

## Che cos'è il social software? Architettura delle reti e politiche del nuovo discorso scientifico<sup>1</sup>

Francesca Di Donato  
(Università di Pisa)

«Il *social software*», ha affermato Clay Shirky, docente di New Media alla New York University, «è l'ala sperimentale della filosofia politica, una disciplina inconsapevole di avere un'ala sperimentale». E prosegue: «nei nostri strumenti (*tools*) stiamo letteralmente codificando i principi di libertà di parola e di libertà di espressione. Abbiamo perciò la necessità di discutere gli obiettivi espliciti di quello che stiamo sostenendo e tentando di fare, poiché si tratta di una discussione importante»<sup>2</sup>.

Ma che cos'è il social software? Con l'espressione si intendono applicazioni orientate ad aiutare la collaborazione fra gruppi che lavorano a distanza, includendo tanto le tecnologie sottostanti, quanto gli aspetti sociali implicati. Da un punto di vista tecnico, sono applicazioni che comprendono sia mezzi relativamente antichi (come le mailing list o Usenet, la cui invenzione risale agli albori di Internet) sia le più recenti piattaforme del cosiddetto Web2.0, come i blog o i wiki. Esso è riferito a numerosi modelli di interazione in cui gli utenti possono formare comunità on-line usufruendo di modelli di comunicazione “uno-a-uno” (e-mail), “uno-a-molti” (blog) e “molti-a-molti” (wiki, peer to peer). Infine, il processo di creazione del software sociale è basato su un modello “bottom-up”, in cui gli obiettivi e l'organizzazione dei contenuti sono stabiliti dai membri della stessa comunità. Ma che cosa significa tutto questo, in termini socio-politici?

Per rispondere a questa domanda, la rete dev'essere considerata come sistema socio-tecnico in cui gli elementi tecnologici si combinano con quelli sociali. È stato osservato che l'architettura è politica, e certamente la struttura di comunicazione determina tanto l'architettura della rete quanto la modalità di partecipazione al suo interno. Ad esempio, la legge di Sarnoff ci dice che nelle reti che obbediscono al modello uno-molti, noto anche come modello broadcast, il valore cresce linearmente con il numero degli utenti. Mentre in una rete peer to peer (molti-molti) il valore cresce secondo il quadrato del numero degli utenti, come dimostra la legge di Metcalfe. Internet è stata concepita considerando il rapporto tra reti matematiche e reti fisiche tenendo conto anche dei gruppi che condividono idee e interessi comuni.<sup>3</sup> Social software è dunque un'espressione complessa che non ha una semplice connotazione tecnica, ma anche culturale, sociale e, appunto, politica.

Nell'intervento da cui è tratta l'osservazione in apertura, Shirky invita ingegneri del software, sviluppatori e web designer a puntare a un obiettivo pratico e assieme elevato: trovare i modi migliori per la collaborazione online, così da produrre tecnologie in grado di contribuire in modo virtuoso alla formazione culturale dei loro utilizzatori. Più in generale, il mass-mediologo li richiama a occuparsi delle implicazioni politiche delle loro applicazioni, ritenendo che sia

---

<sup>1</sup> Questo articolo è stato presentato al Forum International de Philosophie Social et Politique, svoltosi presso l'Université de Toulouse Le Mirail a Tolosa il 6-10 luglio 2009. Desidero ringraziare quanti lo hanno letto, in tutto o in parte, per i consigli e gli spunti di riflessione che mi hanno offerto, in particolare: Marcella Aglietti, Michele Barbera, Alessandro Breccia, Carmelo Calabrò, Brunella Casalini, Matteo D'Alfonso, Riccardo Di Donato, Francesca Menchelli-Buttini, Emanuela Minuto e Maria Chiara Pievatolo. Questo testo è rilasciato con una licenza Creative Commons, Attribuzione – Non Commerciale – Condividi allo stesso modo 2.5 Italia. Leggi le condizioni: <http://creativecommons.org/licenses/by-nc-sa/2.5/it/>

<sup>2</sup> Si veda <http://radar.oreilly.com/archives/2006/03/etech-clay-shirky-1.html>.

<sup>3</sup> M. Berra, *Sociologia delle reti telematiche*, Laterza, Roma-Bari, 2008, pp. 17-20.

compito dell'ala sperimentale della filosofia politica costruire una sorta di contratto sociale che veda sancire i diritti dell'individuo, pena il rischio di precipitare, tanto nello spazio virtuale quanto nell'arena politica, in un dispotismo. “Code is law”, scriveva nel 1999 Lawrence Lessig. «Il controllo del codice è potere. Per i cittadini del cyberspazio, il codice – testo arcano scritto in linguaggio altamente formalizzato, tipicamente accessibile a pochi privilegiati - è il nesso attraverso cui le intenzioni si realizzano e il design prende forma. Esso sta diventando un tema cruciale nel contesto politico»<sup>4</sup>.

Ma perché delegare ai tecnici l'ala sperimentale di una disciplina, la filosofia politica, che ha una forte impronta teorica e possiede tutti gli strumenti per comprendere i meccanismi di potere che il software nasconde o manifesta? È convinzione di chi scrive, infatti, che filosofi e scienziati politici e sociali abbiano in merito molto da dire.

Questo contributo prende sul serio l'affermazione di Shirky e considera le tecnologie del software sociale da un punto di vista filosofico, sociale e politico a partire da un'analisi della sua filosofia tecnica. L'obiettivo delle pagine che seguono è infatti duplice: in primo luogo, definire le caratteristiche filosofiche, socio-culturali e politiche del social software. In secondo luogo, fare da sponda all'invito di Shirky affrontando la questione del modo in cui si definisce il rapporto tra gli utenti e i produttori (architetti) di tecnologie web.

Il paragrafo 1 ripercorre dunque le tappe fondamentali della storia di Internet e della filosofia alla base della sua architettura (i principi architettonici), preconditione essenziale alla nascita delle applicazioni e delle pratiche che ricadono nella definizione di software sociale. L'infrastruttura fisica che collega in rete tramite cavi un'insieme di calcolatori; il software (protocolli e programmi) che permette lo scambio di informazioni tra computer, e le comunità che interagiscono tramite la rete sono infatti un momento preliminare fondante e fondativo di tale processo.

I paragrafi 2-4 concentrano poi l'attenzione sul World Wide Web, culla del social software in quanto luogo e condizione in cui le reti sociali on-line sono nate e si sono sviluppate. Il Web viene qui considerato adottando una prospettiva interdisciplinare che si colloca tra la storia delle idee e la filosofia della tecnica nell'ambito della cosiddetta scienza del Web, una disciplina nata allo scopo di stabilire una corrispondenza tra la progettazione di architetture informative e il modo in cui le infrastrutture tecniche e sociali dell'informazione strutturano i processi culturali e comunicativi.<sup>5</sup>

Il paragrafo 5 è quindi dedicato alla topologia di Internet e del Web, mentre un'apposita successiva sezione, il paragrafo 6, si occupa delle sfide filosofiche che si accompagnano alle recenti evoluzioni tecnologiche del Web (note come web semantico).

Il paragrafo 7 prende in esame un aspetto particolare e relativo al metodo tramite cui si è andata definendo l'architettura delle reti telematiche, vale a dire la politica della comunicazione scientifica sottostante allo sviluppo delle reti (dalle RFC all'open access), passando per alcuni momenti importanti: la nascita del free software e di linux, il progetto creative commons, e la diffusione del concetto di copyleft.

Infine nell'ultimo paragrafo (8) si arriva a definire il software sociale, la tecnologia sottostante e le trasformazioni socio-culturali legate alla diffusione delle reti e delle applicazioni sul web. «Le

---

4 L. Lessig, *Code and other laws of cyberspace*, Basic books, 1999, p. XX.

5 La scienza del Web si vale di un approccio interdisciplinare e risulta essenziale a comprendere come i collegamenti informali tra persone, agenti software, basi di dati, organizzazioni e risorse possano condizionare tanto l'e-science quanto l'e-government. Si vedano T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. Weitzner, “Web Science,” *Science*, vol. 313, August 11th 2006 e soprattutto T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, Foundations and Trends in Web Science, Vol. 1, No 1 (2006) 1–130 2006, p. 72 online all'URL: <<http://eprints.ecs.soton.ac.uk/13347/>>.

reti telematiche, combinandosi con le reti sociali, hanno le potenzialità di far crescere una nuova organizzazione e un nuovo sistema dinamico di relazioni nella società, che ridefiniscono aree, strategie e poteri. Il risultato di queste trasformazioni non è univoco e determinato, ma ambivalente e segnato da contraddizioni.»<sup>6</sup>

## 1. Internet: momenti essenziali della storia di un'idea

«La produzione storica di una data tecnologia, scrive Manuel Castells ne *La galassia Internet*, determina i suoi contenuti e le sue utilizzazioni in modi che durano al di là dei suoi primi passi. [...] La storia di Internet ci aiuta dunque a capire i percorsi del suo futuro procedere nella storia.»<sup>7</sup> Rintracciare i momenti essenziali della storia di Internet<sup>8</sup> ci consente di trarre alcuni insegnamenti utili a interpretare il presente e a comprendere il futuro tanto delle trasformazioni tecnologiche quanto delle pratiche sociali e comunicative che ne derivano. Una storia che considero qui dal punto di vista dell'evoluzione tecnologica, della definizione di una filosofia e degli effetti di tale filosofia nelle pratiche di comunicazione e interazione in rete.

Le prime innovazioni tecnologiche che fanno da premessa alla nascita di Internet risalgono agli anni '60 del secolo scorso e sono dovute a un gruppo di ricercatori, molti dei quali appena laureati, che lavorano presso i centri di ricerca delle migliori università americane (il MIT di Boston e le università della California, in primis Stanford) e DARPA (Defense Advanced Research Projects Agency), l'agenzia governativa del Dipartimento della Difesa degli Stati Uniti incaricata dello sviluppo di nuove tecnologie per uso militare. Oggetto di una ricerca congiunta e parallela, è il progetto di reti di calcolatori in grado di resistere ad attacchi nucleari, per lo scambio di documentazione e informazioni.

La prima innovazione fondamentale è, nel 1961, l'introduzione, per opera di Leonard Kleinrock e Paul Baran, del concetto di rete a commutazione di pacchetto<sup>9</sup>, una rivoluzionaria tecnologia di trasmissione delle telecomunicazioni su cui, negli anni 60, lavorano in parallelo tre centri di ricerca statunitensi<sup>10</sup>. Nell'anno successivo J.C.R. Licklider<sup>11</sup>, ricercatore al MIT, fa circolare la prima descrizione delle interazioni sociali che avrebbero potuto essere rese possibili attraverso il networking, nella serie di appunti scritti sulla "Galactic Network". Sempre nel 1962 Licklider sarà poi a capo di DARPA. Da DARPA viene inoltre costituita un'ulteriore agenzia, orientata alla ricerca e sviluppo, l'ARPA (Advanced Research Projects Agency), per mobilitare risorse di ricerca, in particolare dal mondo universitario, e come polo di aggregazione e di coordinamento dei migliori centri di ricerca statunitensi sul computer networking.

Tra il 1965 e il 1967 Lawrence G. Roberts<sup>12</sup>, al MIT, connette due computer tramite telefono e poi, trasferitosi a DARPA, scrive il primo progetto di rete distribuita a commutazione di pacchetto: ARPANET<sup>13</sup>.

6 M. Berra, *Sociologia delle reti telematiche*, cit., p. VI.

7 M. Castells, *Galassia Internet*, Feltrinelli, Milano 2006, p. 21.

8 Si veda in particolare la storia narrata dagli stessi inventori di Internet: B.M. Leiner, V.G. Cerf V.G., D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts, S. Wolff, *A Brief History of the Internet*, version 3.32, 10 Dec 2003 online all'URL: <<http://www.isoc.org/internet/history/brief.shtml>>.

9 Per una definizione elementare, cfr. Wikipedia <[http://it.wikipedia.org/wiki/Commutazione\\_di\\_pacchetto](http://it.wikipedia.org/wiki/Commutazione_di_pacchetto)>. Per una definizione più approfondita, si confronti A.S. Tanenbaum, *Reti di calcolatori*, Pearson Education Italia, Milano 2003 (quarta ed.), in particolare alle pp. 150-51.

10 Il MIT, la RAND Corporation (<http://www.rand.org/>) e il National Physical Laboratory (NPL) britannico lavorano in modo autonomo e parallelo sul concetto di rete a pacchetti tra il 1961 e il 1967. Il nome "pacchetto" viene concepito a NPL. Cfr. B.M. Leiner, V.G. Cerf V.G., D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts, S. Wolff, *A Brief History of the Internet*, cit.

11 Cfr. J.C.R. Licklider & W. Clark, "On-Line Man Computer Communication", August 1962.

12 Cfr. L. Roberts & T. Merrill, "Toward a Cooperative Network of Time-Shared Computers", Fall AFIPS Conf., Oct. 1966; L. Roberts, "Multiple Computer Networks and Intercomputer Communication", ACM Gatlinburg Conf., October 1967.

13 L. Roberts & T. Merrill, "Toward a Cooperative Network of Time-Shared Computers", Fall AFIPS Conf., Oct. 1966.

La prima rete di computer usa un protocollo che si chiama NCP, e viene inaugurata nel settembre 1969. ARPANET è sia il medium attraverso il quale una generazione di ricercatori si scambia le informazioni relative al progetto di rete di calcolatori<sup>14</sup>, sia l'oggetto, teorico e pratico, della ricerca. Una ricerca che si configura come «learning by thinking»<sup>15</sup>, apprendimento tramite la sperimentazione e la messa in pratica dei principi elaborati dalla comunità di ricerca.

Uno strumento di comunicazione essenziale per la comunità che collabora al progetto sono le Request For Comments (RFC), una serie di note inaugurata nel 1969 come strumento attraverso il quale i ricercatori si scambiano in modo rapido e informale appunti, osservazioni, idee, che vengono discussi, ampliati, definiti e poi messi in pratica. I giovani scienziati fanno dei principi del libero pensiero, del metodo critico e della condivisione e cooperazione con i loro pari, una pratica organizzata di lavoro. Nate in principio su carta, le RFC circolano poi su ARPANET e tra le reti tramite il protocollo FTP (File Transport Protocol) e successivamente sul Web.

L'uso delle RFC scambiate pubblicamente è fondamentale per l'affermarsi dei protocolli, tutti rigorosamente pubblici. La definizione dei protocolli si basa infatti su una procedura fissa, fatta dei seguenti passaggi: La “richiesta di commenti” gira in rete e raccoglie le osservazioni degli altri, dei cui contributi si arricchisce. Una volta raggiunto il consenso della comunità sul suo contenuto, si produce un documento, in termini tecnici una “specificata”<sup>16</sup>, che poi è usata e testata dai gruppi di ricerca come base su cui implementare il software e, se si afferma all'interno della comunità, diventa un protocollo. La pratica delle RFC genera un effetto benefico, sia fornendo le regole della comunicazione alla comunità, sia ampliando il numero di coloro che collaborano alle innovazioni introdotte. Nel tempo, la produzione di RFC si è andata orientando sempre più sulla produzione di standard dei protocolli (“specifiche” ufficiali). L'accesso aperto (libero da copyright e gratuito) alle RFC favorisce la crescita di Internet perché consente alle specifiche di essere discusse, modificate e adottate sia nelle università e nei centri di ricerca, sia nelle imprese.

Ma l'elemento determinante per la nascita di Internet, forse il pilastro fondamentale della sua architettura, è lo “stack” (letteralmente “pila”) di protocolli di comunicazione che si chiamerà TCP/IP dal 1978, progettato a partire dal 1972 da Robert Kahn, anch'egli a DARPA. Kahn dapprima introduce in un progetto di rete radio l'idea di “open-architecture networking”, e crea un primo programma di nome “internetting”; di esso, lo scienziato decide poi di sviluppare una nuova versione che soddisfi le richieste di una rete ad architettura aperta, vale a dire:

- (1) che rispetti l'autonomia tecnica di ogni singola sotto-rete, senza comportare il bisogno di apportare modifiche per connettersi tramite il protocollo;
- (2) in cui la comunicazione avvenga sulla base della regola del *best effort* (migliore prestazione): un pacchetto perso viene ritrasmesso;
- (3) in cui per connettersi alla rete siano usati “black boxes” (gateway<sup>17</sup> e router<sup>18</sup>). I gateway non

---

14 Il Network Measurement Center all'UCLA viene selezionato come primo nodo ARPANET, il secondo nodo sarà quello di Stanford. Nel 1971 i nodi di ARPANET saranno diventati quindici, quasi tutti di centri di ricerca universitari. Il primo nodo italiano risale al 1985/86, presso CNR/CNUCE di Pisa.

15 Questa espressione è usata dal giurista Lawrence Lessig. Cfr. L. Lessig, *Cultura libera, Un equilibrio fra anarchia e controllo, contro l'estremismo della proprietà intellettuale*, Apogeo, 2005 pp. 35 e ss.

16 Col termine “specificata” in ingegneria del software si intende un accordo tra un produttore di servizi ed un utente. A seconda del contesto, si parla di specifiche dei requisiti (tra sviluppatore e committente), specifiche di progetto (tra progettista e sviluppatore, come nel caso delle specifiche dei protocolli), e di specifiche di modulo (tra i programmatori che hanno prodotto il modulo e quelli che lo integrano). Cfr. la relativa voce di Wikipedia: <<http://it.wikipedia.org/wiki/Specificata>>.

17 Dispositivi di rete il cui scopo principale è quello di veicolare i pacchetti di rete all'esterno della rete locale (LAN). Cfr. Wikipedia all'URL: <<http://it.wikipedia.org/wiki/Gateway>>.

18 Letteralmente “instradatore”, è un dispositivo di rete che si occupa di instradare pacchetti. Cfr. Wikipedia:

conservano informazioni sul passaggio dei pacchetti;  
(4) che non preveda alcun controllo globale o centrale sulle operazioni.

