

Internet come fonte?

di Emmanuel Betta e Raffaele Romanelli

Internet muta così rapidamente da rendere estremamente mobili ed effimeri non solo i punti di osservazione, ma la stessa relazione tra i soggetti e i loro oggetti. Su questa mancanza di gerarchie, o quanto meno sulla loro continua trasformazione, vorremmo soprattutto richiamare qui l'attenzione. Siamo peraltro consapevoli che le nostre stesse parole risentono di questa labilità: sono scritte a quattro mani da due autori di generazioni diverse, il cui punto di vista riflette le diverse esperienze nel tempo; prendono lo spunto iniziale dal commento di un testo apparso nel 2002¹, che è a sua volta testimone di problemi e preoccupazioni presto superate da altri problemi e altre preoccupazioni, le quali forse si sono ulteriormente modificate nel pur breve spazio di tempo intercorso tra la discussione e la concreta rifinitura di queste note per la stampa. Va da sé che le nostre stesse osservazioni avranno a loro volta una vita effimera e saranno obsolete in breve volger di tempo, tanto da farci pensare, come molti già pensano, che almeno per questo tipo di riflessioni il mezzo di comunicazione più adeguato non sia la stampa, ma la rete.

Labilità degli assunti, gerarchia delle rilevanze, modi di comunicazione. Ecco alcuni dei nodi da affrontare. Iniziamo con una constatazione preliminare, che riguarda la sempre più penetrante "normalità" dell'uso della rete, e della nostra quotidiana immersione in essa. Internet è diventato un mezzo di cui facciamo uso *normalmente* per prenotare un treno, comprare un libro o rintracciare un indirizzo. Se cerchiamo la correttezza di un'informazione bibliografica che magari negli appunti di un archivio ci è mancata, andiamo a rovistare nella rete e regolarmente, e con relativa facilità, la troviamo. Usiamo la rete per rintracciare qualcosa che sappiamo esistere, perché l'abbiamo già vista o ce ne hanno parlato, ma la usiamo anche per individuare qualcosa che non sappiamo se c'è, lo speriamo, talvolta lo supponiamo. Quel che due ricercatori canadesi, i sociologi Welmann e Hogan, nel 2004 avevano visto come imminente studiando un quartiere di Toronto che era stato cablato con le fibre ottiche e osservando quanto la rete fosse già divenuta uno strumento di uso quotidiano e normale come il frigorifero o il telefono, è oggi divenuto realtà consolidata².

Siamo dunque immersi nella rete in maniera “naturale”, irriflessa, non organizzata. In maniera disorientata. Viene alla mente una delle *Finzioni* di Borges, scritta nel 1942, *Funes, el memorioso*³. Sostanzialmente Funes era un idiota, ricordava tutto nei minimi dettagli ma non era in grado di aver memoria di nulla, era paralizzato dall’incapacità di dare un senso, un’interpretazione qualsiasi, perché non riusciva a operare alcuna selezione nella sterminata massa di informazioni ritenute. Capita di sentirsi così di fronte a un PC connesso in Internet, ma, in realtà, questa prima sensazione di impotente idiotismo di fronte al mare di dati presenti in rete negli ultimi anni è già stata oggetto di numerose riflessioni, che ci consentono prima ancora di discutere la natura e gli effetti di questa apparente disorganizzazione del materiale sulle nostre ricerche e la nostra scrittura, di indagare quale organizzazione sia nascosta in questo flusso di informazioni⁴.

Di fatto, all’utente il materiale si presenta organizzato. Esistono repertori di link, sitografie ragionate, portali generalisti e tematici, *directories*, senza contare i motori di ricerca, che sono il primo strumento usato per entrare in comunicazione con quelle informazioni della rete capaci di soddisfare gli obiettivi di una qualsiasi ricerca. In entrambi i casi, agli strumenti corrispondono diverse logiche di reperimento, di indicizzazione e di presentazione del materiale. E quelle logiche hanno un’influenza sulle modalità con le quali selezioniamo e troviamo le fonti. Entrati nella seconda decade della storia del web già qui possiamo vedere le trasformazioni nei modi di pensare la stessa ricerca in rete⁵.

Nel caso dei repertori, delle sitografie, delle *directories* l’indicizzazione è prodotta da operatori umani, persone che raccolgono, analizzano, valutano e indicizzano i contenuti presenti in rete, fornendo una specie di catalogo ragionato fondato su una verifica diretta. La primissima versione di Yahoo! – che ha inventato e fatto del sistema *directory*, sorta di canale tematico organizzato secondo uno schema simile alla classificazione Dewey, la chiave del proprio successo commerciale – è apparsa nel febbraio del 1994 sul sito dell’università di Stanford e si chiamava *Jerry and David’s Guide to the World Wide Web*; si trattava nient’altro che di un elenco di siti selezionati da Jerry Yang e David Filo, i due inventori di Yahoo!, sostanzialmente i loro *bookmarks*.

Questo modello di indicizzazione presupponeva l’idea che le informazioni in rete fossero in quantità tale da poter essere controllate e ordinate completamente da operatori umani; assumeva, cioè, la possibilità dell’esame esaustivo delle fonti in ragione della scarsità delle fonti esistenti⁶. Bastano poco meno di due anni perché la prospettiva muti in profondità; nasce Google, con una filosofia di segno diverso marcata nel suo stesso nome⁷. Per Sergey Brin e Larry Page, i due studenti di PhD che lo inventarono, il web conteneva ormai una massa smisurata di informazio-

ni, in costante aggiornamento e in crescita esponenziale: l'indicizzazione adeguata, pertanto, non poteva essere che automatizzata⁸. In aggiunta, ritenevano che il web potesse essere inteso come un sistema fondato sulla citazione e l'annotazione: il link e la sua descrizione. L'origine del futuro motore di ricerca più usato nel mondo consistette allora in un sistema chiamato *BackRub*, un algoritmo che scopriva, conservava e analizzava i link esistenti, per verificare chi si collegava a chi⁹, attribuendo ad ogni citazione un peso relativo alla fama di chi citava. Era il 1995, il web consisteva in circa 10 milioni di documenti, con una crescita nell'ordine del 2000% all'anno. Nell'agosto del 1996 sul sito dell'università di Stanford (ancora) compariva la prima versione di Google, e l'anno dopo Brin e Page lo presentarono alla comunità accademica in un articolo destinato a celebrità digitale – *Anatomy of a large-scale hypertextual search engine*¹⁰ – illustrando il cuore dell'efficacia di questo motore di ricerca, quell'algoritmo, ora chiamato *PageRank*, che derivava in qualche misura la propria logica dall'accademia stessa¹¹. Applicava, infatti, alla rete il principio della citazione come metro di misura del valore di un testo. Più un documento aveva link in entrata e in uscita più era presumibile fosse noto, dunque visitato, dunque in qualche misura affidabile per una comunità data come quella della rete. Ma la novità di *PageRank* era soprattutto il pesare la rilevanza assoluta e relativa dei singoli link, valutando sia il numero di link verso un particolare sito, sia il numero di link verso ognuno dei siti collegati al primo. In questo modo, diventava possibile identificare una sorta di autorevolezza algoritmica di chi citava e di chi era citato, la cui precisione migliorava in maniera direttamente proporzionale al numero di documenti su cui era misurata. Si trattava, dunque, del sistema adatto ad una rete in piena espansione.

