (tra gli altri) 110 articoli ciascuno dei quali è stato citato almeno 110 volte; S. Hawking ha un indice di Hirsch  $h = 62$ . Un simile calcolo effettuato per la Biologia e le Scienze Biomediche porta – come ci si potrebbe aspettare – a indici sensibilmente più elevati (il massimo, da un esame del periodo 1983-2002, risulta essere l'indice  $h = 191$  di S.H. Snyder).

Inutile dire che per i matematici, per quanto applicati, la valutazione dell'indice di Hirsch sembra ben più discutibile (e penalizzante) ed il valore di  $h$  ben più modesto. Ad esempio, due matematici certamente molto influenti come Louis Nirenberg e Peter D. Lax avrebbero rispettivamente  $h = 39$  e  $h = 33$ . Oltretutto questi valori si ricavano dal database dell'ISI, ragion per cui molti contributi apparsi su riviste con titolo o in lingua italiana (per esempio Annali della Scuola Normale Superiore di Pisa, Annali di Matematica Pura e Applicata, pub-

blicazioni dell'Accademia Nazionale dei Lincei), o comunque in periodici per qualche motivo non recensiti dall'ISI (per esempio *Differential and Integral Equations*, *Advances in Differential Equations*, per via dei difficili e discontinui contatti con la casa editrice Khayyam Publishing), non entrano nel computo. Come per l'IF, l'indice di Hirsch si dovrebbe quantomeno normalizzare in ciascuna disciplina, anche se questo non basta. Per la matematica c'è anche il limite del (basso) numero di riviste complessivamente considerate dall'ISI.

Valutare l'importanza della produzione scientifica di un dato studioso rimane una questione delicata, e non è neppure detto che il numero di citazioni misuri davvero il valore di un articolo scientifico, tanto meno quello di un dato ricercatore, ma forse vale la pena continuare a cercare di identificare dei parametri utili a tale scopo, e che siano al tempo stesso facili da calcolare, come lo è l'indice di Hirsch.

#### **4. – Controllo di qualità e certificazione negli Open Access e negli Open Archives.**

Per quanto concerne la letteratura scientifica, si dovrebbe incoraggiare la propensione a servirsi di Open Access e di open archives o di «repositories» istituzionali, a causa delle regole spesso insoddisfacenti riguardanti i diritti d'autore che vengono imposte agli autori dagli Editori commerciali. In concomitanza con tali scelte tuttavia si dovrebbero considerare seriamente il controllo della qualità e la conseguente certificazione (QC/C) del materiale reso così accessibile, cfr. [20]. Questo è infatti un punto molto delicato in certe comunità, quale ad esempio quella dei matematici, e corrisponde al processo di una revisione adeguata dei lavori scientifici che vengono resi pubblici (il cosiddetto processo di «peer reviewing»), cfr. [31]. Il altri campi, come la Medicina, un severo controllo dei risultati che appaiono nella letteratura viene percepito come estremamente importante: basti pensare al caso in cui l'esito di una sperimentazione clinica riguardi la salute se non la stessa sopravvivenza di esseri umani, i pazienti.

Si potrebbe suggerire che ogni singolo articolo depositato in un open archive, come pure disponibile in un «Open Access repository», sia accompagnato da una dichiarazione quale: «Questo articolo è stato

letto e approvato da (*nome e cognome di uno o più esperti*)». Anche se questo o questi garanti potrebbero essere il compiacente relatore di uno studente di Dottorato o un amichevole Collega dell'autore, una simile dichiarazione fornirebbe comunque una sorta di certificazione, del tipo di quelle fornite dai referee delle riviste specializzate. In tale modo l'articolo circolerebbe validato da alcune persone che si prenderebbero la responsabilità della sua correttezza. Più famosi e stimati fossero tali garanti, più affidabile sarebbe considerato il lavoro in oggetto. Dopo tutto, questo è quello che accade quando un articolo è giudicato di rango elevato perché è stato pubblicato da un periodico di grande prestigio, il che infatti corrisponde ad aver superato il vaglio di Editor e referee di sicuro valore. E questo è anche quello che accadeva nelle pubblicazioni delle vecchie Accademie, quando i lavori venivano presentati da uno o più membri dell'Accademia.