In pratica, il TCP viene progettato con le seguenti funzionalità:

- «a) Piccole sotto-sezioni con cui l'intera rete sarebbe stata capace di comunicare attraverso computer specializzati il cui compito era solo di indirizzare i pacchetti (gateway, router, switch,...).
- b) Nessuna parte della rete sarebbe stata singolo punto di rottura o sarebbe stata capace di controllare tutta la rete.
- c) Ogni pezzo dell'informazione inviata attraverso la rete avrebbe avuto una sequenza di numeri per assicurare che sarebbero stati inviati nell'ordine giusto alla destinazione e riconoscere la loro perdita.
- d) Un computer che inviava un'informazione ad un altro avrebbe saputo se questa fosse stata ricevuta nel momento in cui il computer di destinazione avesse inviato uno speciale pacchetto detto acknowledgement.
- e) Se l'informazione inviata da un altro computer andava persa, l'informazione veniva ritrasmessa dopo che la perdita veniva intercettata da un timeout che avrebbe identificato che l'atteso acknowledgement non era stato ricevuto.
- f) Ogni pezzo dell'informazione inviata attraverso la rete sarebbe stato accompagnato da un checksum, calcolato dall'originale computer sorgente e verificato dalla destinazione, per assicurare che non fosse rovinato.»<sup>19</sup>

Una volta definiti i principi<sup>20</sup>, nella primavera del 1973 è Kahn a chiedere a Vinton Cerf (al tempo alla Stanford University) di aiutarlo a scrivere l'implementazione dello stack di protocolli TCP/IP<sup>21</sup>, che viene adottato da ARPANET il primo gennaio 1983 e che diventa, nell'arco di circa dieci anni, il protocollo standard di comunicazione tra reti di computer, vale a dire l'insieme delle convenzioni tramite cui i computer si inviano i dati.

L'ultima tappa importante nella storia di Internet che ricordiamo qui è l'introduzione, sempre nel 1983, del DNS (Domain Name System), un sistema che consente di adottare un meccanismo scalabile e distribuito per tradurre gerarchie di nomi di host in un indirizzo internet. Il DNS consente di mappare numeri (indirizzi degli host) a nomi di dominio. A ogni dominio di nomi viene associata un'autorità di dominio. È questo l'unico elemento dell'architettura di Internet che implica una forma di centralizzazione e di accentramento.

Le agenzie federali, negli anni a seguire, danno vita a politiche decisive per la diffusione e lo sviluppo di Internet, dividendo e accollandosi i costi delle infrastrutture comuni, come i circuiti transoceanici, attività per coordinare la quale nasce il Federal Networking Council, che coopera con RARE in Europa. Gli stessi inventori di Internet scrivono due report importanti per il suo futuro sviluppo: "Toward a National Research Network"<sup>22</sup> (1988) commissionato dalla National Science Foundation, rapporto essenziale nell'influenzare Al Gore nel progetto delle autostrade informatiche e, nel 1994, "Realizing the Information Future: The Internet and Beyond"<sup>23</sup>.

Nel 1995 la FNC emana all'unanimità la seguente definizione:

«"Internet" refers to the global information system that -- (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately,

<<http://it.wikipedia.org/wiki/Router>>

19 Cit. da Wikipedia, voce italiana di Robert Kahn: <[http://it.wikipedia.org/wiki/Robert\\_Kahn](http://it.wikipedia.org/wiki/Robert_Kahn)>

20 Cfr. R. Kahn, "Communications Principles for Operating Systems". Internal BBN memorandum, Jan. 1972.

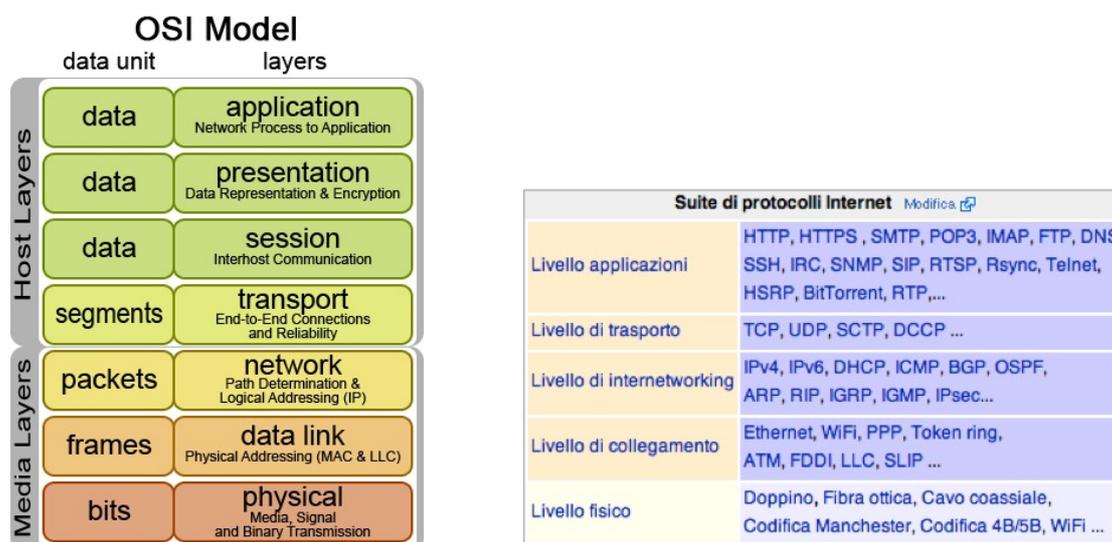
21 Cfr. V. G. Cerf and R. E. Kahn, "A protocol for packet network interconnection" IEEE Trans. Comm. Tech., vol. COM-22, V 5, pp. 627-641, May 1974. Poi perfezionato nel 1978, Darpa fa tre contratti a Stanford (Cerf), BBN e UCL per implementare TCP/IP, dai quali hanno inizio una sperimentazione e uno sviluppo di lungo periodo.

22 Online all'URL: <<http://www.nap.edu/openbook.php?isbn=NI000393>>.

23 Online all'URL: <[http://www.nap.edu/openbook.php?record\\_id=4755](http://www.nap.edu/openbook.php?record_id=4755)>.

high level services layered on the communications and related infrastructure described herein»<sup>24</sup>.

Si osservi che in tale definizione l'unica caratteristica tecnica menzionata è la suite di protocolli (TCP/IP). In essa è inoltre essenziale il concetto di strati, sia tra i protocolli, sia di servizi di alto livello. A partire dallo stack di protocolli TCP/IP, Internet si fonda dunque sul concetto di architettura multi-livello, come si vede nelle immagini che seguono.



In figura: Riproduzione del modello Operating System Infrastructure (standard stabilito nel 1978 dall'International Organization for Standardization che stabilisce una pila di protocolli in 7 livelli) e "pila" dei protocolli TCP/IP.

Il concetto di architettura a strati (*layered network*) è essenziale in un'architettura aperta in quanto consente di introdurre innovazioni senza dover apportare cambiamenti nell'architettura della rete. Il web, ad esempio, si trova al livello delle applicazioni (vedi figura sopra, a dx, il protocollo HTTP), e ha potuto diffondersi sopra e grazie a Internet senza richiedere alcuna modifica alla rete sottostante.

L'architettura di Internet è decentrata, distribuita e prevede una sovrabbondanza di funzioni nel network per minimizzare il rischio di sconnessione. Essa è inoltre aperta in quanto:

- 1) TCP/IP fissa requisiti minimi per il collegamento alla rete, fornendo gli standard compatibili per differenti sistemi di connessione alla rete. La flessibilità di TCP/IP permette di integrare differenti sistemi di connessione alla rete e accogliere altri protocolli (ragione per la quale alla fine si sono imposti come standard comuni).
- 2) Lo stack di protocolli è di dominio pubblico, vale a dire che i protocolli sono liberi da copyright. È questa una preconditione essenziale allo sviluppo della rete, poiché i protocolli e la loro documentazione tecnica hanno potuto essere consultati, discussi, migliorati, e implementati, e perché altri sviluppatori hanno potuto scrivere protocolli e programmi per costruire servizi e applicazioni sopra di esse. Ogni modifica viene comunicata alla rete, e gli utenti di Internet ne sono, dal principio in avanti, i produttori chiave. Senza tale condizione, i 15 nodi di ARPANET del 1971 non sarebbero diventati oggi centinaia di milioni<sup>25</sup>.

24 Cit. in Leiner B.M., Cerf V.G., Clark D.D., Kahn R.E., Kleinrock L., Lynch D.C., Postel J., Roberts L.G., Wolff S., *A Brief History of the Internet*, cit., in particolare nella sezione "History of the Future".

25 Per i dati, si veda questo grafico: <<http://www.totale.it/wp-content/uploads/dimensioni-di-internet.jpg>>. Si noti che

Nata in ambiente scientifico per lo scambio di documenti digitali, Internet si sviluppa senza frontiere in un processo in cui i governi hanno poca parte. Gli indirizzi non hanno una localizzazione fisica fissa, i costi per l'utente non dipendono dal contenuto trasmesso, non esiste un metodo per identificare gli utenti, né un controllo centrale della rete (il solo coordinamento centralizzato è, lo si è accennato, il sistema di indirizzamento).

Si può dunque riassumere che le caratteristiche fondanti di Internet sono due:

- a) in primo luogo, l'architettura della connessione in rete è aperta, decentrata, distribuita e multi-direzionale;
- b) in secondo luogo, tutti i protocolli di comunicazione e le loro implementazioni sono aperti, distribuiti e suscettibili di modifica (anche se non tutti sono modificabili)<sup>26</sup>.

L'infrastruttura fisica e lo stack di protocolli TCP/IP dunque le precondizioni per la nascita di ulteriori applicazioni che si servono di protocolli di livello superiore, tra i quali ricordiamo IMAP e POP per l'email (che nasce nel 1972), il protocollo UUCP di unix che consente la nascita di Usenet (1985), HTTP per il Web, e molti altri ancora.

## 2. Il world wide web

Il progetto del World Wide Web risale alla fine degli anni Ottanta. E' lo stesso Berners-Lee a narrare le circostanze della nascita del Web e a raccontare le tappe della sua evoluzione in un testo divulgativo pubblicato nel 1999 e tradotto due anni dopo in italiano<sup>27</sup>.

Laureatosi in fisica nel 1976 al Queen's College di Oxford, il giovane scienziato inglese si impiega come programmatore alla Plessey Telecommunications prima e poi alla D.G. Nash, presso cui costruisce il suo primo computer. La sua iniziale collaborazione con il CERN risale al 1980, anno in cui scrive *Enquire*, un programma «retiforme» che anticipa una caratteristica essenziale del World Wide Web, la possibilità di collegare pezzi di informazione in forma ipertestuale. Incaricato di occuparsi dei sistemi informativi del laboratorio franco-svizzero che ospitava diverse migliaia di ricercatori delle più differenti nazionalità (i quali lavoravano su numerosi progetti collegati ma indipendenti e che utilizzavano un'ampia gamma di programmi e di calcolatori incapaci di comunicare tra loro), Berners-Lee si scontra con le difficoltà derivanti dal dover gestire una tale mole di informazione: «Preparai Enquire nel tempo libero, per utilizzo personale e al nobile scopo di aiutarvi a ricordare i rapporti intercorrenti tra le varie persone, calcolatori e progetti all'interno del laboratorio»<sup>28</sup>. E continua: «Su Enquire potevo scrivere una pagina d'informazioni a proposito di una persona, una macchina o un programma. Ogni pagina era un "nodo" nel programma, una specie di scheda. L'unico modo per crearne uno nuovo era attuare un collegamento da un nodo già esistente. I link da e verso un nodo apparivano come un elenco numerato in fondo alla pagina, un po' come la lista delle citazioni alla fine di una pubblicazione accademica»<sup>29</sup>. Nella costruzione del suo organigramma, il fisico inglese scopre le potenzialità insite nella possibilità di inserire collegamenti in modo libero. «Per ogni link, potevo descrivere il tipo di rapporto. Per esempio, dichiarando se la relazione tra Joe e un programma

---

esistono eccezioni a questo principio. Alcuni protocolli di routing dello strato della rete e alcuni protocolli dello strato applicativo (per esempio skype) sono proprietari.

26 M. Castells, *Galassia Internet*, cit., p. 39.

27 T. Berners-Lee, *L'architettura del nuovo Web*, Feltrinelli, Milano 2001, scritto in collaborazione con Mark Fiaschetti (ed. or. *Weaving the Web, The original design and ultimate destiny of the World Wide Web, by its inventor*, Harper San Francisco, 1999). Si vedano anche la documentazione disponibile all'URL <<http://www.w3.org/History/>> e l'ipertesto, a cura dello stesso Berners-Lee: <<http://www.w3.org/DesignIssues>>.

28 Ivi, p. 18.

29 Ivi, p. 23.

era il fatto che lo usasse o che l'avesse scritto». Enquire funzionava in base a un principio molto semplice: bastava etichettare ogni pezzo d'informazione con un nome e poi ordinare al computer di trovarlo. Diventando possibile fare riferimento a ogni cosa allo stesso modo, si dava così vita a una rete di informazioni. Un'altra implicazione positiva stava nel fatto che i computer avrebbero potuto seguire e analizzare gli incerti rapporti di connessione che definiscono gran parte dei meccanismi della società, rivelando modalità del tutto inedite di vedere il mondo.

La natura e l'evoluzione del web sono correlate in maniera inestricabile, come spiega il suo stesso autore: «L'invenzione del World Wide Web ha comportato la crescente comprensione da parte mia del potere insito nel disporre le idee in maniera reticolare, una consapevolezza che ho appunto acquisito tramite questo genere di procedura»<sup>30</sup>. Enquire lo porta a concepire una visione fondata sulla crescita decentrata e organica di idee, tecnologia e società. Alla base del sistema stava infatti l'idea che i computer potessero diventare molto più potenti se posti in grado di mettere in connessione informazioni altrimenti scollegate. Potenzialmente, la ragnatela globale dell'informazione si presentava come un tutto collegato a tutto, nel quale i meccanismi della società diventavano simili a quelli in cui funziona il cervello.

La prima collaborazione al CERN dura pochi mesi. Nell'abbandonare il laboratorio, Berners-Lee lascia al suo successore il codice di *Enquire*, che andrà in seguito perduto. Tra il 1981 e il 1984 torna a vivere in Gran Bretagna, dove lavora come ingegnere informatico scrivendo software per stampanti. È in quel frangente che acquisisce dimestichezza con i linguaggi di mark-up, in particolare SGML<sup>31</sup>.

Dall'84 riprende la collaborazione con il CERN, grazie a una borsa di studio che gli consente di lavorare al “controllo e acquisizione dati”, presso il gruppo incaricato al rilevamento e all'elaborazione dei dati risultanti dagli esperimenti dei fisici delle alte energie. Lì scrive un secondo programma, *Tangle* (che significa letteralmente intrico, groviglio). «In *Tangle*, spiega ancora, se ricorreva una certa sequenza di caratteri, il programma creava un nodo che la rappresentasse. Quando essa ricompariva, invece di ripeterla *Tangle* attivava un semplice rimando al nodo principale. Man mano che altre frasi venivano immagazzinate come nodi, e altri puntatori le indicavano, si formava una serie di collegamenti. Il nocciolo della questione era che l'aspetto fondamentale sono le connessioni. Non sta nelle lettere, ma in come esse si uniscono a formare parole. Non sta nelle parole, ma come si uniscono per formare frasi. Non sta nelle frasi, ma come si uniscono in un documento. Immaginai di inserire in questa maniera un'enciclopedia, quindi porre una domanda a *Tangle*. La domanda sarebbe stata ridotta in tanti nodi, che poi avrebbero fatto riferimento ai punti in cui apparivano all'interno dell'enciclopedia. Il groviglio risultante avrebbe contenuto tutte le risposte relative»<sup>32</sup>. Il risultato dell'esperimento è fallimentare. Berners-Lee dismette il programma e tuttavia non abbandona il progetto, per il cui sviluppo l'ambiente del CERN si rivela particolarmente fecondo: la compresenza di migliaia di persone, che utilizzavano un numero imprecisabile di macchine su cui giravano i programmi più disparati e comunicavano grazie a una moltitudine di protocolli di rete, anticipava quella diversità interconnessa che di lì a pochi anni si sarebbe ritrovata anche nel mondo esterno. «Oltre a tener conto dei rapporti tra le varie persone, esperimenti e macchine, volevo facilitare l'accesso ai vari tipi di informazione, come gli scritti tecnici dei vari ricercatori, i manuali dei diversi moduli di software, i resoconti delle riunioni, gli appunti e così via. Inoltre, mi toccava

---

30 Ivi, pp. 16-7.

31 Un linguaggio di mark-up descrive i meccanismi di rappresentazione di un testo. Mark-up (letteralmente: marcatura) è un termine nato in ambiente tipografico per segnalare in forma standardizzata le parti del testo che necessitano una correzione. Tale tecnica richiede una serie di convenzioni, cioè un linguaggio a marcatori di documenti.

HTML, al pari di SGML, è un linguaggio di mark-up di tipo descrittivo, vale a dire che la scelta del tipo di rappresentazione da applicare al testo è lasciata al software. SGML (*Standard Generalized Markup Language*) è stato il primo linguaggio di mark-up descrittivo standardizzato a livello internazionale e ha avuto ampio utilizzo nella produzione di documentazione tecnica.

32 T. Berners-Lee, *L'architettura del nuovo Web*, cit, p. 26.

rispondere di continuo alle medesime domande poste da tante persone distinte. Sarebbe stato molto più semplice se tutti avessero potuto consultare il mio database.»<sup>33</sup> Sono probabilmente queste le ragioni per cui Berners-Lee persevera nella sua idea di creare un sistema di documentazione universale.

Il successore di *Tangle* è *Remote Procedure Call* (RPC), un programma scritto per facilitare la comunicazione tra i computer e le reti. Berners-Lee ha ben chiaro che costringere gli scienziati a riadattare il loro sistema di lavoro in base alle sue necessità lo porterebbe necessariamente all'insuccesso. Viceversa, avrebbe dovuto pensare a un sistema di documentazione che avrebbe permesso a ciascuno di conservare il proprio metodo organizzativo e i suoi programmi. «Dovevo creare, scrive, un sistema con regole comuni, accettabili per tutti, cioè il più possibile vicino alla mancanza di regole»<sup>34</sup>. RPC assolve al compito richiesto.

Il modello che sceglie per il sistema minimalista che ha in mente è l'ipertesto. Il termine, coniato nel 1965 da Ted Nelson, indica un formato nuovo, non lineare, scritto e pubblicato tramite macchine cosiddette "letterarie"<sup>35</sup>. Con Xanadu<sup>36</sup>, il progetto ipertestuale di Nelson, ogni informazione sarebbe stata pubblicata in forma ipertestuale. Ogni citazione sarebbe stata dotata di un link alla fonte, garantendo agli autori originali una piccola ricompensa ogni volta che la fonte venisse letta.