L'efficacia della ricerca e dell'*Information retrieval* è divenuta un elemento decisivo nell'economia digitale. Nella maggior parte dei casi, il motore di ricerca è il primo strumento con il quale si cercano le informazioni, sia che si tratti di informazioni già note, sia che si tratti di informazioni ambite ma non conosciute. Dal 2003, ogni giorno ci sono state nel mondo 550 milioni di ricerche, con una crescita annuale nell'ordine del 20%. E le modalità delle interrogazioni sono mutate, se è vero che in quasi la metà dei casi si usano due o tre parole, in un caso su cinque una sola parola, e soltanto il 5% delle interrogazioni vede l'uso di più di sei parole. Nonostante ciò, i risultati prodotti dagli algoritmi e dalle combinazioni booleiane sono sempre più precisi, e si approssimano alla ricerca perfetta, vale a dire alla risposta giusta al giusto quesito.

La crescente precisione nelle risposte agli interrogativi più vari ha quindi un rapporto diretto anche con la valutazione dei risultati e, nel caso che qui ci interessa, con la valutazione delle fonti. Perché dietro

l'apparente neutralità matematica, la lista di risultati è il prodotto sofisticato di un insieme complesso di calcoli, che incrociano variabili e parametri differenti, in parte noti e in parte segreti commerciali, ma tutti esclusivamente quantitativo-formali. Benché la prospettiva di sviluppo di Internet sia quella del cosiddetto "web semantico", capace cioè di riconoscere un qualche significato alle sequenze alfanumeriche che compongono i termini di una interrogazione di ricerca, per il momento i motori agiscono semplicemente sul piano della forma testuale, calcolando e ponderando la presenza in rete della sequenza di caratteri alfanumerici inseriti nella maschera di interrogazione di un motore di ricerca. E, allora, per attribuire rilevanza a un documento conta se un termine vi compare, quante volte, in che parti del testo, e conta anche quante volte il sito che lo ospita è aggiornato, quanti documenti contiene, quanto è citato e da chi. La lista dei risultati della ricerca in ordine decrescente di rilevanza è dunque il prodotto dell'incrocio algoritmico di questi e altri parametri. Eppure, benché il significato dei termini non sia ancora considerato, nella maggior parte dei casi anche l'utente più avvertito informaticamente non faticerà a concordare con quella valutazione automatica, in qualche caso attribuirà più peso a un documento piuttosto che a un altro, ma difficilmente considererà rilevanti i documenti presenti oltre la quarta, quinta pagina dei risultati di ricerca.

Nella valutazione delle informazioni e delle fonti, dunque, la stessa sequenza dei risultati richiede una critica e una ponderazione, in quanto la precisione dei risultati è il prodotto combinato dello sviluppo dei sistemi dei motori di ricerca e del fatto che i webmaster e i produttori di contenuti, commerciali o non commerciali – facendo ricorso alle tecniche di marketing e ottimizzazione attraverso i motori di ricerca (*SEM-SEO, Search engine marketing-Search engine optimization*) che, nella loro versione fraudolenta, corrispondono agli *spam* – hanno appreso come organizzare e presentare i propri contenuti in rete in modo che essi vengano valutati adeguatamente dai motori e adeguatamente valorizzati ai primi posti dei risultati di ricerca.

Ma come, da chi, e soprattutto con qualche autorevolezza è costruita questa prima organizzazione? Quando guardiamo una lista di risultati possiamo valutare il peso del motore di ricerca nell'attribuzione di rilevanza al sito e alla fonte, e possiamo considerare quanto hanno fatto i produttori di quel contenuto per farsi notare e valorizzare. Una cosa comunque è certa, che non possiamo rifarci alle autorità canoniche che regolano il lavoro critico.

Vale la pena di sottolineare la distanza che separa questo ordine nascosto dai criteri che ci sono familiari e che poi sono alla radice del pensiero critico. Dove risiede l'autorevolezza delle fonti alle quali normal-

mente ci rivolgiamo? Se pensiamo ad esempio all'opera del tempo, alla stratificazione temporale delle fonti, ci accorgiamo che questa manca del tutto nella rete a causa dell'altissima mortalità dei documenti digitali. In futuro una ricerca sull'oggi si troverà di fronte a un problema di scarsità di documenti a partire dal fatto che gli indici di sopravvivenza di quelli digitali sono estremamente labili. Vi sono aspetti tecnici della questione che pure hanno sempre avuto a che fare con l'autorevolezza delle fonti. Sappiamo ad esempio quanto deve all'invenzione della stampa la stratificazione autorevole del sapere. E sappiamo le preoccupazioni oggi suscitate dalla deperibilità della carta. Ora, se un documento su carta senza acidi ha una vita di circa 100 anni circa, un documento digitale su supporto magnetico ha una vita media di 10-15 anni.

Ma c'è di più. Gli stessi linguaggi adoperati per la scrittura digitale (*software*), nonché gli strumenti necessari per la lettura (*hardware*) hanno vita assai breve. Una serie di strumenti con i quali fino a ieri venivano scritti i documenti non esistono più. Solo dieci anni fa i *browser* più diffusi erano Netscape e Mosaic. Explorer era poco usato e Firefox non esisteva, le pagine html non erano ancora in grado di interagire con le applicazioni di scrittura e di calcolo, né, tanto meno, di mostrare immagini in movimento o far sentire suoni. Se superiamo la prima decade di Internet e andiamo ancora più indietro, i sistemi di videoscrittura usati nei primi anni Ottanta sono oggi forse reperibili, ma con grande difficoltà. Nel 1986, la BBC decise di celebrare i 900 anni dal *Domesday Book*, il censimento sulle terre commissionato nel 1086 da Guglielmo il Conquistatore, varando un progetto che ne ricalcasse l'idea. Furono prodotte 25.000 cartine, 50.000 fotografie, un filmato di 60 minuti, che rappresentavano la fotografia dell'Inghilterra del XX secolo. Mentre il libro del XI secolo è rimasto leggibile e tuttora consultabile al *Public Record Office*, nel 2002 ci si è resi conto che il materiale elettronico del progetto della BBC non era più decifrabile perché mancava la tecnologia *hardware* in grado di farlo, e soltanto con un investimento apposito è stato possibile recuperare questo materiale e renderlo nuovamente fruibile¹². Se tutto ciò non bastasse, si aggiunga l'intrinseca mortalità del documento digitale. Una pagina web vive tra i sessanta e i cento giorni, ed è noto quanto ciò incida sulla possibilità di citarla tra le fonti; non a caso, è invalso l'uso di aggiungere al riferimento la data della consultazione – come facciamo anche noi in queste note – ma si tratta di un provvisorio espediente che non risolve affatto la questione.