## 5. – Ripulire il Web e identificare i documenti immessi.

Un altro problema di carattere generale è rappresentato dalla talvolta indesiderata proliferazione di pagine nel Web. La necessità di rimuovere dalla rete ogni sorta di «immondizia» potrebbe diventare sempre più importante. I matematici conoscono bene l'importanza di *non* conservare tutto. Stan Ulam racconta che nel celebre «Scottish Café» di Lwów alcuni matematici, tra cui Banach, Mazur ed egli stesso, si trovavano per discutere di matematica e là «i tavolini avevano la superficie superiore di marmo bianco su cui si poteva scrivere con la matita e, *cosa ancor più importante, da cui si potevano facilmente cancellare le annotazioni precedentemente scritte*» [34, p. 67].

Come si è ricordato sopra, vi sono attualmente vari miliardi di pagine in rete e il loro numero cresce esponenzialmente. Ripulire il Web, cancellando le pagine obsolete o contenenti informazioni rivelatesi errate, potrebbe diventare tanto importante quanto depositarvi nuovo materiale. Nessuno sembra preoccuparsi di rimuovere pagine che riguardano vecchi convegni o versioni preliminari di svariati documenti. Quest'opera di pulizia dovrebbe essere fatta al fine di rendere la rete più efficiente, cioè più veloce. È una questione di civiltà, di educazione, come pulire le strade dalle immondizie. Sarebbe anche opportuno che

ogni documento depositato sul Web contenesse delle «coordinate» quali URL e data in cui il documento è stato preparato e messo online: non sempre queste informazioni sono reperibili sul testo. Da pochi anni è stato introdotto l'uso del DOI (Digital Object Identifier system) [10]. Il DOI, come è ben spiegato in [25], è uno standard costituito da una stringa univoca di caratteri alfanumerici, che permette di identificare permanentemente, all'interno di una data rete digitale, un qualsiasi oggetto di proprietà intellettuale, e di associarvi i relativi dati di riferimento, i metadati. Attraverso questi ultimi, il DOI non solo si presta a identificare un documento in forma digitale, ma rappresenta anche un potente e inequivoco strumento per la conservazione e la condivisione di documenti. È superfluo sottolineare l'importanza di un tale strumento per la letteratura scientifica, ma si deve osservare che il DOI relativo a un dato documento viene assegnato da alcune agenzie autorizzate, le quali richiedono un pagamento annuale. Giunti nel dominio del profitto, vi sono casi in cui si tende a spingere molto la cosiddetta «granularità», cioè la frammentazione del documento in varie parti più piccole al fine di aumentare i guadagni.

Oggi giorno infine si osserva con frequenza crescente il fenomeno denominato «SPAM», cioè quello di ricevere messaggi di posta elettronica indesiderati, non sollecitati. È stato osservato che messaggi di questo tipo rappresentano fino ai due terzi di quelli che si ricevono quotidianamente (in realtà spesso sono molti ma molti di più). Si tratta per lo più di annunci commerciali pubblicitari, riguardanti software e hardware, medicinali, (presunte) lotterie e pornografia. Alcuni messaggi trasmettono virus informatici (dai quali si è sempre protetti se si usa Linux, ma non se si utilizza Windows), altri sono annunci mirati a un uditorio specifico, come quelli che propongono libri e abbonamenti a riviste scientifiche. Questi ultimi hanno origine dagli elenchi dei partecipanti a convegni specializzati, o da quelli di Dipartimenti universitari, che sono liberamente accessibili sul Web. Sistemi di protezione quali «firewall», filtri, programmi antivirus, nonché leggi adeguate, possono difenderci, almeno in parte, da tutti questi fastidi. Anche se riusciamo a filtrare molti SPAM però, quando la posta si accumula per giorni o settimane, si deve sprecare molto tempo per un pur rapido controllo e quindi per cancellare la posta indesiderata.