Berners-Lee immagina l'ipertesto come uno strumento per il lavoro di gruppo e collaborativo, in cui ciascuno è in grado di scrivere e di leggere in un reticolo di documenti. Il sistema che ha in mente avrebbe dovuto essere decentrato, in modo che ogni nodo fosse intrinsecamente equivalente agli altri e che fosse possibile collegarsi a un nodo (pagina) qualsiasi senza richiedere alcuna autorizzazione ad un'autorità centrale. È su tale base che progetta un sistema in grado di combinare i link esterni con l'ipertesto e con gli schemi di interconnessione sviluppati per RPC. «L'ipertesto sarebbe stato molto potente, aggiunge ancora, se fossimo riusciti a indirizzarlo verso il nulla più totale. Ogni nodo, documento o come preferite chiamarlo, sarebbe stato intrinsecamente equivalente agli altri. Tutti avrebbero posseduto un indirizzo di riferimento. Sarebbero esistiti insieme nel medesimo spazio, lo spazio dell'informazione»<sup>37</sup>.

Una tappa fondamentale per la definizione del suo progetto è la penetrazione del sistema operativo Unix e di Internet all'interno del CERN. In Europa, in quegli anni, l'interesse verso Internet era scarso; le istituzioni scientifiche e politiche europee stavano infatti tentando di progettare, con la scarsa lungimiranza di un eccessivo campanilismo, una propria rete alternativa. Ma l'esistenza di protocolli standardizzati è per lui l'occasione per dimostrare la possibilità di creare un ponte tra la pluralità di sistemi operativi e di reti esistenti.

La prima proposta ufficiale di un sistema informativo ipertestuale che il giovane collaboratore inoltra al CERN risale al marzo 1989; nonostante non riceva alcuna risposta, nel maggio 1990 Berners-Lee presenta una seconda proposta<sup>38</sup>, anch'essa ignorata. Tuttavia, ottiene l'appoggio informale del capo della sua divisione, che gli permette di acquistare una macchina NeXT sulla quale comincia a programmare il software del suo progetto, che chiama World Wide Web. Con il sostegno pratico di Robert Cailliau, un ingegnere impiegato al CERN e particolarmente sensibile al problema dell'interoperabilità tra hardware e software che si convince immediatamente della bontà dell'idea, decide di cercare un appoggio esterno al laboratorio di fisica, di nuovo senza successo.

Al 1990 risale la definizione dei tre principi e pilastri del Web: lo schema per definire gli indirizzi

---

33 Ivi, p. 27.

34 Ivi, p. 28.

35 Si vedano: T. Nelson, *A File Structure for the Complex, The Changing and the Indeterminate*, ACM 20th National Conference, pp. 84-100, 1965; Id, *Literary Machines*. Swarthmore, PA. 1981 (prima edizione) – 1987, 1990 e 1993, Mindful Press, Sausalito, CA. Tr. it. *Literary machines 90.1. Il progetto Xanadu*, Muzzio, Padova 1992.

36 Si veda la relativa voce di Wikipedia, all'URL <[http://it.wikipedia.org/wiki/Progetto\\_Xanadu](http://it.wikipedia.org/wiki/Progetto_Xanadu)>.

37 T. Berners-Lee, *L'architettura del nuovo Web*, cit, pp. 28-29.

38 La proposta originale è disponibile sul web all'URL: <<http://www.w3.org/History/1989/proposal.html>>

dei documenti (Universal Resource Identifier o URI<sup>39</sup>), il protocollo di trasmissione dei dati (Hypertext Transfer Protocol o HTTP) e il linguaggio di contrassegno che definisce la formattazione delle pagine contenenti link ipertestuali (Hypertext Mark-up Language o HTML): «Il segreto stava nel definire poche regole base, comuni, di 'protocollo', per permettere ai computer di dialogare tra di loro, di modo che quando tutti i computer di ogni luogo avessero fatto altrettanto, il sistema si sarebbe arricchito, invece di collassare. Per il Web tali elementi erano, in ordine decrescente di importanza, URI, HTTP e HTML, cioè gli identificatori universali, il protocollo di trasferimento e il linguaggio»<sup>40</sup>. Poiché i tentativi di trovare collaboratori disposti a sviluppare un browser, programma essenziale a rendere l'utilizzo del Web effettivo, naufragano, Berners-Lee decide di scriverlo da solo. Lo stesso accade per il primo server Web, il programma che conserva le pagine in una parte del computer accessibile dall'esterno.

Finalmente, registra il suo computer col nome info.cern.ch, mettendo a disposizione i suoi appunti, le specifiche di URI, HTTP e HTML e le informazioni sul progetto in corso<sup>41</sup>. Per dimostrarne gli usi possibili all'interno del CERN, si serve dell'elenco telefonico del laboratorio, che diviene così accessibile a tutti. «Quel che avevamo ottenuto fino a quel momento si basava su pochi principi chiave appresi con l'esperienza. Il punto cruciale era l'idea di universalità, la rivelazione che un solo spazio dell'informazione potesse inglobare tutto, regalandoci un potere e una coerenza inauditi. Da qui derivavano molte decisioni tecniche. [...] Era una rivoluzione copernicana rispetto alla filosofia dei precedenti sistemi informatici. La gente era abituata ad andare a cercare le informazioni, ma di rado faceva riferimento ad altri computer, e anche in tal caso doveva ripetere una lunga sequenza complessa di istruzioni per accedervi»<sup>42</sup>.

Viceversa, il nuovo sistema di documentazione avrebbe permesso di trovare l'informazione sfruttando i meccanismi mentali che tutti usiamo naturalmente per ricordare, basandosi su una modalità di lavoro consueta agli accademici e, più in generale, agli scienziati, l'uso delle citazioni: «La comunità dei ricercatori usa da sempre dei collegamenti del genere tra documenti cartacei: tavole dei contenuti, indici analitici, bibliografie e sezioni di consultazione e rimandi sono tutti quanti link ipertestuali. Però, sul web, i link ipertestuali possono essere seguiti in pochi secondi, invece che in settimane di telefonate e inoltri della posta. E d'un tratto gli scienziati possono sottrarsi all'organizzazione sequenziale di ogni pubblicazione e bibliografia, scegliendosi un percorso di riferimenti che faccia al caso loro»<sup>43</sup>.

### 3. Il Web, premesse e antefatti: dal mundaneum al memex

All'importanza del collegamento nei sistemi di documentazione e all'idea di ipertesto si erano già dedicati, prima di Ted Nelson e di Berners-Lee, scienziati di diverse discipline. Tra la fine dell'Ottocento e la prima metà del Novecento sono in particolare documentaristi e ingegneri ad affrontare, la questione della catalogazione e dell'accesso all'informazione. Tra questi vale la pena ricordare in particolare Paul Otlet e Vannevar Bush.

Nato a Bruxelles nel 1868 da un'agiata famiglia di avvocati, Paul Otlet intraprende nella prima giovinezza studi giuridici. Tuttavia, egli mostra da subito una vocazione agli studi speculativi, vocazione che viene incoraggiata da Edmond Picard e Paul Héger, la cui frequentazione nell'ambiente familiare lo spingerà ad abbracciare la filosofia positivista. Nel 1891 un gruppo di

---

39 Contrariamente alla volontà dello stesso inventore del Web, l'URI sarà rinominato URL (Unique Resource Locator). Sul dibattito in seno al W3C su questo punto, che ha una considerevole rilevanza filosofica, si veda: T. Berners-Lee, *L'architettura del nuovo Web*, cit, pp. 63-65;

40 T. Berners-Lee, *L'architettura del nuovo Web*, cit, p. 44.

41 Tutta la documentazione è accessibile nella sezione "Design Issues" sul sito del World Wide Web Consortium (W3C), all'URL: <<http://www.w3.org/DesignIssues>>

42 T. Berners-Lee, *L'architettura del nuovo Web*, cit, p. 42 e pp. 44-45.

43 Ivi, p. 45.

eruditi belgi fonda a Bruxelles la *Société des Etudes Sociales et Politiques*, che il giovane Otlet vede nascere con grande interesse. All'interno della *Société*, Henri La Fontaine dirige la Sezione di Bibliografia. L'amicizia e la collaborazione con La Fontaine, premio nobel per la pace nel 1913, ne incoraggiano la creazione di repertori bibliografici in materia giuridica<sup>44</sup> e, più in generale, la formulazione di un progetto innovativo che si propone la costituzione di centri di documentazione capaci di mettere a disposizione dell'umanità la grande varietà di documenti prodotti nelle diverse discipline. Il suo primo saggio di bibliografia (*Un Peu de Bibliographie*), in cui esprime apprezzamento per una bibliografia di matrice positivista e per il positivismo in generale, risale al 1892.

Convinto sostenitore della necessità che le scienze umane e sociali imitino le cosiddette scienze dure, occupandosi del problema del registro delle conoscenze umane, Otlet si occuperà in seguito della catalogazione delle fonti sociologiche. Con la creazione del catalogo, Otlet si persuade della necessità di una classificazione, cioè di un vocabolario controllato che consenta di descrivere con esattezza i diversi materiali.

Nel 1893, la Sezione di Bibliografia si trasforma nell'Istituto Internazionale di Bibliografia Sociologica (IIBS). La bibliografia viene intesa da Otlet e La Fontaine come una formula sintetica che permette di rimediare ai grandi mali che affliggono le scienze sociali e, più in generale, l'umanità. Entrambi coltivano infatti parallelamente due grandi passioni: la documentazione e il pacifismo<sup>45</sup>.

È su tali premesse che la creazione di una disciplina che permetta una comunicazione di tipo universale si colloca come parte fondamentale del lavoro quotidiano degli uomini di scienza. La scienza della documentazione, come enunciata da Otlet, nasce dunque allo scopo di studiare i problemi della ricerca e del suo avanzamento, nel piano generale della comunicazione scientifica applicata al progresso umano. A tal fine si rendono necessarie una scienza e una tecnica della documentazione.

Paul Otlet fonda così la teoria della documentazione come scienza che si occupa della classificazione dei dati relativi alla produzione, alla conservazione, alla circolazione e all'uso (o impatto) degli scritti e documenti di ogni specie. Il fine ultimo della sua impresa è unire l'umanità attraverso l'opera scientifica.

Otlet descrive la nascita di una nuova scienza che si scompone in un processo triadico: 1. la descrizione dei libri (biblioteconomia); 2. le leggi della loro produzione, conservazione e diffusione (bibliografia); 3. le applicazioni pratiche delle conoscenze bibliografiche (documentazione). Le fasi del processo si propongono di risolvere altrettanti problemi metodologici che riguardano le scienze bibliografiche: il registro, la moltiplicazione e la disponibilità dell'informazione.

La scienza della documentazione si configura come disciplina che si occupa dello studio del libro e del documento e che assume come sue parti biblioteconomia e bibliografia, che il documentarista teorizza e sistematizza nel *Traité de documentation*, il primo trattato scientifico di teoria della documentazione pubblicato a Bruxelles nel 1934.

Con una metafora efficace, Otlet definisce la sua nuova disciplina «metallurgia documentale»: un processo di estrazione, separazione e trattamento dei materiali intellettuali grezzi, che avviene tramite l'elaborazione e l'interscambio di conoscenze. Così, la documentazione viene intesa come corpo epistemologico che è al contempo oggetto di studio e strumento

---

44 Nel 1891, Paul Otlet pubblica, assieme a Pierre Blanchemerle, Joseph Cassiers e Max Hallet, *Le Sommaire périodique des revues de droit* (Il Sommario periodico delle riviste di diritto), tavole mensili di tutti gli articoli giuridici pubblicati nei mensili belgi. Per una bibliografia completa dell'opera di Otlet, a cura di W. Boyd Rayward, si veda:

<<https://www.ideals.uiuc.edu/html/2142/652/Paul%20Otlet%20Bibliography.htm>>.

45 La sintesi di tale duplice interesse viene pubblicata in due saggi: il *Traité de documentation. Le livre sur le livre. Théorie et pratique*, del 1934 e *Monde Essai d'universalisme* dell'anno successivo. Il primo è disponibile online all'indirizzo: <<https://archive.ugent.be/handle/1854/5612>>, il secondo all'URL: <<https://archive.ugent.be/handle/1854/8321>>.

metodologico.

Su un piano applicativo, Otlet e La Fontaine danno vita a un insieme di creazioni innovative destinate a orientare la scienza documentale dei decenni successivi: l'Istituto internazionale di bibliografia, centro dell'attività bibliografica europea; il Repertorio bibliografico universale, un catalogo di cataloghi che già nel 1897 contiene un milione e mezzo di registri bibliografici; la scheda bibliografica, un innovativo strumento documentale che ancora oggi costituisce uno standard della biblioteconomia; la Classificazione Decimale Universale, un adattamento della classificazione universale di Dewey di cui Otlet viene a conoscenza nel 1894. Elementi, questi, che danno origine al paradigma documentale otletiano.

Ma l'esperienza più importante e significativa dell'opera di Otlet è il Palais Mondial, che sfocia successivamente nel Mundaneum, una sorta di *universitas* universale, un tempio del sapere consacrato alla conoscenza scientifica. Il Mundaneum, che avrebbe ospitato una Biblioteca Internazionale, l'Istituto Internazionale di Bibliografia, il Repertorio Bibliografico Universale, una Enciclopedia di Documentazione, l'officina di Associazioni Internazionali e l'Università Internazionale, si configura come un luogo animato dai principi di totalità, simultaneità, gratuità, volontarietà, universalità, e da un ideale di vita internazionale, sostenuta dalla ragione e dedicata alla cooperazione e alla pace nel mondo. Un tempio consacrato all'intelletto, nel quale i ricercatori e gli intellettuali del mondo potessero avere la possibilità di studiare tutto quanto prodotto e pensato dagli uomini nel corso dei secoli.

Il progetto ottiene grandi riconoscimenti internazionali, ma uno scarso appoggio istituzionale. Otlet getta tuttavia le basi del lavoro documentale, indicando gli attrezzi del mestiere necessari al lavoro di ricerca. Diversi studi su Otlet lo hanno indicato come precursore dell'idea di ipertesto<sup>46</sup>. L'uso delle schede e dei cataloghi proposto dal padre della scienza documentale dà vita a un nuovo paradigma, in cui la scheda bibliografica corrisponde al nodo e i link che creano un sistema di navigazione ipertestuale sono dati dai cataloghi. Il sistema di classificazione proposto da Otlet si distingue come un nuovo dispositivo di accesso all'informazione e al documento, da molti considerato come l'inizio della rottura con il modello della struttura lineare. Il suo interesse si concentra sugli aspetti di un problema centrale per la società, le cui conseguenze sono tanto tecnico-organizzative quanto teoretiche: il problema di rendere la conoscenza facilmente accessibile e consultabile dall'umanità.

È questo un tema che lo avvicina a Vannevar Bush, che nel 1945 pubblica il famoso saggio *As we may think*<sup>47</sup>. «La conoscenza, scrive, per essere utile alla scienza deve poter essere continuamente ampliata, in primo luogo archiviata e soprattutto consultata». Preside della facoltà di ingegneria al MIT tra il 1932 e il 1938, in seguito Bush diventa direttore capo dell'Ufficio americano per la Ricerca e lo Sviluppo scientifico e consigliere scientifico del Presidente. Nello stesso anno in cui pubblica il saggio sopra citato, indirizza al presidente

---

46 Su questo argomento si vedano: W. Boyd Rayward, *Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext*, "Journal of the American Society of Information Science," Vol. 45, 1994, pp. 235-250, online all'URL: <<http://people.lis.uiuc.edu/~wrayward/otlet/xanadu.htm>>; P. Arnau Reved, *Trascendencia de la vida y obra de Paul Otlet*, online all'URL: <<http://www.ucm.es/BUCM/revistas/byd/11321873/articulos/RGID9595220153A.PDF>>; R. Day, *Paul Otlet's book and the writing of social space*, Journal of the American Society of Information Science, Vol. 48, n. 4, 1997, pp. 310-17; I. Rieusset-Lemarié, *P. Otlet's Mundaneum and the International Perspective in the History of Documentation and Information Science*, Journal of the American Society of Information Science, Vol. 48, n. 4, 1997 pp. 301-310; M. Buckland, *What is a Document?*, Journal of the American Society of Information Science, vol. 48, no. 9, 1997, pp. 804-809, online all'URL: <<http://people.ischool.berkeley.edu/~buckland/whatdoc.html>>; C. van den Heuvel, *Mundaneum. Architectures of Global Knowledge: The Mundaneum and the World Wide Web*, online all'URL: <<http://www.virtuallknowledgestudio.nl/staff/charles-vanden-heuvel/vdheuvel-mundaneum.pdf>>; J. M. Zurita Sánchez, *El paradigma otletiano como base de un modelo para la organización y difusión del conocimiento científico*, Tesi di laurea in biblioteconomia, Universidad nacional autónoma de México, 2001, online all'URL: <[http://etdindividuals.dlib.vt.edu:9090/358/1/paradigma\\_otletiano.pdf](http://etdindividuals.dlib.vt.edu:9090/358/1/paradigma_otletiano.pdf)>.

47 V. Bush, *As we may think*, "The Atlantic Monthly", vol. 176, no. 1, luglio 1945, pp. 101-108. Il testo in lingua originale: <<http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>>, tr. it. *Come possiamo pensare*, in T. Nelson, *Literary machines* 90.1, cit. pp. 1/38-53.

Roosevelt un rapporto<sup>48</sup> in cui, con sguardo lungimirante, sostiene l'importanza della ricerca di base; un documento che nei decenni successivi si rivelerà determinante nell'orientare la politica americana in materia di innovazione e sviluppo.

Se durante la seconda guerra mondiale Bush aveva messo la propria opera al servizio dello Stato e, al pari di altri scienziati tra cui vale la pena ricordare il matematico inglese Alan Turing, padre del moderno calcolatore e artefice della decrittazione del codice tedesco *Enigma*, aveva ricoperto un importante ruolo nella ricerca militare, il saggio del 1945 è una riflessione sul rapporto tra la tecnologia e il problema, particolarmente attuale in epoca post-bellica, della pace nel mondo. Come può la tecnica, si domanda l'ingegnere americano, contribuire al benessere dell'umanità? Bush risponde all'interrogativo sciogliendo, nel corso del saggio, un'altra questione che sta, a suo parere, a monte del problema: la questione riguarda il modo in cui il bagaglio di conoscenze umane possa creare le condizioni per la pace nel mondo. Perché le tecnologie possano mettere l'uomo in grado di costruire condizioni di vita migliori, una «casa ben arredata» in cui «vivere in buona salute»<sup>49</sup> suggerisce Bush con un'immagine dall'assonanza kantiana<sup>50</sup>, è necessaria un'indagine sul metodo della ricerca e sul processo di costruzione della scienza. Il saggio non affronta dunque un tema meramente tecnico: viceversa, l'argomentazione sui cui è imperniato è anche e soprattutto una riflessione filosofica e politica sulle modalità di produzione e di trasmissione del sapere.