Continuando nel paragone tra documento su carta e documento in rete, ricordiamoci che la stampa non è soltanto uno strumento tecnico per la riproduzione e la diffusione del sapere; è un sistema che col tempo ha sedimentato una propria gerarchia di autorità, complessa ma immediatamente riconoscibile allo studioso che vi ha plasmato i propri strumenti

critici. La storia dell'editoria, la biblioteconomia, la salvaguardia del diritto d'autore ce lo dimostrano. Del resto, il nostro seminario ha preso lo spunto da un libro che ci è parso significativo perché ci è noto l'autore, che è legittimato ai nostri occhi dalla sua collocazione accademica, perché è pubblicato in una lingua diversa da quella dell'autore – che è già un segnale di attenzione della comunità scientifica internazionale – ed è apparso per i tipi della *Presses Universitaires de France*, editore conosciuto, con un buon standard tipografico, e così via. Vi sono insomma le ragioni per rendercelo di primo acchito meritevole di considerazione. Chiunque insegni sa che oggi, quando uno studente è incaricato di proporre una relazione o tesina, spesso presenta un testo ricco di riferimenti e di fonti che non ci sono familiari, o che ci sembrano immediatamente poco affidabili. Molte volte sono trascritte dalla rete. Di fronte alle domande dello studente, è però difficile formalizzare il sospetto di inaffidabilità. Che l'autore di un testo, o l'editore, ci siano ignoti, il titolo poco convincente, la forma grafica non familiare, la bibliografia non indicata, le tesi sostenute inconsuete ecc., non è sufficiente ad argomentare la nostra perplessità, almeno in maniera didatticamente efficace, e ciò ci costringe a ricorrere a parametri tradizionali di affidabilità.

In conclusione, l'intero complesso di valori, istituzioni e conoscenze che ci guida nel rapporto con le fonti difficilmente è applicabile alle fonti in rete. Abbiamo fatto riferimento a una serie di parametri che ci guidano nella valutazione delle fonti a stampa: la familiarità con la letteratura scientifica che ci deriva dalla nostra stessa formazione, il retroterra accademico delle opere che consultiamo, la notorietà degli autori di riferimento, e così via, insomma il sistema di segni costruito da sapere accademico, sistema editoriale, forma del prodotto bibliografico, dimensioni e complessità dei testi, apparati di note ecc. Potremmo dire che questo complesso di valori e criteri, questa tradizione culturale stratificata, hanno le loro radici e la loro fonte di autorità nello sviluppo secolare di una "società civile". Si tratta di una società civile strettamente intrecciata con "lo Stato", ovvero con l'altra fonte di validazione dei documenti, in questo caso archivistici più che biblioteconomici. Ci riferiamo agli apparati pubblici della conservazione documentaria, agli archivi. Ricordiamoci che la scienza archivistica nasce legata all'esercizio di un'autorità amministrativa e giuridica, che gli archivi sono costituiti da documenti certificati dall'autorità pubblica, siano atti amministrativi, atti privati sottoposti a certificazione notarile, o anche carte private ricevute però in deposito e ritenute rilevanti dall'autorità pubblica.

Nulla di tutto ciò nella rete. Per riprendere i riferimenti appena fatti, diciamo che la rete è figlia di un'altra società civile e di altri tipi di pubblici poteri. Non siamo certo in grado di definire la natura di questa

diversa organizzazione della società e del potere, anche se vi sarebbero molti spunti per indagarla a partire dalla rete, o per i riflessi che essa ha nella rete. In ogni caso ci pare che l'ipotesi di restituire alla comunità accademica o alle istituzioni la loro funzione certificante chiedendo loro di operare da "filtro" del materiale elettronico – un'ipotesi adombrata tra gli altri da Rolando Minuti – possa essere forse una provvisoria soluzione di emergenza, ma non colga affatto l'essenza del problema, così come ci sembra che i progetti di archiviazione della massa di materiali elettronici – dei quali parleremo adesso – nulla abbiano a che fare con la funzione certificante propria degli archivi quali noi li conosciamo.

Del resto, questa diffusa mancanza di autorità, di centro o di gerarchia non è casuale, e solo in apparenza nasce spontaneamente dall'accumularsi e dal susseguirsi delle iniziative e delle innovazioni. Ha invece molto a che fare con la natura non gerarchica della rete stessa in quanto rete. Ricordiamo che Internet ha origini militari, e come tale ha avuto fin da subito la richiesta-necessità di non avere un centro, in modo da poter sopravvivere ad un eventuale attacco¹³. Quando l'11 settembre sono cadute le Twin Towers alcuni *providers* distrutti hanno interrotto le comunicazioni, che però sono state ripristinate nel giro di dodici ore, appunto perché la rete è una rete e non ha un centro nevralgico. E proprio in quella situazione parabolica e grazie alla articolazione della rete, Google salì alla ribalta facendosi notare per la propria innovativa efficacia, inaugurando in quelle ore il servizio Google *news*, con il quale in tempo reale riuscì a fornire agli utenti gli articoli e le notizie che i siti delle agenzie giornalistiche non erano in grado di dare perché i loro servers erano intasati da milioni di contatti¹⁴. E ci riuscì sfruttando a pieno la natura policentrica della rete.

Qualche spunto per comprendere la natura del potere nell'età dell'informatica potrebbe venirci dallo studio dei vari processi di archiviazione del materiale in rete. Il web è archiviato ogni giorno dal 1996, grazie a Internet Archive, progetto *no profit* che attraverso uno *spider* quotidianamente raccoglie e conserva il materiale della rete¹⁵. Le cifre non semplificano il lavoro. Secondo uno studio dell'università di Berkeley, nel 2002 il genere umano aveva prodotto cinque exabyte di dati immagazzinati su carta, dove un exabyte corrisponde a un quintilione di bytes, ovvero a un 10 seguito da 18 zeri. L'equivalente di 500.000 nuove *Library of Congress* ogni anno. Più del 90% di questi dati erano memorizzati su *hard disk*, dispositivo il cui prototipo fu inventato da IBM nel 1956 ma che divenne familiare a partire dal 1980 quando un *hard disk* della capacità di 5 megabyte fu applicato ad un personal computer AT&T. Il che significa che ogni giorno vengono prodotte e memorizzate in formato digitale più informazioni di quante ne siano state immagazzinate per gran parte della storia.