## 6. – «Open Access» e «open archives repositories».

Molti studiosi, in particolare molti matematici, sono ancora riluttanti ad utilizzare, e ancor più a privilegiare, gli «open archives». Alcuni sono contrari allo sfruttamento stesso dei sistemi di circolazione e di pubblicazione elettronici rispetto ai sistemi più tradizionali di stampa su di un supporto cartaceo. Eppure nel corso dei secoli intere biblioteche sono state distrutte da catastrofi naturali o provocate dall'uomo, inondazioni e incendi. Una mancanza improvvisa di elettricità o l'esposizione a forti campi magnetici o termici non sono eventi più probabili dei precedenti, e di regola si prendono delle precauzioni piuttosto efficaci al fine di proteggere i dati immagazzinati su supporto elettronico. Per ciò che riguarda l'hardware, ci sono dispositivi atti a proteggere i sistemi da picchi o cadute improvvise di potenza, il che potrebbe comportare la perdita di dati, e vengono effettuati frequenti salvataggi, formando copie (backup), per difendersi dalla distruzione di dati in formato elettronico.

Si devono piuttosto considerare i vantaggi che derivano dal disporre di materiale scientifico in formato elettronico, perché così facendo si può non solo conservare l'informazione, ma anche ritrovare dati e stamparli ogni volta che si voglia farlo, ed effettuare ricerche bibliografiche anche incrociate sulla base di informazioni parziali, quali il solo nome di un autore, alcune parole chiave, il nome della rivista, l'anno di pubblicazione. Chiaramente quest'impresa sarebbe ben più ardua se fosse condotta manualmente tra gli scaffali di una Biblioteca.

Sfruttare maggiormente i sistemi di pubblicazione e di raccolta elettronici può rendere meno serio il problema della cosiddetta «letteratura grigia», cioè di quell'insieme di documenti non facilmente reperibili, in quanto pubblicati nella forma di atti di convegni o su riviste oscure, su periodici a carattere nazionale o comunque soggetti ad una diffusione limitata. Questo materiale, a volte importante, rischia di restare virtualmente sconosciuto ai più.

Qual è lo «stato dell'arte» degli Open Access e degli open archives, ovvero degli archivi (repositories) istituzionali, riguardo alla ricettività dei potenziali utenti? Assai di recente Elsevier ha reso disponibile alla ricerca specifica di materiale scientifico un motore di ricerca deno-

minato «Scirus» [30]. Questo sistema, il cui nome preciso è «Scirus Repository Search», costituisce un nuovo servizio, concepito per supportare archivi istituzionali. Il primo di questi si trova presso l'Università di Toronto, ed è la collezione denominata «T-space». L'Università di Toronto è forse la migliore e più grande Università canadese, e possiede una delle maggiori Biblioteche del Nord America. Un accordo simile con Google (Google Scholar), aveva preceduto di molti mesi il sistema Scirus. Elsevier ha anche il sistema «Science Direct» [12] (accessibile a pagamento), e pure Springer ha predisposto un potente sistema, «SpringerLink» [32] (pure a pagamento), mentre la Society for Industrial and Applied Mathematics (SIAM) americana, ad esempio, consente di scaricare gratuitamente gli articoli pubblicati nelle sue riviste, tramite JSTOR [23], ma solo dopo alcuni anni dalla pubblicazione. L'APS invece ha messo a disposizione in rete, tramite il suo Physical Review Online Archive (PROLA) [29] *gratuitamente tutti* gli articoli apparsi sulle sue riviste (ad esempio tutte le Physical Reviews) sin dal 1893. Gli articoli più vecchi, che non si trovavano in formato elettronico, sono stati scannerizzati dalla forma cartacea.

Ricordiamo anche il celebre sito denominato arXiv [3], dove i fisici possono depositare gratuitamente i loro preprint. Certo che se ciò viene fatto all'atto della sottomissione ad un periodico, non sono ancora stati vagliati da alcun referee. Esiste anche un ArXiv gestito dalla UC Davis che riguarda gli articoli matematici contenuti in arXiv (che attualmente è gestito dalla Cornell University) [4]. Questo materiale è accessibile a tutti.

