

Un accesso universale alla conoscenza

Parla il fondatore dell'Internet Archive

Fabio Mettieri

giornalista, Milano
yukali@tin.it

L'Internet Archive (<http://www.archive.org>) ha superato l'estate scorsa la notevole dimensione di un petabyte (un milione di gigabyte), con 40 miliardi di pagine web. A Brewster Kahle, co-fondatore di questo progetto nel 1996 e suo attuale direttore, "Biblioteche oggi" ha chiesto come viene gestito, quali sono i suoi obiettivi e come si potrebbero evolvere le biblioteche tradizionali con l'avvento della Rete, incalzate da un lato dalla continua crescita di archivi on line che, come l'Internet Archive, sono di fatto delle biblioteche virtuali, e dall'altro dai numerosi progetti di digitalizzazione di libri e documenti.

Mr Kahle, com'è nato il progetto dell'Internet Archive, e come viene gestito?

Già da studente universitario avevo il desiderio di costruire una biblioteca digitale, ma a quei tempi mancavano ancora gli strumenti necessari per farlo. Così per alcuni anni mi sono dedicato a soluzioni per la pubblicazione di documenti on line. Nel 1996 mi sembrò che finalmente i tempi fossero maturi. Come prima cosa io e miei collaboratori decidemmo di leggere *La biblioteca scomparsa* di Luciano Canfora (Sellerio di Giorgianni editore, 1986), dedicato alla Biblioteca di Alessandria, per trarne ispirazione. Oggi abbiamo una struttura con circa



Brewster Kahle a fianco di uno degli apparati di storage dell'Internet Archive

1.500 server e 25 dipendenti, e moltissime altre persone collaborano con noi come volontari. Oltre che nella sede principale di San Francisco, teniamo i dati in più copie e in diversi continenti, allo European Archive di Amsterdam e nella nuova Biblioteca di Alessandria, in Egitto. Dobbiamo poi tenere conto della fragilità dei bit. Se la carta era in grado di durare per secoli, non esiste un hard disk indipendente dalle configurazioni e capace di tanto, quindi ogni tre anni dobbiamo spostare i

nostri archivi su sistemi più aggiornati.

Oggi quali sono gli obiettivi del progetto?

L'obiettivo fondamentale è garantire un accesso universale alla conoscenza. Dal mio punto di vista, il nostro lavoro è appena agli inizi. I tre fattori necessari sono:

- 1) la conversione e l'archiviazione in digitale di tutto ciò che è stato pubblicato dal genere umano;
- 2) Internet, per portare questo materiale vicino alle persone;
- 3) la volontà politica di vivere in una società aperta.

Su questo non stiamo lavorando solo noi, ma anche noi facciamo la nostra parte. Sul versante dell'accesso, per esempio, in Ruanda abbiamo una *bookmobile* sperimentale, un van con un collegamento satellitare a Internet, una stampante e una rilegatrice, che gira per le zone rurali e permette ai bambini di realizzare da soli i libri che desiderano. Ma il vero problema non è tanto fornire accessi Internet, per esempio in biblioteca, che è banale, quanto piuttosto avere tutto il materiale digitalizzato, in modo da potervi costruire intorno dei servizi da offrire on line.

Un altro obiettivo importante è cercare di avere delle collezioni il più possibile internazionali. Rac-

coogliamo materiale da tutto il mondo e abbiamo accordi con le biblioteche nazionali di dodici paesi. A volte si tratta solo di accordi per la definizione di standard, per favorire lo scambio dei documenti, ma in altri casi la collaborazione è più stretta: per gli Stati Uniti e per la Francia, per esempio, raccogliamo on line dati che vengono poi conservati dalla Library of Congress e dalla Bibliothèque nationale de France.

Quali sono i principali ostacoli per la realizzazione di progetti di questo tipo, per esempio dal punto di vista legislativo?

Oggi ci sono molti soggetti, purtroppo, che non vogliono un libero accesso alla conoscenza. Negli Stati Uniti, per esempio, soltanto adesso ci si sta rendendo conto dei pericoli legati alla legislazione sul copyright, così come si è evoluta negli ultimi quarant'anni. Progetti come le licenze aperte Creative



Commons, in fondo, stanno solo cercando di rimediare a queste storture. D'altra parte, le major hollywoodiane oggi hanno un grande potere, che è molto difficile contrastare. Di fatto, stiamo cedendo alle grandi corporation il controllo di settori molto importanti, come l'istruzione. Per ripristinare una situazione accettabile ci vorrà molto tempo.