In realtà, se il web non ha memoria, il tempo e il passato hanno sempre più un ruolo strategico nell'economia della ricerca, per tutti gli attori di Internet, commerciali e non. E molti dati stanno acquisendo un grado di sostanziale immortalità, in gran parte fuori dal controllo dell'utente singolo. Le informazioni sui comportamenti in rete degli utenti nel corso del tempo sono registrate e archiviate da diversi soggetti della rete stessa, perché questi dati, accumulati, formano il capitale conoscitivo essenziale dell'economia digitale della ricerca. Lo si può verificare facilmente tutti i giorni usando gli strumenti della ricerca più comuni, i motori, le barre degli strumenti dei principali tra essi, le maschere di interrogazione. Ognuno di questi servizi usa la storia delle nostre interazioni con il web: il *clickstream*, vale a dire la sequenza dei link che abbiamo usato, i *cookies*, i microfiles di testo che identificano la macchina che usiamo, indicano quando, dove e quanto tempo essa è stata in Internet, cosa ha acquistato o chiesto ai diversi motori di ricerca o alle maschere di interrogazione. Senza contare il tracciato lasciato dal nostro numero IP. Dati che a loro volta si accumulano, creando un'ulteriore massa di informazioni che si intreccia con le ricerche di ogni altro utente quando noi facciamo una qualsiasi interrogazione, a un motore di ricerca, a un qualunque catalogo, o banca dati. In tal modo, *clickstream*, storia degli URL visitati, IP, *cookies* ecc., dialogano tra loro, componendo il profilo degli interessi, delle intenzioni elettroniche del singolo utente, e intrecciandole con gli interessi di utenti che abbiano in comune porzioni più o meno ampie di questi dati creano una nuova massa di informazioni, capitale conoscitivo usato dagli algoritmi per capire le intenzioni dell'utente e costruire la banca dati delle intenzioni. È così integrare le nostre richieste – completando, ad esempio, una nostra interrogazione, o facendoci tornare dove avremmo voluto se non ci fossimo scordati l'indirizzo – oppure indovinare le nostre intenzioni, indicandoci quel libro o quel film o quell'oggetto che corrisponde in modo stupefacente a quanto avremmo cercato se avessimo saputo che esisteva. Qui il tempo e la conservazione del passato sono capitale strategico per costruire questa sorta di *database* delle intenzioni e renderlo efficace e fruttifero.

In una prospettiva completamente diversa, il tempo e l'aggiornamento dei dati stanno acquisendo una funzione rilevante in riferimento al cosiddetto diritto all'oblio e alla conseguente richiesta di cancellazione di dati o informazioni precedentemente pubblicati. Una persona inquisita e poi assolta ha facoltà di chiedere la cancellazione dalla rete delle informazioni riferite al suo arresto o alla propria incriminazione. Non ha ancora, però, la facoltà di chiedere che quei dati vengano cancellati da una ricerca storica pubblicata in Internet. Salvo poi che in rete non è sempre immediatamente percepibile quale sia il confine tra una ricerca

storica e una che non lo è¹⁶. Qui il problema investe direttamente la capacità del motore di ricerca di far riemergere pagine o dati anche molto tempo dopo la loro cancellazione e, d'altra parte, investe l'interesse della stessa ricerca, che in una notizia poi svuotata da un giudizio finale di un tribunale può trovare motivi di interesse euristico che ne giustificano la salvaguardia.

Internet Archive ha archiviato 85 bilioni di pagine web fino ad oggi. Ora, questi dati non sono tutti disponibili. In parte perché sono in quel cosiddetto web invisibile, o *deep web*, cioè quella parte della rete costituita da archivi digitali, banche dati, archiviazioni sistematiche di dati che richiedono chiavi di accesso o di interrogazione per far emergere le informazioni. Ma in parte perché quei dati non sono conservati. Difficilmente avremo la pagina di un quotidiano online del 23 settembre 2000, a meno che quel giorno non sia successo qualcosa di particolare da dover essere ricordato e archiviato. Conservata l'informazione, sorge poi un problema che la critica classica delle fonti ha sottolineato: la coincidenza tra l'organizzazione logica dell'informazione e la sua organizzazione fisica, cioè tra il contenuto del documento e la sua struttura formale. Perché il come ha un significato anche sul che cosa. In rete il come è scritto un documento è particolarmente rilevante; basta aprire un documento web per rendersi conto che i colori, gli sfondi, i segnali java, i link ecc., ne sono una componente essenziale. L'archiviazione operata da Internet archive non risolve il problema di conservare questo complesso di segnali, dagli hyperlink agli elementi grafici. Internet Archive funziona un po' come la funzione *cache* di Google, dove viene conservato il testo nudo e crudo, un'informazione non da poco ma pur sempre un'informazione ridotta rispetto a quella originaria.

C'è di più. Il documento in rete non solo è effimero, ma è anche "aperto". A differenza delle tradizionali note – che pure sono degli ipertesti – il link non torna necessariamente al testo iniziale, ma conduce il lettore a costruire un proprio percorso, e dunque ad un altro testo. Vi è stato il tentativo di utilizzare gli ipertesti come un mero arricchimento del testo stesso. Il saggio di Robert Darnton sulla Parigi di fine Settecento, nel quale la mappa della città apriva a delle prospezioni sui luoghi della sociabilità, sulle attività teatrali e musicali nella Parigi dell'epoca dei Lumi è forse il più significativo, anche perché, pur utilizzando a pieno le nuove possibilità analitiche offerte dalla tecnologia, non ha di fatto aperto una nuova via alla scrittura storica, e, salvo un caso, non risulta aver avuto imitatori né ulteriori sviluppi¹⁷.

Nei termini in cui ne parliamo, l'utilizzo della rete rivoluziona invece il concetto di testo e mette in discussione lo stesso concetto di autore sia dal lato del testo stesso sia da quello del suo fruitore. In altre parole, il

testo perde ogni oggettiva autorevolezza, non è chiaramente identificabile, a volte è collettivo, a volte è anonimo e *si concretizza nel rapporto con il lettore*, mentre il lettore perde autonomia e distanza dal testo, rischia di confondersi con esso o di diventare un mero manipolatore di testi in rete. Una ventina di anni fa, quando il più giovane di noi non aveva ancora iniziato il suo percorso di storico, il più anziano andò a trovare un nipote che frequentava un college in Illinois e vide che le sue ricerche consistevano nell'assemblare dati, notizie e commenti sul tema assegnatogli senza che egli dovesse necessariamente leggerli e acquisirli ad un proprio discorso. Del resto, chi ha esperienza di insegnamento, sa quanto sia difficile spiegare a uno studente la differenza fra una citazione e un plagio, tra un "taglia e incolla" e una scrittura.

Non vogliamo con ciò rinunciare a ristabilire una distanza dalla fonte e una sua valutazione critica. Soltanto che questo compito è assai più complesso di quanto possiamo pensare e l'idea di applicare alle fonti in rete un sano, tradizionale sistema di critica testuale richiede tecniche e conoscenze del tutto nuove, non "curricolari" nei nostri ordinamenti didattici.