Secondo una recente indagine, condotta dalla Key Perspective Ltd per conto del Joint Information Systems Committee in Gran Bretagna, oltre l'80% dei ricercatori accademici di tutto il mondo accetterebbe di depositare copie dei lavori in un qualche archivio aperto all'accesso da parte di chiunque [33]. Sembra che soltanto il non sapere neppure che una tale possibilità esiste, a parte la difficoltà tecnica incontrata da alcuni, siano le cause principali che hanno limitato fino ad ora il numero dei potenziali utenti. Questi risultati sembrano però troppo ottimistici, dato che il campione consultato (di 1296 persone) è davvero molto piccolo rispetto al numero di ricercatori attualmente attivi nel mondo intero.

## 7. – Wikipedia.

Merita una menzione a parte il fenomeno della Wikipedia, l'enciclopedia aperta [9, 35]. Il termine significherebbe rapido, veloce, in lingua hawaiana, e difatti cercando la traduzione da tale lingua in inglese, il dizionario online «foreignword.com» [17] risponde che «awiki wikiwiki» si traduce con quick, fast. Wikipedia è una enciclopedia multilingue, disponibile gratuitamente online, aperta alla collaborazione di tutti. Chiunque può contribuirvi liberamente (anche se c'è un qualche controllo) e il pericolo di inserire errori sembra essere sempre corretto da ulteriori interventi. Una statistica condotta da *Nature* avrebbe mostrato che il numero di errori presenti in Wikipedia supera solo di poco quello che affligge la prestigiosa Enciclopedia Britannica [26]. In realtà, l'Enciclopedia Britannica ha contestato la validità sia del metodo che dei risultati di questa indagine, per cui non è affatto certo che gli errori incontrati su Wikipedia siano solo poco più numerosi di quelli rilevabili sull'Enciclopedia Britannica. Non c'è dubbio che molti, soprattutto studiosi dell'ambito accademico, siano alquanto sospettosi riguardo all'affidabilità di Wikipedia, perché la correttezza delle informazioni che questa fornisce sono frutto di una collettività che non è affatto garantito sia onnisciente, come sembrerebbe credere chi ritiene che le varie voci di Wikipedia vengano rapidamente corrette. Attualmente è in corso un acceso dibattito non solo sull'affidabilità ma anche sui diversi scopi che Wikipedia avrebbe. Uno dei fondatori di questa, Larry Sanger, sta organizzando una versione più certificata di quanto lo sia Wikipedia, denominata «Citizendium» (da Citizens' Compendium), che avrà maggiori controlli di Wikipedia e farà ricorso a esperti [6], di fatto un misto di enciclopedia aperta e di una tradizionale con i contenuti certificati.

## 8. – I giornali online.

Accanto all'idea degli open access, si dovrebbero considerare anche i giornali puramente elettronici. L'aspetto economico della stampa scientifica non costituisce un elemento trascurabile, anche se i costi tipici relativi ad un periodico prodotto da una casa editrice commer-

ciali sono più elevati, senza contare la questione del copyright. Un prezzo lo si deve pagare comunque. Merita di essere citata l'iniziativa che ha portato a fondare il «Journal of Pancreas» (JOP), il primo giornale di pancreatologia puramente elettronico [22]. Questo giornale è stato fondato nel 2000 ed è pubblicato ad opera di un pugno di persone, che vi lavorano gratuitamente e con abnegazione per assicurare al giornale un rapido e serio sistema di valutazione dei lavori che vi pervengono e la susseguente rapida pubblicazione, ovviamente elettronica, dei lavori accettati. Sarebbe necessario disporre di un supporto finanziario per pagare (almeno) una persona a tempo pieno e un paio di collaboratori a tempo parziale, a livello di segretarie e di personale tecnico. Il costo sarebbe relativamente modesto, ma non nullo, e qualcuno deve pagarlo. Chi deve farlo? Gli autori? Forse. È una possibilità da considerare. Finanziatori quali le case farmaceutiche vogliono avere una ricaduta, e questo è un punto assai delicato se JOP deve contribuire alla guarigione di malati al pancreas senza per questo generare dei profitti. Un'altra possibilità è quella di passare a Editori commerciali. JOP rischia di cadere nelle mani di chi vuole comunque ricavarne un utile. Gli ideatori stanno aspettando una soluzione perché alcuni, pochi, studiosi, non possono svolgere tutto questo lavoro.