Più in generale, oggi dobbiamo decidere se vogliamo ancora le biblioteche oppure no. L'Internet Archive su questo ha una posizione precisa: le biblioteche devono continuare a

esistere ed essere accessibili a tutti. Quindi occorre sia finanziarle sia garantire un quadro normativo che permetta loro di operare, nel nuovo contesto offerto dalla Rete.

Come si evolveranno le biblioteche, quelle di mattoni e calce, nell'era di Internet?

Negli Stati Uniti le biblioteche sono forse l'ultima istituzione civica che gode della fiducia della gente e stanno anche diventando dei luoghi di ritrovo. Il digitale offre a tutte le biblioteche una grandissima opportunità, ma occorre vedere se e quanto decideranno di sfruttarla. Le biblioteche hanno sempre avuto i due valori fondamentali della conservazione e dell'accesso. Se in futuro si limiteranno alla sola conservazione, diventeranno sempre meno rilevanti.

Secondo me c'è il rischio reale che le biblioteche diventino semplici musei dei libri. Negli Stati Uniti, tra l'altro, le ultime leggi rendono sempre più difficile la realizzazione di una biblioteca che nel mondo digitale abbia le stesse caratteristiche delle biblioteche che io ho apprezzato da ragazzo. Oggi ci sono molti soggetti che cercano di contrastare la visione aperta delle biblioteche, o per motivi politici, di censura, o per motivi commerciali. Sul versante commerciale, però, occorre sottolineare che le biblioteche hanno saputo sempre coesistere con il mercato, per secoli, e che oggi nel mondo si spendono per le biblioteche 31 miliardi di dollari l'anno; dato che ben il 25% di questa cifra va direttamente nelle tasche degli editori, non dovrebbe esistere un reale conflitto di interessi.

Qual è il suo giudizio su Google Print e sugli altri progetti di digitalizzazione di libri e documenti?

Google Print è un progetto entusiasmante. Sono cose di cui le biblioteche discutevano da anni, ma era necessario un grande soggetto

Internet Archive, la memoria della Rete

Alexa (<http://www.alexa.com>), ora di Amazon, e Internet Archive, che rimane un'organizzazione non profit, con finanziatori quali la Library of Congress (LOC), Hp e il National Science Foundation (NSF) statunitense, offrono un ricco e variegato insieme di risorse, che sono tutte accessibili dal sito <<http://www.archive.org>> e che si presentano senza troppe distinzioni circa la loro effettiva proprietà, insieme a diverse collezioni che appartengono alla già citata LOC. Vi si trovano testi, audio e video, raccolte tematiche di particolare rilievo come i filmati "effimeri" del vasto Prelinger Archive, e la consociatissima Wayback Machine, l'archivio più grande, che dal 1996 rac-

colle le vecchie pagine web. Il petabyte raggiunto la scorsa estate ora cresce al ritmo di 25 terabyte al mese, cioè di un volume che tutti i mesi è pari ai testi contenuti nei 26 milioni di libri della LOC. Lo scopo di questi progetti è far uscire Internet da quella che lo scienziato Danny Hillis ha definito la "digital dark age", un'epoca in cui il digitale, scomparendo senza essere conservato, lascia l'umanità priva di memoria. Proprio come la mitica Biblioteca di Alessandria, distrutta con tutti i suoi 700 mila volumi nel 646 d.C. (nel 640, secondo altre versioni) per ordine del conquistatore della città, il califfo Omar, che era convinto che l'unico libro utile fosse il Corano.

commerciale, come Google, per realizzare concretamente l'operazione e per stimolare ulteriormente le biblioteche stesse. Certo, il progetto è in parte chiuso, dato che il materiale di Google Print per ora non è scaricabile e non è stampabile, ma potrebbe diventare più aperto in futuro. Si tratta comunque di un'iniziativa che sta catalizzando l'attenzione delle biblioteche, tanto che oggi quasi tutte hanno al riguardo un'impostazione di maggiore apertura rispetto a qualche anno fa. Noi speriamo che numerose biblioteche lancino altri progetti di questo tipo, più aperti di Google Print. Dal canto nostro, oggi offriamo alle biblioteche negli Stati Uniti e in Canada un servizio di digitalizzazione dei documenti che costa solo 10 centesimi a pagina; per le biblioteche più grandi operiamo nelle loro sedi, per quelle più piccole in centri regionali. Tra gli altri progetti di digitalizzazione attualmente esistenti, apprezzo soprattutto quelli incentrati sull'accesso alle informazioni. Creare le collezioni è importante, ma occorre anche renderle disponibili; chi crea collezioni solo per conservarle in cantina non mi interessa. Le iniziative più valide, in questo periodo, sono spesso avviate dalle università e non dalle biblioteche, che purtroppo tendono a concentrarsi molto sulla conservazione e non abbastanza sull'accesso.