Torniamo dunque al problema di come costruire una rilevanza per la massa sterminata e apparentemente informe dei documenti offerti dalla rete. È esperienza comune quella di lanciare una ricerca in Google e di esaminare i primi documenti offerti, che sono comunque una frazione minima di quelli presenti. C'è di più: il dato offerto da un motore di ricerca è a prima vista fortemente decontestualizzato. Prevale il sistema del *deep linking*, si arriva direttamente ad una determinata pagina di un determinato sito, senza passare per la sua *home page*. Né, nella maggior parte dei casi, l'utente medio ripercorre l'indirizzo URL fino alla radice – il cosiddetto *backbrowsing* – per capire dove si trova la pagina alla quale è arrivato. Solo il 30% degli accessi al sito dedicato alla memoria dell'11 settembre¹⁸ è passato attraverso la *home page*, nella quale sono descritti il sito in generale e i materiali che si possono trovare. Quasi tre quarti degli accessi sono direttamente al singolo documento, nel quale peraltro può essere semplicemente presente un termine cercato che ha poco o nulla a che fare con il contenuto proprio del documento. Lo rivela anche la stampa quotidiana quando con una tecnica "puntilista" condisce un articolo con una serie di frasi o citazioni di autori famosi distribuiti nell'arco di secoli. Per operare una nostra selezione possiamo ricorrere a una serie di criteri, che in parte sono anche quelli adottati da Google e che già riducono l'aleatorietà della ricerca che solo pochi anni fa soprattutto preoccupavano Rolando Minuti. Per i soggetti che li gestiscono, istituzionali o non istituzionali, o per un'affidabilità che ci hanno dimostrato nel corso del tempo alcuni siti ci si presentano

come affidabili. Del resto, le domande rituali del ricercatore – perché quella fonte è stata scritta, da chi, come, quando – possono spesso aver risposta nella rete, purché la si sappia trovare. Si tratta ad esempio di quella serie di parametri minimi perché un documento possa avere una identificabilità e una classificazione corretta, vale a dire i *meta-tag*, cioè quei dati invisibili nel *layout* della pagina web ma leggibili nel suo codice sorgente, e nei quali chi ha prodotto la pagina fornisce informazioni sul suo contenuto che possono essere utili per la sua archiviazione e catalogazione: chi è l'autore, come si contatta, quali sono le parole chiave che definiscono l'oggetto della pagina, quando è stata prodotta, dove è ospitata, quando è stata aggiornata ecc. Tutta questa serie di informazioni è l'applicazione in scala al documento digitale di valutazioni concepite in relazione al documento cartaceo. I siti istituzionali, anche quelli che producono informazioni, offrono in genere questo tipo di dati. E non a caso la maggior parte dei siti istituzionali che producono materiale scientifico di scienze sociali e anche di storia sono gestiti da bibliotecari che adeguano alla rete i parametri e le modalità analitiche e classificatorie della biblioteconomia. Tuttavia, dal punto di vista della ricerca storica, non necessariamente il sito cotto è migliore di quello crudo¹⁹; o, fuori di metafora, non necessariamente è più ricco e interessante come fonte storica il sito che corrisponde in pieno agli standard di definizione dei metadati stilati dalla *Dublin Core Metadata Initiative*²⁰ rispetto a quello che ha una forma più rozza e meno mediata.

Tali criteri non sono estranei alla costruzione delle sequenze offerte da Google, che li elabora mediante algoritmi. Ovviamente non sono i soli, e qui ci interessa sottolineare altre variabili che entrano nel calcolo matematico. Tra i criteri almeno in parte trasparenti rientra l'affidabilità. Un'affidabilità a sua volta costruita, e tutta da indagare. I motori di ricerca sono e rimangono aziende commerciali che vendono visibilità e reperibilità, e le strategie di marketing più o meno corrette, nonché la presenza a pagamento nei risultati di ricerca – cosa che avveniva ad esempio in Altavista – sono fattori che incidono in modo più o meno esplicito sull'affidabilità dei risultati di ricerca. Ma a prescindere da ciò, va detto che l'ingresso forte di istituzioni prevalentemente anglofone, inglesi e americane soprattutto, che sono entrate in gioco e hanno scandagliato la rete per organizzare le risorse e metterle a disposizione dell'utente, richiama un aspetto non secondario della questione: la certificazione di affidabilità passa attraverso l'egemonia di una cultura e di un investimento di risorse di tipo "imperialistico".

Anche in questo caso, una concezione tradizionale che identificasse tale potere imperialistico con la concentrazione di capitale sarebbe quantomeno incompleta e non ne coglierebbe la natura profonda. A prima vista

la rete si presenta come una sorta di spazio orizzontale, aperto a chiunque, in un certo senso uno spazio *democratico*. In parte è così: la messa a disposizione gratuita di servers, software, web-editors sempre più semplici, permette a chiunque di far vedere a chiunque ciò che vuole. Ma molti altri fattori complicano terribilmente questa semplicità e questa nozione di democrazia. Chi scrive in rete deve anche costruirsi una “reputazione” che permetta all’informazione che ha messo nel web di poter essere incrociata dai motori di ricerca e quindi di poter essere fruita da altri. Di poter, cioè, emergere dalla massa di dati per intercettare l’attenzione di un utente. Senza contare che deve fare pure i conti con la cosiddetta *danza di Google*, gli aggiornamenti dei propri algoritmi che a cadenza regolare Google produce per contrastare il fenomeno dello *spamming*²¹. Internet appare allora anche come riflesso di rapporti di forza tecnologici ed economici. Quando nel novembre del 2006 Google inaugurò il suo servizio *Google Print*, pubblicazione in rete di una biblioteca digitale, in Francia vi furono molte reazioni critiche; indubbiamente la presenza di quella biblioteca o dei singoli testi selezionati da Google avrebbe risposto a una serie di parametri di selezione stabiliti fuori dall’ambito culturale francese, e in particolare in quello del mondo anglosassone e del suo ambiente accademico. Google Print si appoggia infatti alla rete universitaria americana. Google ha un contratto con la biblioteca dell’università di Berkeley e con le biblioteche di altre dodici università americane per la scansione del loro patrimonio librario. Ad un ritmo di 3.000 libri al giorno, migliaia di persone stanno attualmente scansando libri per rispettare i termini di quel contratto, che se rispettato porterà alla scansione e relativa pubblicazione di 10 milioni di libri all’anno. Posto che il patrimonio librario mondiale è stimato nell’ordine di 65 milioni di titoli esistenti, non ci vorrà molto perché la totalità delle pubblicazioni sia disponibile in qualsiasi angolo del mondo connesso in rete²².

Qualcosa di simile si può dire rispetto a molte raccolte organizzate di fonti bibliografiche. Come spazio aperto, la rete ha abbattuto i costi di molte transazioni. Ma proprio la necessità di fornire materiale selezionato e ordinato e l’organizzazione oligopolistica dei centri accademici maggiori ha prodotto un forte aumento dei costi. Il fenomeno è particolarmente rilevante nel campo scientifico, dove l’accesso tempestivo alla letteratura è determinante per la ricerca. Le riviste sono in rete, ma il prezzo degli abbonamenti è salito vertiginosamente. Questo avviene anche nel campo della ricerca storica. JSTOR, per fare un esempio, non è di libero accesso, e non è raggiungibile se non attraverso le istituzioni che l’hanno acquistata. La fase del libero accesso è stata assai breve, e anche gli editori italiani, che hanno visto restringersi il mercato degli acquisti su carta anche per la concorrenza della rete, non danno più accesso libero al

materiale elettronico. In questo contesto sull'onda della prospettiva *open* inaugurata sul piano del software da Linux, e poi trasferito sul piano dei contenuti tra tutti da *Wikipedia*, un fenomeno analogo sta iniziando ad investire anche l'editoria periodica scientifica, dove sta prendendo piede la produzione di riviste scientifiche *open access*, a partire dall'ambito delle scienze naturali²³.