Nel campo della matematica segnaliamo che anche l'AMS [1] ha lanciato alcuni periodici *puramente* elettronici; al sito [11] l'AMS elenca i giornali esistenti relativi alla ricerca matematica.

## 9. – Tecnologie «grid» orientate ai servizi.

Tra i vari possibili modi di utilizzo dell'elaborazione distribuita («distributed computing») è stato proposto assai di recente di considerare le cosiddette architetture orientate verso i servizi. Questa terminologia si riferisce a sistemi strutturati come reti di servizi intercomunicanti, debolmente accoppiati fra loro. Si tratta di architetture che definiscono delle interfacce e dei protocolli standard che consentono agli operatori di incapsulare strumenti di informazione come servizi a cui i clienti possono accedere senza conoscere o controllare il modo interno in cui essi lavorano [18]. La forma in cui l'elaborazione distribuita entra qui in gioco è tramite le tecnologie del

cosiddetto «grid computing». Progetti che si rifanno a tali tecnologie non mancano neppure in Europa: ricordiamo ad esempio l'«European Network of Excellence» (NoE), denominato CoreGRID, istituito dalla European Commission's 6th Framework Program [8], e le iniziative francesi ad opera dell'INRIA denominate GridCoord e Grid-RMI [19].

In [18], Ian Foster ricorda che Paul Erdős dichiarava che il matematico è una macchina che trasforma caffè in teoremi, e aggiunge che, analogamente, gli scienziati possono essere considerati come macchine che trasformano dati in cose comprensibili. Tuttavia, il modo in cui un tale ruolo è svolto cambia con il progredire delle tecnologie dell'informazione, e difatti l'impresa scientifica è già stata fortemente influenzata dall'avvento e dai successivi progressi di tali tecnologie, soprattutto a causa di Internet. Il fatto è che i risultati e gli strumenti della Scienza non si trovano soltanto nella letteratura scientifica ma anche nelle banche di dati e nei programmi, e poter accedere a tutte queste sorgenti di informazione è essenziale. La scienza orientata verso i servizi ha il potenziale di accrescere la produttività scientifica rendendo disponibili facilmente e automaticamente a tutti gli scienziati ogni tipo di informazione e di servizi come quelli sopra descritti.

Disporre di un completo accesso ad una vasta quantità di dati scientifici e a biblioteche globali è e sarà di importanza fondamentale nel prossimo futuro, ma è stato osservato che è l'accesso automatizzato mediante programmi software che sarà rivoluzionario per via della maggior velocità alla quale tali programmi sono in grado di lavorare [18]. Creare una rete distribuita *di servizi* è quello che serve a tale scopo. È l'uniformità delle interfacce, il formato uniforme (come HTML, HyperText Markup Language) che fa la differenza – per esempio – tra le liste assemblate manualmente di Yahoo e gli indici calcolati automaticamente di Google [18].

## **Ringraziamenti.**

Devo a Kim Veltman alcune informazioni sul contributo Europeo, basato sulle attività di Paul Otlet e ai progetti dell'INRIA che tendono a sfruttare le tecnologie GRID. Sono grato a Edoardo Sernesi per

avermi segnalato tempestivamente l'articolo di J. Ewing e ad Antonella De Robbio per avermi fornito molti dati recenti su Open Access, Scirus, Wikipedia, e DOI.