Vannevar Bush fonda il proprio ragionamento su una importante premessa condivisa da buona parte della tradizione filosofica occidentale da Platone in poi<sup>51</sup>, vale a dire il fatto che la conoscenza umana è un insieme collegato in un tutto che, in quanto tale, ha una dimensione universale la quale trascende la singola vita individuale. Il sapere è tale in quanto frutto di un processo cumulativo, e si struttura, grazie alla collaborazione degli scienziati, in sistemi collegati che includono l'intero patrimonio delle conoscenze umane. Per questo, l'accesso all'informazione scientifica è una condizione necessaria alla stessa possibilità della scienza.

Se la diffusione della stampa ha permesso all'umanità di produrre enormi quantità di informazioni a un ritmo sempre crescente, l'accrescersi progressivo dei dati e delle conoscenze raccolti in libri, articoli, e testi di varia natura non implica la possibilità di condividere tale patrimonio; inoltre, la crescente specializzazione del sapere, prosegue Bush, confligge con l'esigenza di sistemi di comunicazione veloci ed efficienti.

Il principale ostacolo all'accesso risiede negli alti costi di riproduzione, ostacolo che può essere tuttavia facilmente superato con l'ausilio di tecnologie in grado di comprimere grandi quantità di libri in dispositivi di piccole dimensioni. La proposta di Bush individua nell'uso delle macchine una soluzione al problema dell'archiviazione; ma il problema della consultazione del sapere è assai più serio e di difficile soluzione. Esso infatti coinvolge l'intero processo grazie al quale l'uomo trae vantaggio dal suo bagaglio di conoscenze: la selezione. Bush riconosce che il cuore del problema è più profondo del semplice ritardo nella meccanizzazione delle biblioteche. Il processo di selezione è «come uno scalpello nelle mani di un ebanista», che dev'essere pertanto considerato e affrontato con riguardo al metodo della ricerca scientifica, cui l'ingegnere americano dedica ampio spazio nella parte centrale del saggio.

Il cuore del problema deve essere ricondotto all'artificialità dei sistemi di indicizzazione in uso negli archivi, che organizzano l'informazione gerarchicamente, in base a una struttura ad albero. Poiché le informazioni si trovano in un unico punto dell'archivio, a meno di duplicati, si devono avere regole per decidere quale cammino ci porterà a ciò che cerchiamo. Il limite di tale

---

48 V. Bush, *Science: The Endless Frontier. A Report to the President*, Washington, U.S. Government Printing Office, 1945, online all'url: <<http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>>

49 V. Bush, *Come possiamo pensare*, in T. Nelson, *Literary machines 90.1*, cit. p. 1/53.

50 Nell'*Architettura della ragion pura* Kant paragona l'edificio della conoscenza scientifica a una casa d'abitazione (I. Kant, *Critica della ragion pura* A 707- B 735, Utet, Torino 1967).

51 Si veda in particolare il mito dell'*anamnesis* nel *Menone*, 81c-d (Platone, *Menone*, Laterza, Roma-Bari, 2004, p. 34.)

struttura è che a un documento corrisponde un unico percorso. «La mente umana non funziona in questo modo. Essa opera per associazioni. Una volta che essa abbia un elemento a disposizione, salta istantaneamente all'elemento successivo suggerito, in base a un intreccio di piste registrate nelle cellule del cervello, dalla associazione dei pensieri.»<sup>52</sup> Bush suggerisce così di ripensare la selezione dell'informazione a partire dal modo in cui funziona la nostra mente, e indica nel *memex*<sup>53</sup>, una macchina fotoelettronica in grado di attuare e seguire riferimenti incrociati su microfilm servendosi di codice binario, fotocellule e fotografia istantanea, un esempio di meccanizzazione della selezione per associazione piuttosto che per indicizzazione. L'accesso virtualmente universale al patrimonio culturale dell'umanità e l'impiego delle macchine nella ricerca hanno, ai suoi occhi, un potenziale creativo dirompente. Se infatti non esistono sostituti meccanici per il pensiero creativo, ma pensiero creativo e pensiero essenzialmente ripetitivo sono molto diversi; tuttavia ogni qualvolta si combinano e si registrano fatti sulla base di processi logici stabiliti, l'aspetto creativo «riguarda solo la scelta dei dati e del processo da impiegare, e da quel momento in poi la manipolazione è di natura ripetitiva, e dunque adatta per essere affidata alle macchine»<sup>54</sup>.

Il *memex* è un dispositivo meccanico, una scrivania meccanizzata formata da schermi traslucidi inclinati, una tastiera, e gruppi di bottoni e di leve. Al suo interno sono archiviati gli strumenti di lavoro del ricercatore (i libri e gli articoli su cui lavora, come gli appunti che produce), proiettabili velocemente su schermi inclinati. Grazie al *Memex* è possibile consultare il proprio archivio tramite lo schema di indicizzazione tradizionale: basterà digitare il codice del libro desiderato, e la sua copertina sarà rapidamente proiettata su uno dei visori. Ma la caratteristica fondamentale del “memory extender” è che esso consente al ricercatore di collegare stabilmente due testi qualunque premendo un semplice tasto<sup>55</sup>. La valenza pratica della macchina ideata da Bush è descritta dall'autore come segue: «Il proprietario del memex, per esempio, è interessato alle origini e alle proprietà dell'arco. In particolare, sta ricercando il motivo, nelle battaglie delle Crociate, per cui l'arco corto turco fosse superiore all'arco lungo inglese. Ha a disposizione dozzine di libri e articoli pertinenti nel suo memex. Dapprima sfoglia un'enciclopedia, trova un articolo interessante ma non approfondito, lo lascia proiettato. Poi, in un resoconto storico, trova un altro argomento pertinente, e lo congiunge al precedente. Procedo in questo modo, costruendo un percorso di molti elementi. Di quando in quando inserisce un proprio commento, sia congiungendolo alla pista principale sia connettendolo come pista laterale a un elemento particolare. Quando diventa evidente che le proprietà elastiche dei materiali disponibili erano fondamentali per il rendimento dell'arco, egli devia su una pista laterale che lo porta attraverso libri di testo sull'elasticità e tavole di costanti fisiche. Inserisce una pagina di analisi scritta di suo pugno. In questo modo egli costruisce una pista della sua ricerca attraverso il labirinto del materiale che ha a disposizione»<sup>56</sup>.

Così, quando numerosi elementi sono stati congiunti per formare una pista, grazie al *memex* possono essere esaminati l'uno dopo l'altro, velocemente o lentamente, muovendo una leva come quella usata per girare le pagine del libro. È come se gli elementi, parti di libri preesistenti, venissero raccolti separatamente e quindi rilegati nuovamente assieme, in modo da formare un nuovo libro.

Affrontando il problema della selezione dell'informazione, Bush restituisce al ricercatore un ruolo attivo e fondante. «Nascerà la nuova professione di battitore di piste, persone che si

---

52 V. Bush, *Come possiamo pensare*, in T. Nelson, *Literary machines 90.1*, cit. p. 1/49.

53 Il nome deriva dalla contrazione dell'espressione “Memory extender”.

54 V. Bush, *Come possiamo pensare*, cit. p. 1/49.

55 A differenza dei link sul Web, i collegamenti immaginati da Bush sono bidirezionali; vale a dire che ogni volta che un elemento viene proiettato, un altro elemento, ad esso collegato, può essere istantaneamente richiamato premendo un bottone, e visualizzato.

56 V. Bush, *Come possiamo pensare*, cit. p. 1/51.

ingegneranno a tracciare percorsi significativi attraverso l'immane mole dell'esperienza umana. L'eredità del maestro ai suoi discepoli non saranno più solo i suoi contributi alla conoscenza comune, ma l'intera impalcatura di conoscenze sulla quale essi sono stati costruiti»<sup>57</sup>. L'attività della ricerca si traduce praticamente nell'inserimento di ogni singolo elemento in una o più piste; percorsi che restano così patrimonio dell'umanità, tanto quanto i libri, e danno vita a enciclopedie di concezione radicalmente nuova, dotate di una trama di piste associative che le attraversano. L'utilità insita nella possibilità di conservare traccia dei passaggi del processo di ricerca è esemplificata con riferimento alla ricerca sulle origini e le proprietà dell'arco, come segue: «Qualche anno dopo, durante un suo colloquio con un amico, il discorso cade sui modi strani in cui le persone resistono alle innovazioni, anche se di vitale interesse. Egli dispone di un esempio, ed esattamente di come gli europei, pur riconoscendo che la loro arma aveva una gittata minore, si rifiutarono di adottare l'arco turco: dispone addirittura di una pista al riguardo. Un tocco mostra l'elenco dei codici. La pressione di alcuni tasti porta all'inizio della pista. Una leva permette di scorrelerla a piacimento, fermandosi sui punti interessanti e dipartendosi in escursioni laterali. È una pista interessante, pertinente alla discussione. Così aziona un riproduttore, fotografa l'intera pista e la passa all'amico perché la inserisca nel proprio memex, congiungendola opportunamente ad altre sue piste»<sup>58</sup>.

#### 4. L'architettura del Web

Nel pensiero e nei progetti di Otlet e di Bush è possibile rinvenire gli elementi essenziali a caratterizzare la filosofia e i principi che stanno alla base del Web. Le implicazioni di un sistema di documentazione ipertestuale universale sono discusse dallo stesso Berners-Lee in una lezione tenuta in Giappone nel 2002<sup>59</sup>.

«Il concetto del Web ha integrato molti diversi sistemi informativi disparati, creando uno spazio immaginario astratto in cui le differenze tra essi fossero ininfluenti. Il Web doveva includere ogni sorta di informazione su qualunque sistema. L'unica idea comune necessaria a mettere ciò assieme fu identificare un documento tramite lo Universal Resource Identifier (URI). Direttamente da ciò sono come piovute architetture (*design*) di protocollo (come HTTP) e di formati di dati (come HTML) che hanno permesso ai computer di scambiarsi informazioni mappando i propri formati locali in standard in grado di offrire interoperabilità globale»<sup>60</sup>. In termini semplificati, Berners-Lee spiega che il Web non è un programma ma un insieme di protocolli che richiede a ciascuno di assegnare un nome unico ai propri documenti. Si tratta di una richiesta inaggirabile in quanto necessaria al suo funzionamento, ma ampia: è infatti l'unico limite che l'architettura del Web impone alla rappresentazione e alla organizzazione dei dati.

Nella sua lezione, Berners-Lee insiste sul fatto che il passato, il presente e soprattutto il futuro del Web dipendono dal modo in cui viene garantita e protetta l'universalità del sistema. Sul piano tecnico, ciò è possibile postulando la separazione tra contenuto dell'informazione e forma in cui essa viene veicolata da due punti di vista principali: in primo luogo, mantenendo valida la condizione che ha permesso la nascita della ragnatela ipertestuale su Internet, vale a dire la sua indipendenza dal sostrato materiale (*hardware*) e dal modo in cui le idee sono codificate e veicolate (*software*); in secondo luogo, definendo con maggior rigore i confini tra "documenti" e "dati", i primi dedicati alla lettura da parte degli uomini, i secondi interpretabili

---

57 Ivi, p. 1/52.

58 Ivi, p. 1/51.

59 T. Berners-Lee, "The World Wide Web - Past, present, future. Exploring Universality" (2002), online all'URL: <<http://www.w3.org/2002/04/Japan/Lecture.html>>.

60 Ivi. Le specifiche URI, HTTP e HTML formano cosiddetti "bus" che connettono i programmi che risiedono sul computer dell'utente (detti *client*) ad altri programmi installati su computer remoti (cosiddetti *server*) che forniscono le informazioni.

dalle macchine.

Nella conferenza giapponese, Berners-Lee ricorda l'esigenza che ha dato origine al World Wide Web, quella di mettere a disposizione degli ospiti del CERN un comune strumento che permettesse loro di condividere documenti in rete. Si è osservato come l'idea di Berners-Lee si fondasse su un principio filosofico semplice e, assieme, forte: creare un sistema di condivisione dei documenti "leggero", fondato sulla sottrazione, piuttosto che sull'addizione di requisiti. Così operando, il sistema pensato dallo scienziato inglese sarebbe stato facilmente condivisibile da diversi computer, sistemi operativi e programmi applicativi, e, in quanto tale, universale, cioè aperto a tutti coloro che volessero connettersi. Da un punto di vista pratico, l'universalità del World Wide Web è garantita dalle specifiche del W3C<sup>61</sup>, che raccomandano:

1. l'indipendenza dall'hardware e dal software, importanti nel prevenire tanto la frammentazione in molti ipertesti sconnessi e proprietari quanto il pericolo che il Web finisca sotto il controllo monopolistico di una data società o di un solo stato;
2. l'adozione di standard per la codifica dei caratteri che permettono la presenza sul Web alle diverse lingue del mondo;
3. la creazione e la diffusione di standard per l'accessibilità, che tutelano chi è svantaggiato in termini di capacità (di vedere, sentire, muoversi e capire) e chi dispone di macchine con tecnologie obsolete.

La riflessione del padre dell'ipertesto globale si sposta dunque su un *topos* della teoria informatica, la dicotomia tra cervello e macchina, tra "rima" e "ragione". Se il cervello umano funziona tramite un complesso sistema di associazioni, e possiede la capacità di effettuare collegamenti che rientrino in un sistema coerente e consistente di conoscenze, le macchine, viceversa, hanno enormi capacità di calcolo, e possono processare l'informazione in forma di tabelle, vale a dire secondo una struttura gerarchica.

Per disporre di strumenti informativi più ricchi, l'architettura del Web deve pertanto definire con precisione i confini tra i "documenti" (i contenuti accessibili agli uomini) e i "dati"; una reale separazione permetterà alle macchine di processare le informazioni, una possibilità resa oggi complicata dal fatto che la gran parte dell'informazione collegata in forma di ipertesto nella ragnatela condivisa "World Wide", contiene entrambi gli elementi mescolati<sup>62</sup>.

Berners-Lee ha cura di precisare che non intende sostenere la possibilità di un'intelligenza artificiale - ne è conferma il presupposto su cui si fonda il suo ragionamento, la netta differenza tra il funzionamento del cervello umano e della macchina. Egli vuole semplicemente sottolineare i notevoli vantaggi che la possibilità di trattare informazione semanticamente strutturata reca con sé, favorendo l'interoperabilità, la portabilità e la durabilità dell'informazione (il riuso della conoscenza), e aprendo nuove possibilità di selezione del sapere.

Il discorso di Berners-Lee affronta la questione della selezione del sapere secondo criteri di "qualità". Se «è noto che una raccolta di testi, come un insieme di report tecnici o una biblioteca, include soltanto quegli articoli che raggiungano un certo livello di qualità» e «alcuni ritengono l'assenza di simili sistemi una limite del Web», tuttavia, precisa, è importante che il Web in sé non tenti di promuovere una singola nozione di qualità, ma continui a raccogliere

---

61 Il W3C (acronimo di World Wide Web Consortium) è un organismo misto pubblico-privato con sede al MIT e nato allo scopo di promuovere e sviluppare tecnologie che garantiscano l'interoperabilità sul Web. Sul sito del W3C (<http://www.w3.org/>) si trovano le specifiche e le linee guida del consorzio, oltre a software e applicazioni specifiche.

62 La formattazione dei dati in HTML non permette che l'informazione sia processabile dalle macchine; è questo un limite strutturale del linguaggio, che non consente di separare pienamente il contenuto dalla sua struttura. Tuttavia, la nascita di nuovi formati di codifica (primo tra tutti XML, acronimo di Extensible Markup Language) e di nuove tecnologie di cui il W3C si è fatto promotore (che rientrano nel cosiddetto Web semantico) stanno favorendo l'accentuarsi di questa separazione. Cfr. su questo il paragrafo 6.

ogni informazione, vera o falsa che sia. È questa una caratteristica che può certamente rappresentare un limite. Tuttavia, un'autorità centrale che esercitasse un controllo sulla qualità sarebbe assai più dannosa; e se nessuno può essere in alcun modo obbligato a leggere letteratura di bassa qualità, è altresì vero che, come mostra la storia della scienza, appunti che oggi sono marginali potrebbero, un domani, essere a fondamento di nuove idee dalla portata rivoluzionaria. Pertanto, il problema della selezione dell'informazione secondo criteri di qualità dev'essere così riformulato: «Come possiamo fornire all'utente la percezione soggettiva di qualità elevata, e allo stesso tempo mantenere il Web aperto a persone i cui criteri di giudizio sono diversi?». La questione, risponde il suo inventore, dovrà essere affrontata dotando i lettori di strumenti di filtro in grado di sfruttare l'organizzazione semantica dell'informazione, che trasformeranno i ricercatori in veri e propri “battitori di piste” i quali, come nell'esempio di Bush, potranno scambiarsi e condividere i percorsi e crearne di nuovi, mantenendo comunque l'ipertesto globale svincolato da una autorità centrale che decida quale informazione è appropriata, e per chi.

## 5. Cenni di topologia delle reti

Prima di concludere questa introduzione alla filosofia del sistema universale di documentazione, è utile dare uno sguardo alla struttura di Internet e del Web da un punto di vista matematico. La conoscenza della topologia delle reti risulta infatti essenziale sia alla comprensione delle reti telematiche, sia all'uso dei metodi e degli strumenti, in primo luogo i motori di ricerca, che caratterizzano la professione del ricercatore nell'età delle reti.