Quali saranno gli sviluppi più interessanti di Internet? Quanto è ancora lontana la realizzazione del Web semantico? E, più in generale, che cosa prevede per il futuro della Rete?

La semantica nel web forse non si sta realizzando, non come il sogno romantico che alcuni avevano preconizzato, ma sul versante dei *web services* si sta verificando qualcosa di molto interessante: i computer iniziano a comunicare tra loro. Se il web di anni fa era fatto di persone

Brewster Kahle: *un semplice bibliotecario digitale*

Brewster Kahle, che oggi con modestia si definisce soltanto un "bibliotecario digitale", occupa a pieno titolo un posto di rilievo tra i grandi nomi che hanno contribuito alla realizzazione di Internet. Se i più noti Vinton Cerf, Robert Kahn, o lo scomparso Jonathan Postel hanno implementato i protocolli di comunicazione della Rete, Kahle ha invece lavorato in modo meno appariscente ma su un campo non meno importante, quello della pubblicazione e della ricerca delle risorse on line. Laureato al celebre Massachusetts Institute of Technology (MIT), Kahle contribuì a fondare Thinking Machines, una società specializzata in ricerche full-text sui documenti. Sulla base di questa importante esperienza, nel 1989, quando i Gopher e i web non erano ancora stati varati, implementò i Wide Area Information Server (WAIS), con il primo sistema client server e

ad architettura distribuita per la pubblicazione on line di documenti e per la loro ricerca. I WAIS, una soluzione che venne poi acquistata da America On Line (AOL), adesso non si usano più, ma di certo hanno contribuito sia a spianare la strada agli altri sistemi di tale tipo, web compreso, sia a evidenziare l'importanza del protocollo su cui erano basati, lo Z39.50, che oggi si trova in molte interfacce http per l'interrogazione di sistemi catalografici, dalla Library of Congress al Servizio bibliotecario nazionale, e negli Zclient, che possono essere considerati gli eredi naturali dei vecchi WAIS.

Nel 1966 Kahle co-fondò Alexa Internet, un servizio di ricerca e indicizzazione che venne presto acquisito da Amazon e che tuttora fornisce informazioni e dati a diversi motori di ricerca, e Internet Archive, di cui è direttore.

che parlavano con i computer, oggi i computer parlano con altri computer, per costruire nuovi servizi su quelli già esistenti, per esempio sfruttando quelli di Amazon, o di Google Map, o anche dell'Internet Archive. Questo favorisce la collaborazione tra organizzazioni, con la realizzazione di progetti importanti senza il bisogno di costruire grandi strutture. E avere molte piccole organizzazioni è sempre positivo; anche nel caso dei motori di ricerca, a me piacerebbe che ci fossero centinaia di motori in competizione tra loro per affermare le migliori soluzioni. Ogni scenario in cui ci sono solo due o pochi concorrenti è pericoloso. La vera competizione non si ha quando esistono Coca e Pepsi, ma quando i concorrenti sono centinaia, senza barriere per entrare

nel mercato e per proporre nuove idee. Lo stesso discorso si può applicare a Internet, che secondo me è in costante pericolo. L'evoluzione del web continuerà solo fino a quando Internet riuscirà a restare aperto, malgrado i tentativi di vari soggetti di capitalizzare una sua non apertura, ricercando profitti che nel breve periodo potrebbero essere enormi. Che si tratti di una telco, di un gestore di telefonia mobile, del produttore di un sistema operativo o di un browser, nessuno dovrebbe riuscire a controllare Internet. Noi abbiamo la grande opportunità e il compito di far sì che i grandi monopoli non riescano a esercitare alcun controllo sulla Rete, né con mezzi legali né con sistemi commerciali, e su nessuno dei suoi componenti.