## BIBLIOGRAFIA

- [1] <http://www.ams.org>
- [2] <http://www.aps.org>
- [3] <http://arxiv.org>
- [4] <http://front.math.ucdavis.edu>
- [5] V. BUSH, *As we may think*, The Atlantic Monthly, July 1945, 14 pp.  
<http://www.theatlantic.com/doc/194507/bush>
- [6] The *Citizendium* Project, <http://www.citizendium.org>
- [7] A. CONTI, *Scovare le gemme nascoste*. L'algoritmo di Google potrebbe valutare meglio la qualità della ricerca, Le Scienze, n. 454, Giugno 2006, p. 43.
- [8] CoreGRID, <http://www.coregrid.net>
- [9] A. DE ROBBIO - A. MARINI, *L'enciclopedia aperta Wikipedia e matematica*, Notiziario SIMAI, N. 1/2006, pp. 4-9, <http://www.iac.rm.cnr.it/simai>
- [10] <http://www.doi.org>
- [11] <http://www.ams.org/mathweb/mi-journals2.html>
- [12] <http://www.sciencedirect.com>
- [13] D. ENGELBART, <http://www.tecnoteca.it/sezioni/comunicare/iper/augmen>
- [14] J. EWING, *Measuring Journals*, Notices Amer. Math. Soc. **53**, N. 9, October 2006, pp. 1049-1053.
- [15] R. N. MAYO, <http://www.hpl.hp.com/techreports/Compaq-DEC/WRL-TN-60.html>
- [16] A. FIGÀ TALAMANCA, *Come valutare «obiettivamente» la qualità della ricerca scientifica: Il caso dell'«impact factor»*, Boll. Unione Mat. Ital. Sez. A, La Matematica nella Società e nella Cultura, Serie VIII, Vol.II-A, Dicembre 1999, 249-281.
- [17] A. FIGÀ TALAMANCA, *The «impact factor» in the evaluation of research*, Bull. Group Int. Rech. Sci. Stomatol. Odontol. 2002, Jan.-Apr., **44** (1), 2-9.
- [18] <http://www.foreignword.com>
- [19] I. FOSTER, *Service-Oriented Science*, Science, Vol. 308, 6 May 2005, pp. 814-817.
- [20] GridCoord and Grid-RMI, <http://www-sop.inria.fr/oasis>
- [21] S. HARNAD - L. A. CARR - T. BRODY, *How and why to free all refereed research from access- and impact-barriers online, now*, Libraries Webzine (High Energy Physics Libraries Webzine), Issue 4, June 2001. <http://library.cern.ch/HEPLW/4/papers>
- [22] J.E. HIRSCH, *An index to quantify an individual's scientific research output*, Proc. Natl. Acad. Sci. USA **102**, no. 46, November 15, 2005, pp. 16569-16572,  
<http://www.pnas.org/cgi/reprint/102/46/16569>
- [23] <http://www.joplink.net>
- [24] <http://www.jstor.org>
- [25] <http://www.ams.org/mrcitations/search.html>
- [26] <http://www.medra.org/it/faq.htm>
- [27] <http://www.nature.com/news/2005/051212/full/438900a.html>
- [28] T. NELSON, <http://www.ibiblio.org/pioneers/nelson.html>
- [29] *Paul Otlet et le Mundaneum: inventer un Internet de papier*,  
[http://www.boxesandarrows.com/view/forgotten\\_forefather\\_paul\\_otlet](http://www.boxesandarrows.com/view/forgotten_forefather_paul_otlet)

- [30] <http://prola.aps.org>
- [31] <http://www.scirus.com/srsapp>
- [32] R. SPIGLER, *Peer-reviewing and electronic publishing*, High Energy Physics Libraries Webzine, issue 6, March 2002,  
<http://library.cern.ch/HEPLW/6/papers/5/>
- [33] <http://www.springerlink.com>
- [34] A. SWAN - S. BROWN, *Open access and self-archiving: An author study*, Technical Report, External Collaborators, Key Perspectives Inc., 2005,  
<http://eprints.ecs.soton.ac.uk/10999/>  
<http://www.iam.ecs.soton.ac.uk/news/752>
- [35] S.M. ULAM, «Avventure di un matematico», Sellerio ed., Palermo, 1995 [trad. di «Adventures of a Mathematician», Charles Scribner's Sons, New York, 1976].
- [36] [http://en.wikipedia.org/wiki/Vannevar\\_Bush](http://en.wikipedia.org/wiki/Vannevar_Bush)

Dipartimento di Matematica, Università «Roma Tre»,  
Largo S. L. Murialdo, 1 – 00146 Roma, Italy  
[spigler@mat.uniroma3.it](mailto:spigler@mat.uniroma3.it)