Da un punto di vista topologico Internet e il web sono grafi, vale a dire insiemi di vertici uniti da archi. La teoria dei grafi è stata inaugurata da Eulero nel 1736<sup>63</sup> per risolvere un problema pratico che affliggeva gli abitanti di Königsberg. La città di Königsberg ha una geografia particolare: si trova alla confluenza di due fiumi, comprende un isolotto ed è divisa in quattro parti che al tempo di Eulero erano unite tramite sette ponti. Si narra che gli abitanti si divertissero a scommettere sulla possibilità di trovare un percorso che, partendo da una qualsiasi delle quattro zone della città, permettesse loro di attraversare ciascun ponte soltanto una volta, ritornando in fine al punto di partenza. Il matematico rappresentò ciascuna delle quattro zone della città con un cerchio (“vertice”), e indicando ogni ponte con una linea (“arco”) e fornì una dimostrazione matematica dell'impossibilità di trovare un tale percorso<sup>64</sup> dando così origine alla cosiddetta teoria dei grafi.

Il problema fu dunque risolto, e, per buona pace dei suoi abitanti, nel 1875 a Königsberg fu costruito un ottavo ponte.

---

63 Eulero, *Solutio problematis ad geometriam situs pertinentis*, 1736, testo online all'URL:

<<http://math.dartmouth.edu/~euler/docs/originals/E053.pdf>>

64 Eulero basò la sua dimostrazione sul fatto che i vertici con un numero dispari di spigoli debbono trovarsi al principio o al termine del percorso, e che un percorso che comincia in un punto e finisce in un altro non può avere più di due nodi siffatti. È facile osservare che il grafo di Königsberg ha quattro nodi con un numero dispari di collegamenti e che quindi tale percorso non esiste.

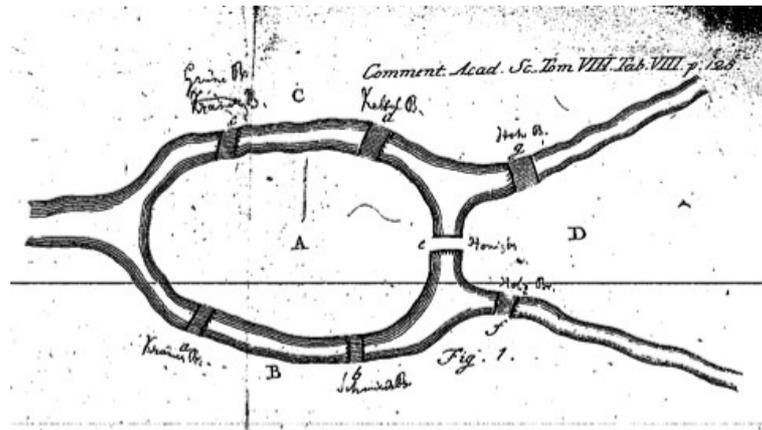


Illustrazione 1: Il disegno originale di Eulero

La teoria inaugurata da Eulero è oggi considerata il fondamento della attuale concezione delle reti. Vale a dire che «nella loro architettura, i grafi o le reti nascondono proprietà che possono limitare o favorire ciò che possiamo fare con loro»<sup>65</sup>.

In un saggio recente, il fisico rumeno Albert-László Barabási ripercorre le tappe essenziali nella storia della scienza delle reti, e definisce un particolare modello teorico, le reti a invarianza di scala, di cui mostra esempi significativi. Internet, il World Wide Web, la rete delle citazioni scientifiche, le presenze sul set degli attori di Hollywood e persino la diffusione dei virus hanno una topologia comune, ovvero sono riconducibili a quest'unico modello. Sul web, i vertici si chiamano “nodi”, e gli archi “link”. Si tratta di reti distribuite, dinamiche e in crescita, tenute insieme da una gerarchia di connettori, che formano una tela senza il ragno, vale a dire autorganizzata. Che cosa significa, esattamente? È possibile dare una risposta a questa domanda analizzando alcune caratteristiche delle reti anche senza entrare in complesse definizioni matematiche.

1. Il web è un grafo orientato, cioè una rete diretta i cui link sono monodirezionali, vale a dire che sono orientati in un'unica direzione. In pratica, ciò significa che nel momento in cui inserisco un link ipertestuale in una pagina web, io stabilisco un collegamento dalla mia pagina alla pagina che cito, in modo analogo a quanto avviene in un libro, in cui le citazioni vanno in un'unica direzione, indietro nel tempo.

2. Il web è un “mondo piccolo”, in cui il numero di passaggi necessari a raggiungere un punto qualunque del suo spazio è molto basso. La teoria dei mondi piccoli, nota anche come principio dei “Sei gradi di separazione”, è stata definita dal sociologo americano Stanley Milgram come esito di un esperimento condotto nel 1961. Milgram si propose di rispondere alla questione: qual è la distanza tra due cittadini qualsiasi degli Stati Uniti? In altri termini: quanti contatti (link) sono necessari per connettere tra loro due individui che non si conoscono scelti a caso? Selezionò due destinatari finali (la moglie di uno studente di teologia nel Massachusetts e un agente di cambio di Boston) e scelse casualmente alcuni abitanti di piccole città del Kansas e del Nebraska, a cui inviò una lettera in cui spiegò gli obiettivi dell'esperimento e le istruzioni da seguire<sup>66</sup>. Il risultato fu sorprendente: in media, negli Stati Uniti, il numero minimo di

65 A. L. Barabási, *Link. La scienza delle reti*, Einaudi, Torino 2004, p. 14.

66 Milgram inviò la lettera seguente:

“COME PRENDERE PARTE A QUESTO STUDIO

1) AGGIUNGETE IL VOSTRO NOME ALLA LISTA CHE TROVATE IN FONDO A QUESTO FOGLIO, affinché chi riceve per primo la lettera possa sapere da chi proviene.

2) STACCATE UNA CARTOLINA POSTALE, COMPILATELA E RISPEDITELA ALL'UNIVERSITA' DI HARVARD. L'affrancatura non è necessaria. La cartolina è molto importante: ci permetterà di seguire le tracce del documento nel suo viaggio verso il destinatario finale.

intermediari necessari per collegare due sconosciuti qualsiasi attraverso persone di loro conoscenza è circa sei. Vale a dire che, più in generale, la società è una rete di sei miliardi di nodi in cui la distanza media tra un nodo e l'altro non supera una manciata di link. È stato dimostrato che tale proprietà è scientificamente fondata e valida in diverse reti a invarianza di scala. Nel 2004, i gradi di separazione del web erano diciannove.

3. Il web è caratterizzato dalla presenza di connettori (*hub*). La sua architettura è dominata da pochissimi nodi altamente connessi (detti appunto “hub”) che tengono insieme molti nodi poco connessi. Anche tale caratteristica lo accomuna alla rete della società umana: «Disseminate in varie occupazioni, esiste una manciata di persone che possiede l'abilità davvero straordinaria di stringere un numero eccezionale di amicizie e conoscenze. Sono i connettori»<sup>67</sup>. I connettori sono presenti in tutte le grandi reti complesse.

4. Sul Web, il numero di link dà la misura di visibilità di un nodo.

5. Il Web è caratterizzato dalla legge 80/20, nota anche come principio di Pareto, e che è sintetizzabile nell'affermazione secondo cui, su grandi numeri, la maggior parte degli effetti (indicativamente l'80%) è dovuta a un numero ristretto di cause (indicativamente, il 20%). L'economista dimostrò infatti che in una determinata regione solo pochi individui possedevano la maggior parte della ricchezza. La legge è stata tuttavia verificata grazie a osservazioni empiriche di numerosi fenomeni. Essa è dunque applicabile sia al Web, in cui l'80% di link proviene dal 20% delle pagine, sia alla rete della comunicazione scientifica, in cui l'80% citazioni proviene dal 20% di scritti.

6. Il Web è una rete a invarianza di scala, i cui nodi sono distribuiti secondo leggi di potenza in base a due principi: il principio della crescita (sempre nuovi nodi vengono aggiunti progressivamente nel tempo) e il principio del collegamento preferenziale (è preferibile aggiungere collegamenti a pagine che hanno molti link, che sono cioè altamente connesse).

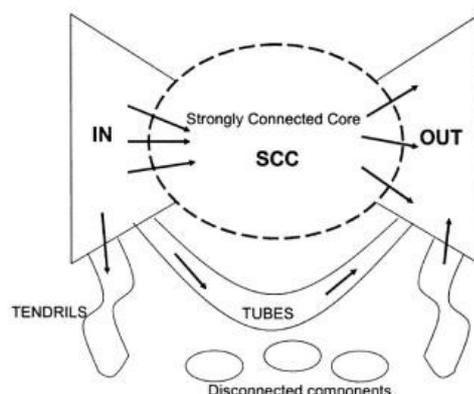
A causa di tali caratteristiche e proprietà, dunque, il Web non forma una singola rete omogenea ed è costituito da numerosi “sottomondi”, in modo analogo a quanto avviene per la rete delle citazioni scientifiche. Il suo spazio è diviso in quattro continenti di dimensioni pressoché equivalenti come rappresentato in figura.

---

3) SE CONOSCETE DI PERSONA IL DESTINATARIO FINALE, SPEDITEGLI/LE DIRETTAMENTE IL DOCUMENTO. Fatelo soltanto se lo avete già incontrato in precedenza, e se vi date del tu.

4) SE NON CONOSCETE DI PERSONA IL DESTINATARIO FINALE, NON CERCATE DI CONTATTARLO DIRETTAMENTE. SPEDITE INVECE QUESTO DOCUMENTO (COMPLETO DI CARTOLINA POSTALE) A UN VOSTRO CONOSCENTE CHE RITENETE ABBA MAGGIORI PROBABILITA' DI CONOSCERE IL DESTINATARIO FINALE. Potete spedirlo a un amico, a un parente o a un conoscente, ma dev'essere qualcuno a cui date del tu.” Cit. in A. L. Barabási, *Link. La scienza delle reti*, cit., pp. 31-32.

<sup>67</sup> M. Gladwell, *Il punto critico. I grandi effetti dei piccoli cambiamenti*, Rizzoli, Milano 2000, cit. in A. L. Barabási, *Link. La scienza delle reti*, cit., p. 61.



*Illustrazione 2: I "quattro continenti" del Web*

Il primo è il “corpo centrale”, comprende circa il 25% delle pagine, ed è facile da navigare perché molto interconnesso. Esso contiene i grandi connettori come i motori di ricerca (primo tra tutti google) e i grandi siti altamente linkati e visitati (per fare alcuni esempi: Amazon.com, ebay, youtube). Il secondo, detto “continente in”, è più difficile da navigare poiché consente di muoversi verso il “corpo centrale” ma non viceversa (come accade al pesce in una nassa). Allo stesso modo, il terzo continente (detto “continente out”) è raggiungibile dai nodi del corpo centrale, ma una volta usciti da tale zona non ci sono link per tornare indietro (esso contiene prevalentemente siti aziendali). Il quarto continente è fatto di tentacoli e isole separati, gruppi di pagine collegate tra loro ma non al “corpo centrale”.

Recenti studi hanno stimato che soltanto metà del Web è navigabile, cioè il “corpo centrale” e il “continente out”. Questa caratteristica è una sua proprietà topologica: vale a dire che la possibilità di trovare un percorso tra due nodi è indipendente dalle capacità dell'uomo e dei migliori motori di ricerca, esistenti o possibili (proprio come a Königsberg).

## **6. Sfide: verso la ragnatela semantica dei dati**

Sappiamo dunque che è possibile navigare soltanto in una frazione del Web, vale a dire che la maggior parte delle sue pagine può essere rintracciata soltanto conoscendone l'indirizzo (URL) e che solo una minima parte di esse è raggiungibile a partire da altri nodi della ragnatela. Ma come giungiamo a un sito che ci interessa? Lo strumento essenziale a reperire le informazioni contenute nelle pagine Web sono i motori di ricerca. Digitando una o più parole su google, il sito di google ci restituisce un elenco di link a pagine che contengono informazioni pertinenti ai fini della nostra ricerca. Tuttavia, non sempre le informazioni che troviamo corrispondono a quanto cerchiamo, viceversa, di solito la parte maggiore dei risultati è decisamente “fuori tema”. Infatti, se i motori di ricerca si sono dimostrati utilissimi nel setacciare rapidamente un numero molto elevato di pagine, non si rivelano altrettanto utili nel valutare la qualità del documento. Assieme a informazioni rilevanti, i motori ci restituiscono moltissimi link a pagine il cui contenuto è estraneo a quanto cerchiamo. Un problema che non dipende dalle caratteristiche topologiche della rete, ma da un suo limite architettonico.

In pratica, le difficoltà che incontrano i motori di ricerca dipendono dal fatto che questi si limitano a controllare la presenza nei documenti di determinati termini (parole chiave), un indizio che dice molto poco sul reale contenuto delle pagine che troviamo<sup>68</sup>. Perché dunque non

<sup>68</sup> T. Berners-Lee, *L'architettura del nuovo Web*, cit., p. 155.

si investe nel migliorare le tecniche e gli algoritmi di ricerca? In parte, lo si fa. Tuttavia, l'architettura originaria del Web presenta alcuni limiti che non dipendono da un difetto negli algoritmi di ricerca utilizzati dai motori. Il limite principale consiste piuttosto nel fatto che, sul Web, le diverse informazioni di un documento HTML sono mescolate al suo interno e non sono strutturate semanticamente, rendendo impossibile alle macchine trattare i dati sparsi nelle pagine secondo il significato che attribuiamo loro nel contesto in cui sono inseriti - mentre il significato dei dati all'interno del contesto è essenziale in qualsiasi ricerca.

Prendiamo ad esempio il caso in cui io decida di vendere la mia automobile tramite un annuncio in rete. L'annuncio che decido di postare in una pagina Web è più o meno del tipo "Vendo Toyota Yaris nera, ottimo stato, buon prezzo, Pisa, Italia". Tuttavia, se cerco "Yaris nera in vendita a Pisa, Italia", i motori di ricerca mi restituiranno un elenco sterminato di pagine che contengono quelle parole, ma il cui contenuto, nella maggior parte dei casi, sarà molto lontano da quello che desidero, informazione che viene definita *noise*, "chiasso", poiché distoglie la mia attenzione dai risultati che effettivamente mi interessano. Otterrò invece risultati migliori decidendo di inserire la mia offerta in un sito specializzato che contiene un modulo i cui campi di ricerca siano predefiniti e indichino, ad esempio, la casa automobilistica, il modello, l'anno di immatricolazione dell'auto e altri dati rilevanti ai fini della ricerca. Questo perché la pagina è scritta in un formato leggibile da una macchina, la quale sarà in grado di conservare il significato delle singole parti del documento.

Nel processo di ricerca non conta soltanto trovare la risposta a una domanda; ad assumere un rilievo fondamentale sono le strutture in cui è inserita la singola informazione. Il progetto originario del Web prevedeva in effetti che tali strutture fossero in qualche modo visibili e rintracciabili tanto dagli umani quanto dai computer<sup>69</sup>. Viceversa, sul Web come è stato implementato e come lo conosciamo, i link non sono "etichettati". Questa caratteristica ne indebolisce le potenzialità.

Dal 1999 il World Wide Web Consortium (W3C)<sup>70</sup>, il consorzio fondato e diretto da Tim Berners-Lee allo scopo di promuovere standard che assicurino l'interoperabilità del Web, ha concentrato le ricerche sui modi in cui risolvere il problema, dando vita a un orientamento noto come Web semantico<sup>71</sup>. Scopo di tale orientamento è progettare e sviluppare una estensione del Web che lo trasformi in una ragnatela di dati elaborabili dalle macchine. La scommessa fatta propria dal W3C è far diventare la rete in grado di interpretare le nostre richieste. Il web diviene semantico nel momento in cui, nella rete di collegamenti tra dati di natura differente ed espressi in forma diversa, i computer diventano in grado di trattare i dati in modo da inferire nuova conoscenza a partire da quella nota. Da un punto di vista tecnico, ciò è possibile strutturando l'informazione in modo tale che i documenti non restino "isole di dati" ma diventino "data base aperti" da cui un programma possa attingere informazioni.

Per comprendere il senso di tale affermazione, è utile introdurre la differenza tra "information retrieval" (recupero di informazione) e "data retrieval" (vale a dire: risposta automatizzata alle domande). Obiettivo della prima è produrre documenti che sono rilevanti per una query; questi documenti non devono essere unici, e da interrogazioni successive possono dare luogo a risultati completamente diversi. Obiettivo del secondo è invece produrre la risposta corretta a

---

69 Si osservi con attenzione la figura 1 nella prima proposta presentata da Berners-Lee al CERN (Information Management: A proposal, marzo 1989-maggio 1990) oggi on-line all'URL: <<http://www.w3.org/History/1989/proposal.html>>. Nella figura, gli archi che collegano le singole pagine (e che corrispondono ai link) sono contrassegnati da "etichette" che definiscono il tipo di collegamento che esiste tra due risorse. I tipi di collegamento possono essere diversi: ad esempio "Tim Berners-Lee" "ha scritto" "questo documento". "Questo documento" "descrive" "l'ipertesto". Le relazioni "ha scritto" e "descrive" si traducono banalmente in link. Ma sono link dal significato diverso; un significato che, nell'idea originaria di Web, veniva esplicitato.

70 Il sito del W3C è all'URL: <<http://www.w3.org>>.

71 Si veda in particolare questa sezione del sito del W3C: <<http://www.w3.org/2001/sw/>>.

una domanda. Il Web semantico si occupa di questo aspetto<sup>72</sup>. Ma che cosa significa Web semantico, in pratica? E come ci apparirà il nuovo Web? Una prima risposta si trova nell'esempio che segue:

«Quando il telefono squillò, il sistema di intrattenimento stava cantando a squarciagola “We Can Work It Out” dei Beatles. Nel momento in cui Pete alzò la cornetta, il suo telefono abbassò il volume mandando un messaggio a tutti i dispositivi *locali* con un *controllo del volume*. All'altro capo della linea c'era sua sorella, Lucy, dallo studio medico: “la mamma ha bisogno di una visita specialistica e poi dovrà eseguire una serie di trattamenti. Due volte a settimana, o giù di lì. Ora chiedo al mio agente di fissare gli appuntamenti”. Pete accettò subito di condividere l'impegno.

Dallo studio del medico, Lucy istruì il suo agente semantico attraverso il browser del suo palmare. L'agente di Lucy trovò immediatamente dall'agente del medico le informazioni relative al *trattamento prescritto*, controllò diverse liste di *offerte* e controllò quelle che *rientravano nel piano assicurativo* della madre *nel raggio di 20 miglia* da casa sua, e con un *rating* di fiducia *eccellente o molto buono*. Poi l'agente cominciò a provare a cercare gli *appuntamenti liberi* (offerti dagli agenti dei singoli provider tramite i loro siti Web) compatibili con le disponibilità di Lucy e di Pete (Le parole in corsivo indicano i termini la cui semantica, o significato, vengono definiti per gli agenti attraverso il Web semantico).