In questi casi i privati non possono godere del beneficio dell'universalità della rete. La struttura diventata monopolistica è molto costosa, ben poco aperta e molto orientata. Si pensi a ricerche di tipo sociologico, o demografico. La possibilità di immagazzinare ed elaborare una gran quantità di dati ha costituito una grande innovazione all'epoca della prima informatizzazione della ricerca, che forse non a caso ha coinciso con una stagione in cui la storia economico-sociale quantitativa da un lato e l'analisi testuale dall'altro hanno avuto grande fortuna. In quel caso la quantità poteva dare significato e qualità. Qui la metafora potrebbe essere quella dei *pixels*: tantissimi punti, se addensati in un modo che nessuna persona potrebbe fare, danno un'immagine. Era quanto succedeva con le analisi censuarie e demografiche, o con i catasti. Le "immagini" così ottenute rispondevano anche all'esigenza di elaborare modelli, ovvero di ordini sistemici che consentono di intuire il dato mancante; se si conosce la collocazione di *pixels* o di punti, anche se manca un frammento ne possiamo conoscere la collocazione. La ricerca storico-sociale ha fatto così molti progressi. In quella fase iniziale, ciascun progetto lavorava con i propri dati, e spesso elaborava i propri programmi. Con l'avvento della rete, molte serie di dati sono ormai acquisiti e disponibili, e le ricerche possono raggiungerli e compararli. Ma la prevalenza di serie di dati riguardanti i paesi che hanno maggiormente investito nella ricerca e nella informatizzazione – anche in questo caso, per lo più in area anglofona – non può che orientare anche la comparazione.

In conclusione, ciò che ognuno di noi trova nella rete quando digita in un motore di ricerca è il risultato di una serie di parametri ben poco oggettivi, ed è attraverso quei parametri che è possibile ricostruire una visibilità e una capacità di farsi vedere, di mostrarsi, di trovare informazioni, non soltanto storiche, che siano affidabili e che rispondano ai parametri della professione.

Ma c'è un altro e più importante aspetto da sottolineare nella sorta di selezione darwiniana subita dai nostri siti, ed è l'interazione con l'utente. A costruire l'ordine delle rilevanze concorre infatti la frequenza degli accessi. L'alta mortalità dei documenti in rete – che pone il problema della loro archiviazione – dipende anche dal fatto che il documento in rete non venga visto, non venga fruito. Come sappiamo, la sensibilità di Google alla dinamica degli accessi è immediata. I siti che hanno una più

lunga esistenza e appariscenza contano dunque su un pubblico che è soggetto attivo della visione e dell'esistenza del sito perché nel vederlo, nel servirsene, gli attribuisce una visibilità che è parte rilevante e integrante del suo persistere oggi. Il documento è tale, si può dire, perché è vitale, è oggetto di comunicazione attiva.

È anche noto che in alcuni casi questa partecipazione dell'utente alla costruzione del documento va oltre, e incide sui contenuti stessi del documento. In *Intute* o in *Bubl*, o in altri portali di scienze sociali si può interagire con colui o colei che ha fatto la classificazione di una serie di documenti e chiederne conto²⁴. In *Wikipedia* è lo stesso utente che produce il contenuto, un utente collettivo, composto da tutti coloro che vogliono integrare o aggiornare le informazioni di un singolo lemma. In parte è un approdo di quella prospettiva propria della rete per la quale ognuno ha la possibilità di essere storico di se stesso, congiunta però con la filosofia di lavoro aperto e collettivo propria di Linux, di Firefox, *openoffice*, che è stata di recente adottata anche da Google per un pacchetto *office online* in concorrenza con l'analogo software di Microsoft²⁵. Il singolo lemma, dunque, non ha più un autore singolo, né, pertanto, la responsabilità delle cose scritte può più essere ricondotta ad una individualità. L'attivazione di un'intelligenza collettiva è stata il volano del successo di *Wikipedia*, che, nato nel 2001, è in breve divenuto uno dei punti di riferimento della rete: 20 milioni di pagine pubblicate, 6,8 milioni di voci in 250 lingue diverse, con una crescita esponenziale degli accessi. E questo successo si è riflesso direttamente anche sul piano della qualità, se è vero che "Nature", nel 2005, in un'analisi comparata con l'enciclopedia professionale per eccellenza, l'*Encyclopaedia Britannica*, su 42 voci in comune, analizzate con il metodo del *peer review*, ha rilevato 4 imprecisioni per l'enciclopedia gratuita e *open source*, e 3 per la consorella d'oltremarina, costosa e professionista²⁶. La modalità di lavoro *open source* non ha mancato di avere un impatto problematico sui contenuti delle singole voci e sulla loro affidabilità, che è stata spesso tacciata di relativismo in quanto non fondata su un'autorità e una competenza riconosciuta in quanto individuale e certificata, ma su una competenza collettiva e *in fieri*²⁷.

La richiesta di un'assoluta neutralità del punto di vista, cardine delle linee guida di *Wikipedia*, ha un effetto particolarmente problematico per le voci di interesse storico, per le quali la tendenza all'annullamento di qualsiasi interpretazione porta non di rado la voce a sfiorare l'aneddotica. Effetto che, già problematico di per sé, si moltiplica perché le stesse voci di *Wikipedia* sono riprodotte testualmente in molti altri siti della rete, determinando nei fatti una sorta di uniformazione depurata di ogni parvenza di interpretazione. Nonostante questa richiesta di neutralità,