In pochi minuti, l'agente presentò loro un programma. Pete non lo apprezzò. Lo University Hospital si trovava dalla parte opposta della città rispetto alla casa della madre, e avrebbe dovuto tornare indietro nel caos del traffico dell'ora di punta. Programmò il suo agente in modo da rifare la ricerca con preferenza più strette su *luogo e orario*. L'agente di Lucy, avendo completa fiducia in quello di Pete riguardo a questa specifica ricerca, fornì a quello assistenza immediata fornendogli l'accesso ai certificati e agli shortcut dei dati che aveva trovato.

Quasi istantaneamente fu presentato un nuovo programma: la clinica era molto più vicina e gli orari precedenti, ma c'erano due avvertimenti. In primo luogo, Pete avrebbe dovuto spostare due appuntamenti poco importanti. Controllò di che cosa si trattava, e vide che non era un problema. L'altro riguardava il fatto che la clinica non rientrava nell'elenco di *terapeuti* coperti dalla compagnia di assicurazione: l'agente lo rassicurò “Servizio coperto e piano assicurativo verificato sicuro da altri mezzi” “(Dettagli?)”

Lucy si mise assente nello stesso istante in cui Pete brontolava: “risparmiami i dettagli” e tutto fu risolto<sup>73</sup>.

Abbiamo osservato come, da un punto di vista tecnico, l'architettura originale del Web si basasse esclusivamente su tre principi semplici: URL (la possibilità di puntare a una risorsa dandole un nome dal significato univoco), HTTP (il protocollo di trasporto delle pagine sul Web) e HTML (il linguaggio di codifica delle pagine Web, che permette di inserire link). Per far sì che i computer possano essere in grado di eseguire compiti come quello descritto, si sono rese necessarie alcune trasformazioni nell'architettura del Web. Che cosa significa strutturare semanticamente l'informazione? E come si può farlo?

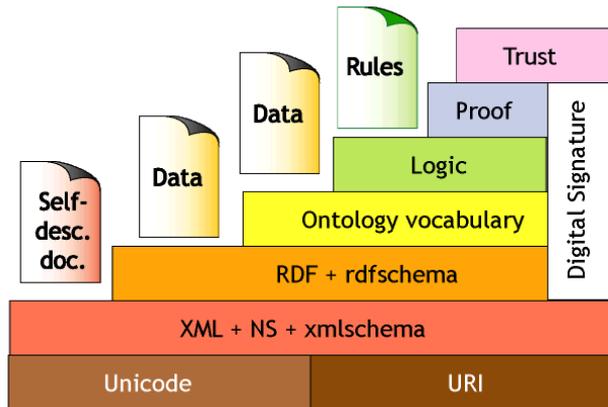
È possibile costruire una risposta a entrambe le domande considerando da vicino le trasformazioni tecnologiche sulle quali si fonda il Web semantico, oggi detto anche Web dei dati. Da un punto di vista architettonico, il Web semantico è stato rappresentato nel 2001 da Tim Berners-Lee<sup>74</sup> come una piramide di sette strati e composta da nove elementi, una piramide che “sorregge” tre tipologie di informazioni (“documenti auto-descrittivi”, “dati” e “regole”).

72 T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., p. 18.

73 T. Berners-Lee, J. Hendler, O. Lassila, *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American Magazine, 17 maggio 2001, online all'URL:

<<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>>

74 La slide è disponibile on-line all'URL: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>>.



Il significato della nuova architettura dev'essere spiegato a partire dalla base della piramide. Gli elementi in basso sorreggono gli strati più alti, e ne sono una preconditione. Infatti, il rapporto tra un piano della piramide e quello successivo può essere interpretato come un rapporto di condizionamento inteso in senso kantiano, vale a dire che il web può essere concepito come un insieme di strati con standard, linguaggi o protocolli che agiscono come piattaforme sulle quali possono poggiarsi formalismi nuovi, più ricchi e più espressivi. Piattaforme che sono intese più neutrali possibile, in cui l'architettura a strati ha funzione di regolamentazione e non è prescrittiva<sup>75</sup>. Un'analogia che può aiutare a comprendere il tipo di rapporto che esiste tra i sette strati della “torta semantica” è quella con il funzionamento di un aeroporto, dove la partenza e l'arrivo di ogni aereo dipendono dalla corretta concatenazione ed esecuzione di una serie di procedure standard (dal check-in dei passeggeri al loro sbarco, dalla spedizione alla riconsegna dei bagagli, dalle fasi di decollo e di atterraggio, seguite dai piloti e monitorate dalle torri di controllo, etc...).

E' inoltre importante osservare che, nel passaggio dagli strati inferiori a quelli superiori (in particolare gli ultimi due, “Proof” e “Trust”), le innovazioni che sono introdotte sono poco definite sul piano tecnologico. Ciò che viene invece illustrato è piuttosto l'impatto socio-culturale che le trasformazioni in atto nell'architettura del Web saranno in grado di produrre nelle comunità che operano sul Web, un aspetto di particolare interesse ai fini del nostro discorso.

Un'analisi dettagliata dell'architettura del Web semantico esula dallo scopo di questo contributo. Qui ci limitiamo a riassumere gli aspetti essenziali delle innovazioni tecnologiche proposte, che possiamo concentrare in quattro punti:

a ) Metadati (“dati sui dati”): servono a descrivere o annotare una risorsa in modo da renderla maggiormente comprensibile agli utenti. Normalmente i metadati sono di natura descrittiva e includono informazioni relative a un documento tra cui l'autore, il titolo o l'abstract, il tipo di file, i diritti d'accesso o il numero di versione. L'aggiunta di dati sui dati è utile per organizzare le risorse, per l'archiviazione e per identificare l'informazione. Tuttavia, la funzione più importante dei metadati è promuovere l'interoperabilità, consentendo la combinazione di risorse eterogenee tra piattaforme senza perdere informazioni rilevanti<sup>76</sup>.

<sup>75</sup> Gli studiosi di Web science sostengono che il rapporto tra gli strati si possa definire come *supervenienza*, termine che in filosofia indica un modo per spiegare la generazione di significati per cui un discorso A è superveniente a un discorso B se un cambiamento in A comporta un cambiamento in B, ma non viceversa. T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., p. 72.

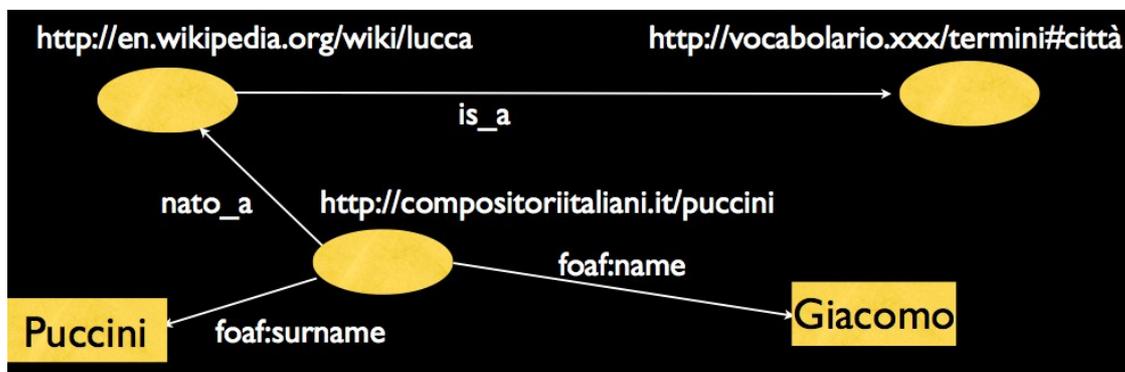
<sup>76</sup> Schemi di metadati sono ad esempio quelli Dublin Core (<http://dublincore.org/>), un sistema costituito da un nucleo di elementi essenziali ai fini della descrizione di qualsiasi materiale digitale accessibile tramite la rete. Per l'uso di chi naviga sul Web, i metadati possono essere non-strutturati; viceversa, per essere comprensibili alle macchine è necessario che tali informazioni siano strutturate. RDF offre meccanismi per integrare schemi di metadati. Cfr. T. Berners-Lee, W. Hall, J.A.

Si parla infatti di “open data” per definire un'importante caratteristica del web semantico, cioè la disponibilità dei dati e la conseguente possibilità di identificarli e citarli. Il Web semantico è perciò un'estensione del Web tradizionale nel senso che è il successivo passaggio nel linking e pensato per funzionare nel contesto di un modello relazionale di dati, in cui il link, da collegamento generico e cieco tra due documenti, diviene capace di esprimere relazioni concettuali, che convogliano significati. In pratica, sul web ogni “dato”, pezzetto di informazione, viene identificato da un URI. Perciò, si usa dire che sul web semantico, tutto è un URI.

b) Gli URI (identificatori di dati) vengono messi in relazione tra loro tramite il linguaggio Resource Description Framework (RDF), che consente di costruire asserzioni tramite triple formate da soggetto predicato e oggetto e di collegare tra loro le triple in un unico grafo (in cui soggetto e oggetto sono nodi, URI, e il predicato è il link). Si vedano le figura sotto, per un esempio. La prima mostra come si costruiscono le triple.



La seconda mostra come si costruiscono grafi con RDF.



c) Per esplicitare in modo formale un determinato campo di conoscenze, è necessario un ulteriore passaggio che permetta di associare i concetti a regole logiche d'uso. Questo compito è assegnato alle ontologie. In filosofia, un'ontologia è una teoria sulla natura dell'essere di tipi di oggetti; in senso più tecnico, è un accordo che si basa su definizioni condivise e che rappresenta la base concettuale su cui operiamo. Il termine è diventato di uso comune tra i ricercatori che si occupano di intelligenza artificiale e del Web, e in questo significato l'ontologia è un documento che definisce in modo formale le relazioni tra termini<sup>77</sup>. Nel grafo della figura sopra è già

Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit, pp. 36-39.

<sup>77</sup> Nel campo delle “computational ontologies”, si parla di ontologie con due accezioni diverse, entrambe corrette. La prima accezione intende con ontologia l'insieme sia dei dati, sia dei vocabolari; nella seconda accezione, l'ontologia è l'insieme dei

presente un'ontologia. Si tratta del link “is\_a” tra “lucca” e “città”. “is\_a” è in effetti una relazione che fa parte dei mattoni costituenti per creare un'ontologia. Relazioni come “is\_a” sono in un certo senso predefinite: sono gli assiomi del linguaggio, i connettori logici. La tipica ontologia per il Web ha una tassonomia e un insieme di regole di inferenza. La tassonomia definisce classi di oggetti e le relazioni tra essi. Così, il significato dei termini in una pagina web può essere stabilito da puntatori che linkano a un'ontologia. Un agente semantico che cerchi l'ospedale di Pisa e trovi “Santa Chiara” deve poter capire che non cerchiamo una biografia di Santa Chiara, né una chiesa o una Chiara qualunque. L'ontologia è un albero che permette di esprimere restrizioni sui termini (condizioni), quindi relazioni. Attraverso le relazioni della logica descrittiva si può esprimere in maniera formale qualsiasi oggetto o concetto. Le ontologie dunque contengono le specifiche dei concetti necessari a comprendere un dominio di conoscenza, il vocabolario corrispondente, e il modo in cui concetti e vocabolari sono collegati e in cui sono definite e descritte le classi, le istanze e le loro proprietà<sup>78</sup>.

Un'ontologia può essere formale o informale. Il vantaggio di un'ontologia formalizzata risiede nel fatto che questa è machine-readable, vale a dire che un computer può compiere dei ragionamenti a partire da essa. Di contro, lo svantaggio sta nella difficoltà di implementare tali costrutti formali.

È intuitivo osservare che due basi di dati possono usare identificatori differenti per il medesimo concetto. Un programma che voglia confrontare o combinare informazioni nei due database deve sapere che più termini possono riferirsi al medesimo oggetto. Idealmente, il programma deve essere in grado di riconoscere due termini come sinonimi. La soluzione a questo problema è data dalle altre ontologie<sup>79</sup> – le quali debbono mettere in correlazione i dati usando una lingua franca in modo da facilitarne la condivisione. Si veda nella figura sotto un esempio..

**Web dei dati**

ArteAlucca.com			
Opera	Titolo	Luogo	Città
9328	Ilaria del carretto	Duomo di San Martino	Lucca

chieseditalia.com			
ID	Nome	Località	Data nascita
2435	Duomo di San Martino	Lucca	22/12/1858

chieseditalia.com			
ID	Nome	Località	Data nascita
2435	Duomo di San Martino	Lucca	22/12/1858

geografiitalia.com			
CAP	Nome	Long	Lat
55100	Lucca	3249784	34322

compositori.com			
ID	Nome	Luogo Nascita	Data nascita
45	Giacomo Puccini	Lucca	22/12/1858

Sul piano dell'implementazione, l'ontologia esistente più diffusa è Wordnet<sup>80</sup>, che descrive i concetti (sinonimi, contrari, relazioni tra concetti) in diverse lingue. La peculiarità di tale sistema

solli vocabolari. In logica descrittiva si fa distinzione tra Abox e Tbox come due componenti distinti delle ontologie. La Tbox definisce un vocabolario e le relazioni tra gli elementi del vocabolario, la Abox è un insieme di asserzioni compatibili con quanto definito dalla Tbox. Qui intendiamo l'ontologia nella seconda accezione.

78 T. Gruber, “A translation approach to formal ontologies,” Knowledge Acquisition, vol. 5, no. 25, pp. 199–200, <http://ksl-web.stanford.edu/KSL Abstracts/KSL-92-71.html>, 1993.

79 Le ontologie sono scritte in RDF Schema, OWL (Ontology Web Language), e in altri linguaggi appositi. Come ad esempio DAML, DAMOIL.

80 Cfr. paragrafo 8.

è che è in grado di effettuare ricerche di “dipendenza concettuale”, come ad esempio le relazioni meronimiche (parte-tutto), sineddotiche o metonimiche (per esempio, se cerco un ospedale nella città di Lucca, è essenziale che il sistema sappia in qualche modo che Lucca è in Toscana e che la Toscana si trova in Italia).

Si osservi che chiunque può creare un propria ontologia di riferimento (sperando che venga condivisa) o adottarne una esistente, optando così per una altrui prospettiva. In genere, le ontologie sono create all'interno di comunità di conoscenza. Un problema, riscontrato e segnalato in particolare da filosofi e sociologi, è che ciascuno definisce il mondo a suo piacimento, e che proprio la creazione di queste definizioni è parte integrante del mestiere del ricercatore in scienze umane e sociali. Dunque, esiste un alto rischio di creare sistemi che si sovrappongono o che sono in conflitto, mentre la ricerca di un'unica ontologia, se da una parte sarebbe auspicabile, può comportare il pericolo insito nella rigidità delle interpretazioni di fatti e fenomeni.

Tuttavia, tale limite o pericolo può diventare una ricchezza. E' infatti auspicabile che vengano create molte e diverse ontologie, poiché più sistemi di classificazione sono in grado di stabilire una corrispondenza tra nomi e oggetti meglio rispetto a un unico vocabolario; inoltre, è possibile creare meta-ontologie che stabiliscano corrispondenze tra sinonimi, e che mettano a loro volta in relazioni vocabolari distinti. Ontologie diverse possono essere intersecate in modo da stabilire analogie, vale a dire relazioni di equivalenza. Spesso, gruppi diversi elaborano indipendentemente concetti molto simili e descrivere la relazione che esiste tra questi porta grandi benefici. Per un software, potrà essere dunque normale avere a che fare con ontologie differenti e magari in parte conflittuali nello stesso dominio di conoscenza, e sarà un compito del programma stesso mostrarne le diversità, aiutando così a migliorare la comprensione del concetto in questione.

Un'altra garanzia di pluralità, per la quale i teorici del web semantico si distinguono dalle teorie dell'intelligenza artificiale, è data dalle implicazioni della cosiddetta “Open World Assumption”, secondo la quale il valore di verità di un'asserzione è indipendente dalle conoscenze dell'osservatore. In altre parole, questo significa che se un osservatore non sa se un'asserzione è vera non può inferire che essa sia falsa (come invece accadrebbe invece in un mondo in cui vale la “Closed World Assumption”).

L'unione dei dati espressi in RDF e dei dati espressi in un linguaggio delle ontologie (ad esempio OWL) permette di inferire nuova conoscenza. È questo lo strato della logica, in cui trova spazio il “reasoning”, l'inferenza di nuova conoscenza: per costruire procedure analoghe al ragionamento software appositi devono essere in grado di collegare i termini, una funzione che sarà resa possibile tramite i linguaggi di inferenza. I linguaggi di inferenza consentono alle macchine di convertire dati da un formato all'altro, riconoscendo due termini come identici e traducendoli, un po' come fa un dizionario bilingue. Si tratta di una funzione fondamentale poiché nessuno ha il potere di definire un termine per tutti; questi linguaggi si occupano pertanto di identificare relazioni tra basi di dati per stabilire la presenza di sinonimi. Per tornare all'esempio di Lucy e Pete alla ricerca di un medico per il trattamento prescritto alla madre, grazie ai linguaggi di inferenza potrà essere trovato l'ospedale giusto anche cercando “clinica” o “hospital”, e via dicendo. Questa specifica funzione è assegnata a motori di ricerca logici, motori in grado di applicare le regole della logica per stabilire se le risposte ottenute in una ricerca iniziale sono utili o no. Nell'esempio, è un motore che riesca a rintracciare le informazioni relative alla loro specifica richiesta e a combinare i criteri da loro definiti: il trattamento richiesto e la copertura assicurativa dell'assistita, la distanza dalle rispettive abitazioni, gli orari in cui si sono resi disponibili, e altre informazioni rilevanti.

d) Infine, un concetto essenziale è quello di fiducia (*Trust*).