i problemi sui contenuti non sono mancati. Fino alla fine del 2005, era possibile collaborare a *Wikipedia* anche in forma anonima, modificando le voci e integrandole senza registrazione preventiva. Sistema efficace per aumentare il numero di collaboratori e la quantità delle voci, ma che favoriva imprecisioni o falsità. Il caso più celebre è quello di John Seigenthaler, direttore editoriale di “Usa Today” ed ex assistente di Robert Kennedy, che nel 2005 nella voce che lo riguardava si vide coinvolto nell’uccisione di entrambi i fratelli Kennedy, nonché ospite in URSS per tredici anni: informazioni false, che tuttavia erano rimaste visibili per 132 giorni prima di essere corrette da qualcuno. A seguito di questo caso, e della relativa minaccia di azioni legali, *Wikipedia* introdusse la registrazione preventiva al sito per poter operare modifiche sulle voci. Soluzione che ha limitato i problemi, ma non li ha cancellati. Regolarmente vi sono attacchi (vandalismi nel linguaggio wikipediano) a singole voci. È il caso, ad esempio, del lemma relativo a Simon Wiesenthal, modificato all’indomani della sua morte, il 20 settembre 2005, da hacker neonazisti. Contro questi attacchi, l’enciclopedia prevede la protezione di singole voci la cui modifica è appannaggio esclusivo dei soli amministratori. E così, tra le voci protette della versione italiana si possono trovare “Partito socialista democratico italiano”, insieme a “Giovani Psdi” e “Gioventù socialdemocratica italiana”, “Anticomunismo” e “Carlo Giuliani”. A un diverso grado di protezione, tra le pagine semi-protette, modificabili solo da utenti “anziani” – registrati cioè da più di 4 giorni (*sic*) – si possono trovare “Lincoln” e gli “Attentati dell’11 settembre”, “L’eccidio delle Fosse Ardeatine” e la “Foiba”, il “Genocidio armeno” e la “Storia del cinema”. Il passaggio ai diversi gradi di protezione, così come la sospensione in attesa di controllo o la cancellazione, vengono decisi dagli amministratori insieme agli utenti, previa discussione e votazione negli appositi forum di discussione presenti sul sito. Questi diversi gradi di controllo non hanno tuttavia eliminato i problemi o i contrasti, al punto da indurre uno dei due fondatori di *Wikipedia*, Larry Sanger, ad inaugurare nel settembre 2006 un nuovo progetto editoriale di enciclopedia gratuita, *Citizendum*, in tutto identico al modello originario di *Wikipedia*, salvo il non trascurabile dettaglio che le singole voci verranno sottoposte a vaglio e approvazione di personale esperto, interno o esterno alla redazione²⁸.

Come si vede, parlare di una intrinseca “democraticità” della rete va ben oltre il nostro campo di osservazione e investe la natura della moderna democrazia, informatica e non. O del mercato aperto, di cui eBay offre un modello di grande interesse teorico, giacché sembra che in quel caso l’iterazione degli scambi conduca non solo alla fissazione dei prezzi ma anche alla fiducia tra contraenti anonimi. Qualcosa di simile si potrebbe dire per l’intervento sui siti, che sono aperti a un controllo

d'opinione che è alla portata di tutti senza mediazioni istituzionali. Il mercato perfetto e la democrazia diretta. Le implicazioni per la ricerca in campo storico sono molteplici. La mancanza di memoria di Funes diventa una risorsa, perché consente, e anzi sollecita, una risistemazione continua. L'autorevolezza dei siti deriva dalla perdita dell'autorità. Si noti che questo spostamento dall'autorità certificata delle istituzioni pubbliche allo spazio aperto della comunicazione sembra lambire lo stesso apparato archivistico. L'incapacità degli archivi pubblici di accogliere, selezionare e conservare una mole crescente di materiale cartaceo – per non dire del materiale informatico o audiovisivo – richiede una collaborazione degli utenti nell'orientare le scelte archivistiche²⁹.

Ma siamo davvero in presenza di un ampliamento della sfera della democrazia informatica? Inutile dire che la possibilità di ridurre al massimo l'intermediazione tra chi produce un testo e chi pubblica ha aspetti sicuramente positivi, che ampliano la sfera della libertà di espressione. Produrre una rivista per farla leggere a dieci persone costa tanto quanto produrla per farla leggere a dieci milioni di persone, mentre un documento cartaceo ha dei costi differenti se è prodotto per dieci o per dieci milioni di persone. È anche vero però che il documento prodotto di anni di ricerca e di una metodologia sofisticata, sulla rete può avere una visibilità uguale o inferiore ad uno scritto ideologico, sommario, prodotto rapidamente senza grossi lavori di ricerca.

Facciamo un esempio al quale ci siamo interessati di recente, quello dei siti negazionisti e di estrema destra. In una recente ricerca sull'"odio in rete" un sociologo bolognese, Antonio Roversi, ha esaminato tre tipi di siti: dei tifosi di calcio, dei neonazisti e neofascisti e della *jihad* islamica più radicale³⁰. La maggior parte di questi siti, in particolare quelli neonazisti e neofascisti, presentano materiali che una volta verificata l'affidabilità possiamo tranquillamente chiamare fonti: video e filmati originali di Mussolini, trascrizioni più o meno complete di molti documenti del periodo fascista, e così via. Nella maggior parte dei casi quei siti sono costruiti con intelligenza degli strumenti della rete, con la consapevolezza di che cos'è la rete, di come funziona, di come restituisce i risultati delle ricerche, e quindi sono strumenti che entrano in gioco pesantemente nella produzione della memoria, della selezione dei ricordi del nostro Funes.

La ricerca in rete può arrivare molto più facilmente a incrociare documenti di questo tipo che a incrociare documenti più affidabili, prodotto di lunghe, complicate e serie ricerche di storia. Perché quei contenuti sono scritti, presentati e organizzati in modo tecnicamente funzionale a farli emergere il più possibile dalla massa di informazioni esistenti per imporsi all'attenzione di quel Funes incapace di selezionare, ma capace di tenere tutto, anche quando egli sta cercando qualcosa di altro, o qualcosa

che ha una lontana parentela con quei contenuti. Tuttavia la caratteristica essenziale della maggior parte di questi siti di neonazisti, che pure trasmettono singole informazioni attendibili, è quella di comunicare un senso di appartenenza, un contenuto identitario più che un contenuto informativo. Nel presentare vicende già note, consolidano nell'interlocutore una lettura della storia semplice e lineare ma molto precisa, che rassicura l'interlocutore di appartenere a un gruppo, a una comunità che non ignora i documenti storici ma che fornisce una lettura differente – spesso dichiarata più veritiera – di determinati percorsi storici.

Ma questo schema – che nei siti negazionisti è portata all'estremo – si ritrova molto frequentemente nei siti di storia, in particolare nei siti amatoriali, nei siti minori, che hanno una vita molto più breve. In questi casi l'informazione non è immediatamente affidabile, ma se attentamente verificata può essere utile per un percorso di ricerca, o per completare una carenza informativa. Da questo punto di vista l'ambiguità della rete risiede nel fatto che i percorsi più o meno casuali del reperimento delle informazioni digitali non di rado conducono la ricerca anche in siti che non avremmo nessuna difficoltà a identificare come inaffidabili, incompleti, storiograficamente discutibili.

Note

1. R. Minuti, *Internet et le métier d'historien*, PUF, Paris 2002; una versione italiana in "Cromohs", 6, 2001, pp. 1-75, URL: http://www.cromohs.unifi.it/6_2001/rminuti.html. Questo come i successivi URL sono stati consultati nel marzo 2007. Il seminario, organizzato dall'École française de Rome, con il Dipartimento di storia moderna e contemporanea dell'Università di Roma "La Sapienza" ha avuto luogo il 6 febbraio 2006.

2. B. Wellman, B. Hogan, *The Immanent Internet*, in J. McKay (ed.), *Netting Citizens: Exploring Citizenship in a Digital Age*, St Andrew Press, Edinburgh 2004, pp. 54-80; anche all'URL: <http://www.chass.utoronto.ca/~wellman/publications/immanent/immanent.html>.