«La rete di fiducia è un modello essenziale del modo in cui lavoriamo realmente come persone. ognuno di noi costruisce la sua rete sin dall'infanzia. Man mano che decidiamo cosa linkare, cosa leggere o comprare sul web, un elemento che entra a far parte della nostra decisione è quanto possiamo fidarci dell'informazione che vediamo. Potremo fidarci del nome del suo editore, delle pratiche di tutela della privacy, delle motivazioni politiche? Certe volte impariamo nel modo peggiore che non dobbiamo fidarci, ma più spesso ereditiamo la fiducia da altri, da un amico o da un insegnante o da un familiare, o da raccomandazioni edite oppure da garanzie di terzi come la banca o il dottore». <sup>81</sup>

La costruzione di reti di fiducia è necessaria a implementare sistemi di filtro dell'informazione secondo criteri di qualità condivisi ma soggettivi, ed è uno dei temi su cui discute da anni il W3C <sup>82</sup>. In particolare, sul Web il concetto di fiducia assume notevole importanza da un punto di vista socio-culturale, sul piano cioè della trasformazione dei comportamenti dei navigatori della rete. Perciò, le implicazioni di questo concetto saranno approfondite nell'ultimo paragrafo.

La rete semantica è dunque in grado di descrivere l'informazione, poi di dedurre nuova conoscenza e infine di ragionare a partire da essa. E' questo, in sintesi, il sistema necessario a far funzionare i nostri agenti semantici nella storia di Lucy e Pete.

«Quando migliaia di moduli saranno collegati in tutto il campo “cognomi”, allora tutti quelli che analizzeranno il Web capiranno che è un importante concetto comune. La cosa bella è che nessuno dovrà compiere realmente questa analisi. Il concetto di “cognome” comincerà semplicemente a emergere come caratteristica importante di una persona. Come un bambino che impara un'idea tramite contatti ripetuti, la Rete Semantica “impara” un concetto tramite contributi ripetuti da diverse fonti indipendenti. [...] Il ragionamento che sta dietro questo approccio, quindi, è che non esiste un magazzino centrale dell'informazione, e nessuna autorità su alcunché. Collegando le cose tra di loro potremo fare molta strada verso la creazione di una comprensione comune. La Rete Semantica funzionerà quando ci saremo messi d'accordo sui termini, ma anche se non ci saremo riusciti.» <sup>83</sup>

La pretesa della rete semantica non è dunque quella di poter rappresentare tutti i dati o il sapere in qualche ristretto insieme di formalismi, ma piuttosto fare sì che la possibilità di linkare i dati a nuovi dati permetta di usarli in modo sempre più ampio. L'ambizione del W3C consiste nell'incrementare i dati a disposizione, e nel valorizzarli tramite nuove tecnologie che si aggiungono ai pilastri del Web. L'estensione del livello di inferenza che può essere ottenuta automaticamente non è dunque uno scopo ma semmai una conseguenza auspicabile <sup>84</sup>.

Questa evoluzione facilita la creazione di gruppi con interessi comuni, vere e proprie comunità aperte in rete. Nonostante le applicazioni del Web Semantico siano per il momento confinate per lo più all'ambito accademico, le loro possibili implicazioni cominciano a diventare evidenti. Restando nei confini della ricerca accademica, non è difficile vedere le implicazioni di tale rivoluzione non solo nella scienza del Web, ma anche nella pratica della ricerca dell'informazione in generale. Grazie alla costruzione di un Web ipertestuale, un gruppo di qualsiasi dimensione può comunicare con facilità, acquisire sapere e veicolarlo velocemente, superare le incomprensioni e ridurre la ridondanza degli sforzi. «L'universalità, scrive ancora Tim Berners-Lee, deve esistere in molte dimensioni. Tanto per cominciare, dobbiamo essere in

---

<sup>81</sup> T. Berners-Lee, *L'architettura del nuovo Web*, cit., pp. 137-38.

<sup>82</sup> Ivi, pp. 134-37.

<sup>83</sup> Ivi, p. 163.

<sup>84</sup> Ivi, p. 30.

grado di collegare tra loro molti documenti, dalle bozze alle successive versioni fino ai documenti finiti»<sup>85</sup>. Un passaggio essenziale a non perdere i passaggi del processo di un ragionamento, perché quando facciamo ricerca il processo tramite cui si raggiunge un risultato è importante al pari del risultato stesso. «Qualora nuove persone entrassero in un gruppo, avrebbero a disposizione tutto un passato di decisioni e motivazioni. Quando lasceranno il gruppo, il loro lavoro sarebbe già stato assorbito e integrato. E come interessante bonus, l'analisi automatica della rete di conoscenza potrebbe consentire ai partecipanti di trarre conclusioni sulla gestione e l'organizzazione della loro attività collettiva, un'impresa che non sarebbe stata possibile altrimenti»<sup>86</sup>. Uno scenario che non è molto distante da quello prefigurato, nel 1945, da Vannevar Bush.

## 7. Una pratica di comunicazione scientifica: dalle RFC all'open access

«Un elemento chiave della crescita di Internet è stato l'accesso libero e aperto alla documentazione di base, e in special modo alle specifiche dei protocolli». Ad affermarlo sono gli stessi inventori di Internet nella già citata *A Brief History of the Internet*, in cui un paragrafo è dedicato in particolare al ruolo della documentazione nel suo sviluppo e nella diffusione della rete<sup>87</sup>.

I giovani laureati che collaborano alla creazione di ARPANET e dei protocolli TCP/IP sono mossi dai medesimi principi che animavano il dibattito sulla repubblica delle lettere sin dal suo nascere, un dibattito di cui adottano metodi e concetti: se infatti il concetto di *humanitas* per gli umanisti era strettamente legato alla comunicazione, intesa come un atto di generosità verso gli altri che non si realizza in mera erudizione, comunicazione e comunicare sono termini che venivano usati per indicare il dovere di trasmettere il sapere ai posteri; non dunque un atto di liberalità, ma di giustizia. Trasmettere le proprie conoscenze costituiva il punto di arrivo e l'anima stessa del lavoro intellettuale: non solo un ideale, ma una condizione indispensabile allo svolgimento dell'attività di ricerca. Non erano dunque tanto l'amore per la ricerca quanto la volontà e la capacità di comunicare il sapere a definire il dotto degno di questo nome. Il dibattito scientifico pubblico si configurava come una rete che subiva una progressiva espansione e che, attraverso continue citazioni reciproche (espresse sulle riviste nella forma di recensioni, note, riassunti, bibliografie e indici, discussioni, lettere aperte) costruiva un discorso comune<sup>88</sup>. Gli scienziati americani prendono tuttavia le distanze dagli strumenti consueti della comunicazione scientifica, ritenendo «il ciclo normale della pubblicazione accademica tradizionale [...] troppo formalizzato e troppo lento per lo scambio dinamico di idee essenziali a creare reti»<sup>89</sup>, e si dotano di nuovi mezzi di comunicazione e pubblicazione, in primis le Request for Comments (RFC), formalizzando così nuove procedure per la costruzione di consenso.

L'invenzione della posta elettronica (e-mail) nel 1972 è in questo quadro essenziale. L'e-mail si configura da principio come una “killer application”, un'applicazione decisiva per la diffusione di Internet che è, ancora oggi, il servizio più usato in rete. La nascita delle e-mail è molto importante anche per la definizione del concetto di “authorship”, letteralmente “paternità intellettuale”, concetto che in un primo momento resta indeterminato a vantaggio di una scrittura collettiva e in cui l'autore assume un'importanza marginale e secondaria. La diffusione

---

85 Ivi, p. 128.

86 Ivi, p. 143.

87 B.M. Leiner, V.G. Cerf V.G., D.D. Clark, R.E. Kahn, L. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts, S. Wolff, *A Brief History of the Internet*, cit., p. 35.

88 Su questo mi permetto di rimandare a F. Di Donato, *Quale futuro per la repubblica scientifica? Kant e il dibattito illuminista su scienza, università e politica*, di prossima pubblicazione presso FUP, Firenze (si veda in particolare il cap. 1).

89 M. Castells, *Galassia Internet*, cit. p.

di mailing list, liste di discussione, integra lo strumento delle RFC sistematizzando le procedure decisionali all'interno dei gruppi di lavoro.<sup>90</sup>

Pensata come strumento per condividere l'informazione, Internet ha così trasformato il principio della libera collaborazione tra scienziati in una pratica di ricerca e di lavoro, arricchendosi nel tempo di sempre nuovi strumenti. Un principio che è stato teorizzato e formalizzato grazie alla sinergia col movimento per il software libero (free software) nelle sue diverse varianti<sup>91</sup>. È Richard Stallman, sistemista al laboratorio di intelligenza artificiale del MIT negli anni 70, il primo a farsi portatore della filosofia che guida il movimento a partire dai primi anni 80 quando, alle prese con una stampante donata dalla Xerox al prestigioso laboratorio statunitense, si scontra con l'impossibilità di modificare il codice sorgente del dispositivo a causa dell'imposizione del copyright sul software da parte della casa di produzione. Un evento scatenante che porterà il fisico a lasciare il MIT, prefiggendosi l'obiettivo di difendere la libertà del software e di creare, a tal fine, un sistema operativo<sup>92</sup> aperto compatibile con Unix<sup>93</sup>, intenzione che comunica al mondo nel 1983 attraverso ARPANET. «La sua convinzione sulla non utilità e, anzi, sulla dannosità di non diffondere il codice di controllo della macchina, basata su premesse insieme etiche e funzionali, trovava una continua conferma nei molti problemi quotidiani connessi all'utilizzo di computer e altra strumentazione elettronica.»<sup>94</sup>

La definizione di software libero si fonda sull'importante distinzione tra codice sorgente, il programma che viene scritto dai programmatori in un linguaggio formalizzato simile al linguaggio naturale, e codice eseguibile, scritto in linguaggio macchina (fatto di zero e uno), molto difficile da scrivere e pressoché impossibile da leggere per gli umani<sup>95</sup>. Il codice sorgente viene convertito (in termini tecnici “compilato”) in codice eseguibile tramite appositi programmi. Un passaggio che consente al programma di funzionare, ma che comporta la perdita di informazioni essenziali al fine di comprenderlo e modificarlo. Pertanto, l'espressione "software libero" implica le libertà di eseguire, copiare, distribuire, studiare, cambiare e migliorare il software.

«Più precisamente, esso si riferisce a quattro tipi di libertà per gli utenti del software:

1. Libertà di eseguire il programma, per qualsiasi scopo (libertà 0).
  2. Libertà di studiare come funziona il programma e adattarlo alle proprie necessità (libertà 1).
- L'accesso al codice sorgente ne è un prerequisito.

---

90 Si osservi ad esempio che la Internet Engineering Task Force ha 75 gruppi di lavoro, ciascuno dei quali lavora su un aspetto della progettazione e implementazione di Internet. I gruppi fanno uso di mailing list per discutere e poi, una volta che sia stato raggiunto consenso su un documento, questo viene distribuito come RFC.

91 Un'espressione alternativa a “Free software” è “Open source software”, con la quale si intende software il cui codice sorgente è aperto, ma rilasciato con licenze meno restrittive dalla licenza GNU/General Public License (GPL), promossa dalla Free Software Foundation. Si vedano i termini della licenza GNU/GPL all'URL: <[http://it.wikipedia.org/wiki/GNU\\_GPL](http://it.wikipedia.org/wiki/GNU_GPL)>. Si veda inoltre la seguente definizione di Open Source Software: <[http://it.wikipedia.org/wiki/Open\\_source](http://it.wikipedia.org/wiki/Open_source)>.

92 Il sistema operativo è “un insieme di programmi che consentono la gestione razionale delle risorse del calcolatore, dalle unità di ingresso-uscita (tastiera, video, stampante) alla memoria centrale (quella che oggi viene chiamata RAM) e periferica (tipicamente, l'hard disk). Il sistema operativo è la parte centrale del software di base, ossia di quell'insieme di quei moduli software che sono generalmente venduti insieme all'hardware [...]”. M. Berra, A.R. Meo, *Informatica solidale*, Bollati Boringhieri Torino 2001, p. 84. Per una definizione tecnica, si veda la voce:

<[http://it.wikipedia.org/wiki/Sistema\\_operativo](http://it.wikipedia.org/wiki/Sistema_operativo)>

93 Il nome del sistema operativo di Stallman, “GNU”, è l'acronimo ricorsivo di “GNU is not Unix”.

94 M. Berra, A.R. Meo, *Informatica solidale*, cit., p. 83. Per una ricostruzione della storia del movimento GNU – free software, si vedano in particolare le pp. 82-97. Nel 1985 Stallman fonda la Free Software Foundation, un'organizzazione nata allo scopo di convogliare le forze (lavoro e denaro) per lo sviluppo e la diffusione di programmi aperti.

95 «I calcolatori della prima generazione dovevano essere programmati in linguaggio macchina, ossia in questo complicato linguaggio fatto di lunghissime sequenze di simboli 1 e 0. [...] La programmazione in linguaggio macchina era operazione molto lunga e complessa, perché richiedeva la conoscenza minuta dell'architettura del calcolatore, dei codici di tutte le istruzioni e il controllo mentale delle molte celle della memoria centrale. Per questo furono presto ideati linguaggi simbolici e sviluppati opportuni programmi “traduttori”, capaci di tradurre il codice simbolico scritto dal programmatore in un programma in linguaggio macchina che facesse le stesse cose”. M. Berra, A.R. Meo, *Informatica solidale*, cit., p. 74.

3. Libertà di ridistribuire copie in modo da aiutare il prossimo (libertà 2).
  4. Libertà di migliorare il programma e distribuirne pubblicamente i miglioramenti (e le versioni modificate in genere), in modo tale che tutta la comunità ne tragga beneficio (libertà 3). L'accesso al codice sorgente ne è un prerequisito.
- Un programma è software libero se l'utente ha tutte queste libertà»<sup>96</sup>.

L'aggettivo “free” si riferisce dunque alla libertà e non al prezzo. Nulla vieta infatti che il software libero possa essere venduto (Stallman stesso chiarisce questo aspetto, sottolineando che il termine è utilizzato come nell'espressione “free speech” e non, invece, nell'accezione di “free beer”). Essenziale a questo proposito è il concetto di “copyleft”, letteralmente diritto di copia, un gioco di parole che sfrutta la contrapposizione tra “right” e “left”, destra e sinistra, e che si traduce praticamente nell'adozione di licenze libere, licenze tramite le quali l'autore cede parte dei suoi diritti (economici) sull'opera stabilendo le condizioni alle quali può questa essere utilizzata<sup>97</sup>.

Ma è la convergenza tra il movimento free software e il progetto Linux dell'hacker finlandese Linus Torvalds a dare ampia diffusione alla filosofia pratica del movimento. Gran parte dell'infrastruttura del web (tra cui, oltre al codice di Linux, è opportuno ricordare Apache, MySQL, Perl, PHP, Python) si fonda sul metodo che trova spazio a partire dalla creazione di Internet e che si diffonde in modo virale tramite la rete.

In tale contesto, «l'etica rappresenta il collante e stabilisce lo standard di un comportamento accettato da chi sente di far parte di questo gruppo in cui la capacità, la conoscenza, l'interesse per il calcolatore sono gli elementi primari della comunicazione e la ragione dello stare insieme, e sono alla base di una organizzazione non gerarchica ma reticolare, basato sullo scambio e sull'interazione»<sup>98</sup>. Si diffonde cioè quella che Pekka Himanen definisce etica hacker, un atteggiamento di cui il mondo accademico può essere considerato il predecessore più antico<sup>99</sup>. E' questo infatti uno dei tratti essenziali della metodologia e della pratica di ricerca che caratterizza lo sviluppo delle reti telematiche, un fenomeno che Pierre Lévy definisce come “intelligenza collettiva”, intelligenza distribuita e connessa in rete che valorizza le competenze individuali attraverso uno spirito collaborativo<sup>100</sup>.

Nonostante l'evidente assonanza di tutti questi principi con la pratica della comunicazione scientifica tradizionale, fatta eccezione per i casi sopra descritti il mondo dell'accademia ha tardato a fare propri gli strumenti di comunicazione e, soprattutto, di pubblicazione che Internet e il Web hanno reso disponibili e accessibili. È solo tra il 2002 e il 2003, infatti, che, sulla scia di alcune significative esperienze pilota<sup>101</sup>, in ambito accademico si è cominciato a parlare di letteratura scientifica *open access*, vale a dire letteratura «digitale, online, gratuita e libera da alcune restrizioni dettate dalle licenze per i diritti di sfruttamento commerciale». Condizioni rese possibili «grazie a Internet e al consenso dell'autore o del titolare dei diritti d'autore»<sup>102</sup>.

Il movimento per l'accesso aperto alla letteratura scientifica nasce nella comunità accademica avviando una campagna in favore della condivisione dell'informazione e della conoscenza, intese come beni comuni<sup>103</sup>, e come soluzione al problema che affligge il mondo delle

---

96 <<http://www.gnu.org/philosophy/free-sw.it.html>>

97 La Free Software Foundation di Stallman crea la General Public License (cfr. nota 90). Le licenze copyleft più diffuse sono le licenze Creative Commons (<http://www.creativecommons.it/>), promosse dal giurista americano Lawrence Lessig. Per le premesse teoriche del progetto di Lessig, si vedano: L. Lessig, *Code and other laws of cyberspace*, cit.; L. Lessig, *Il futuro delle idee*, Feltrinelli, 2006; L. Lessig, *Cultura libera*, cit.

98 M. Berra, A.R. Meo, *Informatica solidale*, cit., p. 94.

99 P. Himanen, *L'etica hacker e lo spirito dell'età dell'informazione*, Feltrinelli, Milano 2001, p. 17.

100 P. Lévy, *Intelligenza collettiva. Per un'antropologia del cyberspazio*, Feltrinelli, Milano 1999.

101 Si veda in particolare l'esperienza di ArXiv, l'archivio di pre-print di fisica, matematica, informatica e biologia creato da Paul Ginsparg a Los Alamos nel 1991 e poi spostato alla Cornell University. Cfr. <<http://it.wikipedia.org/wiki/ArXiv>> e <<http://arxiv.org/>>. Oggi ArXiv contiene circa cinquecentomila articoli.

102 P. Suber, *Breve introduzione all'accesso aperto*, 2004, on-line all'URL <<http://www.aepic.it/docs/OA/brief-italian.htm>>.

103 Si veda C. Hess, E. Ostrom (a cura di), *La conoscenza come bene comune. Dalla teoria alla pratica*, Mondadori, Milano 2009.

biblioteche e noto come “crisi del prezzo dei periodici”, vale a dire l'aumento vertiginoso del costo degli abbonamenti alle riviste scientifiche<sup>104</sup>. Su tale base, l'open access diviene al centro di un ampio dibattito tra ricercatori, bibliotecari, amministratori di università e centri di ricerca, agenzie di finanziamento, editori e studenti.