3. J. L. Borges, *Funes l'uomo della memoria*, in Id., *Finzioni*, Adelphi, Milano 1996, pp. 95-104.

4. Per una densa riflessione sulle implicazioni per la ricerca storica cfr. ora T. Detti, G. Lauricella, *Una storia piatta? Il digitale, Internet e il mestiere dello storico*, in "Contemporanea", 2007, 1, pp. 3-23.

5. A questo proposito cfr. il sintetico bilancio di D. J. Cohen, *History and the second decade of the web*, in "Rethinking history", 2004, 8, 2, pp. 293-301.

6. Sull'impatto della sovrabbondanza di informazioni della rete sulla ricerca storica cfr. R. Rosenzweig, *Scarcity or abundance? Preserving the past in a Digital Era*, in "The American Historical Review", 2003, 108, 3, pp. 735-62, disponibile all'URL: <http://chnm.gmu.edu/resources/essays/scarcity.php>.

7. Il nome Google è variazione del termine inglese *googol*, ideato nel 1938 da Milton Sirota, nipote del matematico americano Edward Kasner e da questi usato per indicare il numero intero esprimibile da 1 seguito da 100 zeri.

8. Nel 2004 l'indice di Google aveva raggiunto la cifra di 6 miliardi di voci.

9. Google occupa il 51% del mercato globale della ricerca, seguito da Yahoo! 24%, Msn 13%, AskJeeves 5%, Aol 5%, dati riferiti al primo trimestre del 2005. Per questo dato

e altri successivi, così come per un inquadramento della ricerca come problema strategico della economia digitale cfr. J. Battelle, *Google e gli altri. Come hanno trasformato la nostra cultura e riscritto le regole del business*, Raffaello Cortina, Milano 2006 (ed. or. 2005); il riferimento è alle pp. 43-4.

10. Reperibile all'indirizzo URL: <http://infolab.stanford.edu/~backrub/google.html>.

11. Per la storia di Google, cfr. D. A. Vise, M. Malseed, *The Google story*, Delacorte Press, New York 2005.

12. Il progetto è visibile a questo indirizzo: <http://www.domesday1986.com>; per una descrizione: <http://www.atsf.co.uk/dottext/domesday.html>. Sui temi della conservazione dei dati digitali in Inghilterra cfr. la *Digital Preservation Coalition*, <http://www.dpconline.org/graphics>.

13. Su questo cfr. K. Hafner, M. Lyon, *La storia del futuro. Le origini di Internet*, Feltrinelli, Milano 1998; Detti, Lauricella, *Una storia piatta?*, cit.

14. Cfr. R. Wiggins, *The effect of September 11 on the leading search engine*, www.firstmonday.org/issues/issue6_10/wiggins; Battelle, *Google e gli altri*, cit., pp. 168 ss.

15. <http://www.archive.org/index.php>.

16. In Italia l'Autorità garante per la protezione dei dati personali è intervenuta a più riprese su un problema di questo tipo. Il provvedimento del 9 novembre 2005 (documento web n. 1200127) riguarda la richiesta di cancellazione di dati conservati nel sito di un quotidiano, relativi ad un'incriminazione che ha poi dato luogo ad assoluzione. Il 28 settembre 2006 (doc. web n. 1361916) la stessa autorità ha respinto invece la richiesta della medesima persona di rendere anonimi i dati relativi al suo proprio arresto presenti in un dossier sugli intrecci tra finanza e politica dal 1943 ad oggi, pubblicato sul sito della Fondazione Luigi Cipriani (<http://www.fondazionecipriani.it>). In quest'ultimo caso, posta l'essenzialità e pertinenza delle informazioni rispetto ad un interesse storiografico ha prevalso la tutela della libertà di ricerca ed espressione. I documenti sono consultabili sul sito dell'autorità: <http://www.garanteprivacy.it>. Un grazie ad Alessandra Pierucci per averci segnalato il caso.

17. R. Darnton, *An early information society: news and the media in Eighteenth-Century Paris*, in "The American historical Review", 2000, 105, pp. 1-35; URL: <http://www.historycooperative.org/journals/ahr/105.1/ah000001.html>. Al numero del dicembre 2000 è stato allegato il solo saggio che ha seguito la proposta di Darnton, in modo ancor più articolato: Ph. J. Ethington, *Los Angeles and the problem of Urban Historical Knowledge*, URL: <http://cwis.usc.edu/dept/LAS/history/historylab/LAPUHK/index.html>.

18. <http://www.911digitalarchive.org>.

19. Cfr. D. J. Cohen, *Digital history: the raw and the cooked*, in "Rethinking history", 2004, vol. 8, 2, pp. 337-40.

20. <http://dublincore.org>. Cfr. anche il progetto *Historical event markup and linking project*, <http://www.heml.org>, che propone metadati relativi alla pubblicazione in rete di materiale di interesse storico.

21. Cfr. Battelle, *Google e gli altri*, cit., pp. 186 ss.

22. In merito cfr. D. Pudles, *Not bound by anything*, in "The Economist", 22 marzo 2007.

23. Cfr. *Public Library of Science*, open access che pubblica dal 2003 riviste specializzate nell'area medico-biologica: <http://www.plos.org/>.

24. <http://www.intute.ac.uk/>; <http://bubl.ac.uk/>.

25. R. Rosenzweig, *Everyone a historian*, in <http://chnm.gmu.edu/survey/afterroy.html>; Id., *Can History be an Open Source? Wikipedia and the Future of the Past*, in "The Journal of American History", 2006, 93, 1, pp. 117-46; disponibile anche all'URL: <http://chnm.gmu.edu/resources/essays/d/42>. Per il pacchetto office *open source* cfr. <http://www.openoffice.org>.

26. Nel numero del 15 dicembre 2005, la rivista scientifica "Nature" ha comparato con il metodo del *peer review* 42 voci presenti in entrambe le enciclopedie, riscontrando

INTERNET COME FONTE?

4 imprecisioni in *Wikipedia*, 3 nella *Encyclopaedia Britannica*. L'articolo, con ulteriori interventi, è disponibile all'URL: <http://www.nature.com/nature/journal/v438/n7070/full/438900a.htm>.

27. A. Spadaro, S. J., *Wiki. Utopie e limiti di una forma di «intelligenza collettiva»*, in "La civiltà cattolica", 2005, 3, pp. 130-8, disponibile all'URL: <http://www.laciviltacattolica.it/quaderni/2005/3722/Articolo%20Spadaro.html>.

28. <http://www.citizendium.org>.

29. Lo osserva Stefano Vitali nel suo *Abbondanza o scarsità? Le fonti per la storia contemporanea e la loro selezione*, in C. Pavone (a cura di), *Storia d'Italia nel secolo ventesimo. Strumenti e fonti*, vol. 1, *Elementi strutturali*, Ministero per i Beni e le attività culturali, Dipartimento per i beni archivistici e librari, Direzione generale per gli archivi, Roma 2006, pp. 21-50.

30. A. Roversi, *L'odio in Rete. Siti ultras, nazifascismo online, jihad elettronica*, Il Mulino, Bologna 2006.