In pratica, il movimento per l'accesso aperto si dota di due canali:

1. Gli archivi elettronici aperti, che non richiedono che i risultati depositati siano stati preventivamente sottoposti a peer review e che possono essere istituzionali (come quelli di università e centri di ricerca), o disciplinari.
2. le riviste ad accesso aperto, che non sono diverse dalle riviste tradizionali se non in quanto rendono i loro articoli selezionati liberamente disponibili al mondo. «I loro costi consistono nella gestione dei processi editoriali, dal referaggio alla manipolazione dei manoscritti, allo spazio sul server. Le riviste ad accesso aperto coprono i propri costi in un modo molto simile alle radio e alle televisioni: coloro che hanno interesse alla disseminazione dei contenuti pagano anticipatamente i costi di produzione in modo che l'accesso alla lettura sia libero»<sup>105</sup>.

Grazie alla diffusione di questi strumenti, una sempre maggiore quantità di dati diviene liberamente disponibile sul Web.

Ma è la filosofia politica alla base del movimento a essere particolarmente interessante ai fini di questo discorso. Scrivono infatti i promotori della Dichiarazione di Berlino<sup>106</sup>:

«La nostra missione di disseminazione della conoscenza è incompleta se l'informazione non è resa largamente e prontamente disponibile alla società. Occorre sostenere nuove possibilità di disseminazione della conoscenza, non solo attraverso le modalità tradizionali ma anche e sempre più attraverso il paradigma dell'accesso aperto via Internet.

[...]

Per mettere in pratica la visione di un'istanza globale ed accessibile del sapere, il Web del futuro dovrà essere sostenibile, interattivo e trasparente. I contenuti ed i mezzi di fruizione (tools) dovranno essere compatibili e ad accesso aperto».

Il collegamento tra tali affermazioni, la filosofia del movimento per il software libero e le premesse che hanno portato all'invenzione di Internet e del Web, che i sostenitori dell'accesso aperto estendono ai contenuti scientifici, in particolare accademici, è evidente.

## 8. Il social software: ragnatele di dati e reti sociali

La disponibilità di una grande quantità di dati è una precondizione allo sviluppo del “web 2.0”, espressione coniata dall'editore americano Tim O'Reilly per definire «un insieme di tendenze economiche, sociali e tecnologiche che formano collettivamente la base per la futura generazione di Internet, che diventa un mezzo più maturo» e definito sulla base di alcune caratteristiche: partecipazione degli utenti, apertura ed effetti di rete<sup>107</sup>.

L'aspetto tecnologico del Web 2.0 è funzionale alle caratteristiche delle reti sociali e della cooperazione tra utenti, e le tecnologie che vengono sviluppate poggiano sui principi architettonici di Internet, Web e web semantico, vale a dire l'apertura e l'universalità che si esplicano nell'architettura a strati, nella decentralizzazione, nella “open world assumption”, nella

104 Si veda J.-C. Guédon, *Per la pubblicità del sapere. I bibliotecari, i ricercatori, gli editori e il controllo dell'editoria scientifica*, PLUS Méthexis, Pisa, 2004, anche online all'URL: <<http://bfp.sp.unipi.it/ebooks/guedon.html>>

105 P. Suber, *Breve introduzione all'accesso aperto*, cit.

106 Nel 2009, la Dichiarazione di Berlino è stata sottoscritta da oltre duecentosessanta istituzioni di ricerca (per un elenco completo dei firmatari, si veda <<http://oa.mpg.de/openaccess-berlin/signatories.html>>).

107 Cfr. T. O'Reilly, *Che cos'è web 2.0*, 2004 online all'URL: <<http://www.awaredesign.eu/articles/14-Cos-Web-2-0>>. Si veda anche il bellissimo video “Web 2.0 ... The Machine is Us/ing Us” di Michael Wash docente di antropologia culturale alla Kansas State University, online su Youtube all'URL: <<http://www.youtube.com/watch?v=6gmP4nk0EOE>>.

centralità degli URI e nel principio del “data linking”.

Si afferma un principio noto come “mash up” (letteralmente: “poltiglia”)<sup>108</sup>, che consiste nella combinazione e nel riutilizzo di informazioni e servizi. Un esempio di ciò è la combinazione di un servizio google come google maps con altri servizi, ad esempio flickr, il sito che consente agli utenti di pubblicare sul web le proprie fotografie; nello specifico, la combinazione di google maps con flickr permette di localizzare e visualizzare le fotografie su una mappa. La tecnologia che permette di farlo sono le API (Application programming interface)<sup>109</sup>, insiemi di procedure, strutture dati, classi di oggetti e protocolli forniti dalle librerie e dai servizi del sistema operativo per supportare l'implementazione di applicazioni, che, quando liberamente disponibili, si dicono aperte (Open API).

I principi della programmazione di software libero/open source includono il fatto che gli utenti siano coinvolti nel processo di programmazione e trattati come co-sviluppatori, che il software sia reso pubblico spesso e non subordinato all'accumularsi di miglioramenti e innovazioni sostanziali, e un modello di progettazione indipendente dall'hardware (ora sempre più aperto anche ai cellulari) e a strati.

Il web fa dunque da specchio alla società umana di cui riflette gli interessi, che si estendono su un arco molto ampio di assunti, valori e culture. La rete è così usata per condividere informazioni, per divertirsi, per fare scienza in modi radicalmente nuovi, per il business<sup>110</sup>. Il Web 2.0 è stato così rappresentato «come un insieme di principi e di procedure che collegano un autentico sistema solare di siti che fanno propri tali principi, in tutto o in parte, a una distanza variabile da tale centro.»<sup>111</sup> Quali sono dunque i principi che caratterizzano la fisionomia del web 2.0? I concetti, o architravi portanti della sua struttura tecnica e sociale, sono sintetizzabili in cinque punti:

#### 1) *"Il web come piattaforma"*.

I servizi web sostituiscono sempre più le applicazioni desktop. Un esempio è l'account di Google, che offre gratuitamente agli utenti collegati l'integrazione di servizi prima disponibili solo in locale, come ad esempio editor di scrittura (tipo word) e suite di applicazioni come Microsoft Office. Google diviene così l'intermediario tra utenti e informazione, in cui le applicazioni non sono concepite come pacchetti in vendita ma fornite come servizio, senza licenze o condizioni di vendita, e senza richiedere il porting su piattaforme diverse affinché i clienti possano utilizzare il software sulle proprie macchine (in termini tecnici: platform independent). Un corollario di questa trasformazione è che il software è concepito sempre più come un servizio e non come un prodotto. Il modello google si contrappone così al modello tradizionale à la Microsoft: «da una parte, un singolo fornitore di software, la cui massiccia base e il sistema operativo strettamente integrato con le API permettono di controllare il paradigma di programmazione; dall'altra, un sistema senza un proprietario, tenuto insieme da una serie di protocolli, standard aperti e accordi di cooperazione.»<sup>112</sup>

Con il Web 2.0 si supera il concetto di sito web come entità autonoma incapace di comunicare con il mondo esterno; si diffondono molte applicazioni sviluppate dal basso da comunità decentralizzate, che rientrano nel cosiddetto social software; inoltre, i dati online diventano indipendenti dalle applicazioni e possono essere utilizzati da servizi diversi.

#### 2) *Centralità dei dati e contenuti prodotti dagli utenti.*

108 In termini informatici, indica un'applicazione che usa contenuto da più sorgenti per creare un servizio completamente nuovo. Cfr. <[http://it.wikipedia.org/wiki/Mash-up\\_\(informatica\)](http://it.wikipedia.org/wiki/Mash-up_(informatica))>.

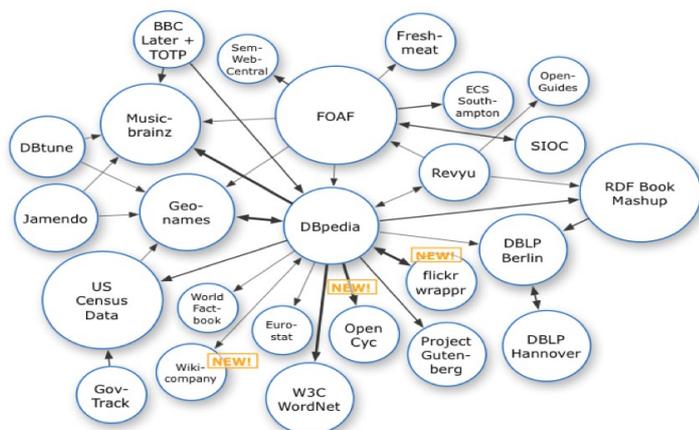
109 Si veda la definizione su Wikipedia all'URL: <[http://it.wikipedia.org/wiki/Application\\_programming\\_interface](http://it.wikipedia.org/wiki/Application_programming_interface)>.

110 A Framework, p. 80.

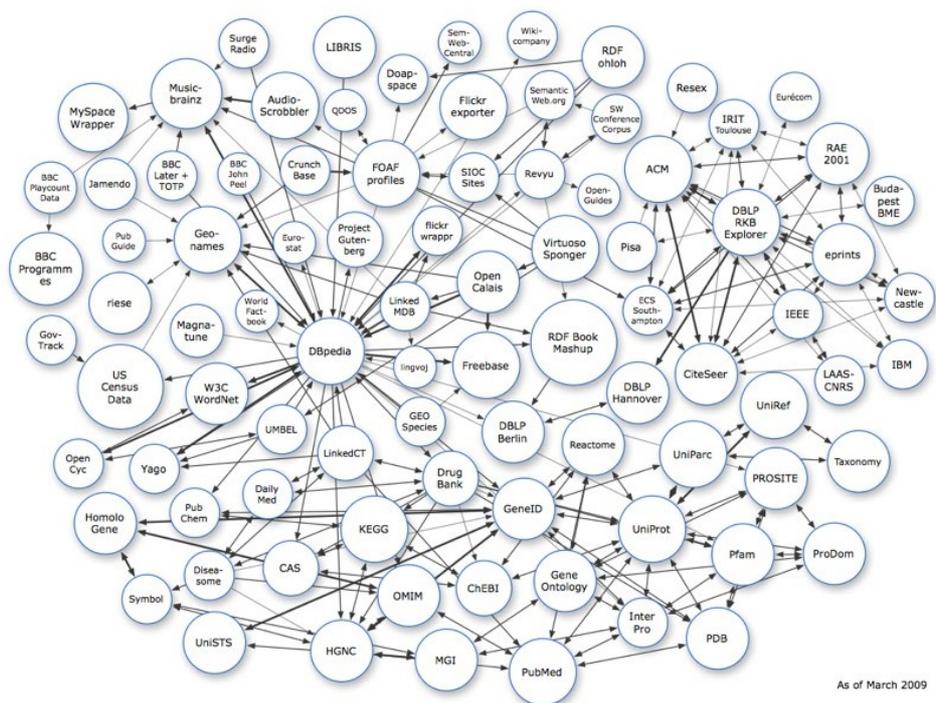
111 T. O'Reilly, Che cos'è web 2.0, cit.

112 T. O'Reilly, Che cos'è web 2.0, cit.

«Senza i dati, scrive ancora O'Reilly parafrasando una nota affermazione di Kant, gli strumenti sono inutili; senza il software, i dati sono ingestibili». Google si configura così come un insieme di strumenti software e un ampio database specializzato<sup>113</sup>. La disponibilità di un'enorme quantità di dati è una caratteristica fondamentale del nuovo web.



In figura: basi di dati sul Web, 2007.



In figura: l'immagine sopra, aggiornata al marzo 2009.

I contenuti sono generati dagli utenti e condivisi all'interno di comunità. In questo modo, sono gli stessi utenti a creare il "servizio"<sup>114</sup>. È questo un elemento caratterizzante il web 2.0 e il social software: «Da "The computer moves in" a "Yes, you. You control the information age. Welcome to your

113 T. O'Reilly, Che cos'è web 2.0, cit.

114 Il web è ricco di esempi. Tra essi, oltre al già citato Flickr <<http://www.flickr.com>> vale la pena menzionare Youtube e ebay, che non hanno bisogno di spiegazioni.

*world*”. Dall’ormai celebre copertina del *Time* del 3 gennaio 1983<sup>115</sup>, che nominava il personal computer come “*machine of the year*”, a quella, quasi un quarto di secolo dopo, del 25 dicembre 2006<sup>116</sup>, in cui il titolo di “*man of the year*” va alla seconda persona, sia singolare sia plurale, della lingua inglese, scritta, o meglio, digitata al centro di un monitor». L'uomo dell'anno, recita il *Time*, sei tu. Benvenuto nel tuo mondo. «Se nel gennaio del 1983 la figura umana veniva presentata stilizzata e glaciale di fronte al computer, sul cui schermo comparivano numeri e grafici, in un ambiente cupo dove, per contrasto, solo la sedia ed il tavolo erano caratterizzati da un aspetto tradizionale e da colori caldi<sup>117</sup>, nella copertina del 2006 questa presenza è totalmente astratta, ridotta all'essenza di un pronome personale, al posto dei grafici nel monitor, ora a cristalli liquidi.»<sup>118</sup> La rete di calcolatori lascia così spazio alle reti sociali sul web, in cui gli utenti producono e condividono contenuti, arricchendoli di ulteriori dati e collegandoli tra loro. I blog sono, in questo contesto, un importante strumento di pubblicazione interamente nelle mani degli utenti che si diffondono anche grazie ai “trackback”, link monodirezionali che creano l'effetto di collegamenti a due vie, i quali incrementano il grado di connessione dei contenuti e creano zone in cui il Web si presenta come una rete indiretta e più facile da navigare. «La “blogosfera” può essere considerata un nuovo equivalente peer-to-peer di Usenet e dei bulletin-board, le aree conversazionali del primo internet. Non solo è possibile iscriversi ai siti degli altri e collegarsi facilmente ai singoli commenti su una pagina: tramite un meccanismo noto come trackback si può anche vedere quando qualcun altro si collega alle proprie pagine e rispondere, con link reciproci o aggiungendo commenti».

Si diffondono infine “folksonomie”, termine che deriva dalla contrazione dei termini “folks” e “taxonomy” e che consistono in stili di categorizzazione collaborativa che utilizzano parole chiave liberamente scelte, generalmente definite tag. Se l'uso di ontologie aggiunge una struttura ai dati, in questo secondo caso la struttura emerge in modo organico dall'organizzazione delle richieste informative dei singoli individui. Le ontologie vengono integrate da folksonomie, che sorgono quando un ampio numero di persone è interessato a qualche informazione e incoraggiato a descriverla – in gergo, a taggarla.

All'aumentare dei tag inseriti dagli utenti, questi tendono a essere riusati e applicati a nuovi elementi da altri utenti. Certamente, la libertà nell'assegnare parole chiave a un contenuto può creare un eccesso di informazione e una certa confusione. Tuttavia, i tag sono generati dalle interazioni che avvengono nel mondo reale e rilevano connessioni reali tra contenuti e utenti. Tali strutture permettono alla semantica di emergere da accordi impliciti (*ex post*), diversamente da quanto avviene nella costruzione di ontologie, le quali indicano accordi espliciti (*ex ante*); il campo delle dinamiche semiotiche ha come premessa l'idea che gli accordi comunicativi o i sistemi di organizzazione dell'informazione spesso si sviluppano attraverso processi decentralizzati di invenzione e negoziazione<sup>119</sup>.

Tuttavia, la distinzione tra ontologie e folksonomie ha un notevole rilievo su un piano filosofico, e merita qualche parola in più. È stato infatti osservato che le folksonomie sono preferibili all'utilizzo di ontologie controllate e centralizzate. Se infatti la possibilità di annotare pagine web usando vocabolari controllati può accrescere le possibilità di trovare ciò che cerchiamo con esattezza, d'altra parte una base ampia ed eterogenea di utenti non consente che siano molte le persone che scelgono di adottare o di mantenere un'ontologia complessa. Creare ontologie richiede un investimento di tempo e di energie molto elevato. Tuttavia è altrettanto

---

115 <<http://www.time.com/time/covers/0,16641,19830103,00.html>>

116 <<http://www.time.com/time/covers/0,16641,20061225,00.html>>

117 Un approccio visuale simile sarà utilizzato l'anno successivo nel famoso spot pubblicitario “1984” di Ridley Scott per l'Apple Macintosh. Il video dello spot è disponibile on-line a diversi indirizzi, per maggiori informazioni vedi <[http://en.wikipedia.org/wiki/1984\\_%28television\\_commercial%29](http://en.wikipedia.org/wiki/1984_%28television_commercial%29)>.

118 F. Meschini, *eContent: tradizionale, semantico o 2.0?*, online all'URL: <<http://dspace.unitus.it/handle/2067/162>>.

119 T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit. p. 32.

vero che le parole chiave scelte da una persona in modo indipendente e scoordinato dagli altri possono essere inutili o inaccurate, e anche se l'uso di folksonomie è assai diffuso, queste rimangono uno strumento di classificazione impreciso. Allo stesso tempo, però, la diffusione di strumenti per l'uso di ontologie è assai più lenta di quanto non stia avvenendo per le applicazioni sociali che si servono delle folksonomie. Dunque, è inopportuno e improprio considerare folksonomie e ontologie come strategie mutuamente escludentisi – molto meglio invece adottarle come complementari. Le prime si rivelano così di grande utilità per la ricerca di documenti da parte degli utenti; le seconde sono essenziali per il recupero automatizzato di dati, in modo particolare in campo scientifico<sup>120</sup>. Gli esempi di convergenza tra l'impostazione del web semantico e quella del web2.0 non mancano. Tra gli altri, vale la pena ricordare *Visual Thesaurus*<sup>121</sup>, un dizionario della lingua inglese, che utilizza il software di visualizzazione *ThinkMap*<sup>122</sup> e che è basato su *WordNet*<sup>123</sup>, un database lessicale della lingua inglese il cui progetto è stato avviato nel 1985, in cui i termini sono organizzati tra di loro sulla base di rapporti semantici che caratterizzano le classificazioni ontologiche (iponimia, ipernimia, omonimia e meronimia). Il vocabolario presenta i vari significati di un termine in forma di grafo i cui rami sono tipizzati in base al tipo di relazione. «Inserendo “*Information*” ad esempio, si può navigare il grafo fino ad arrivare ai termini “*Data*” e “*Knowledge*” da un lato, e “*Entropy*” dall'altro. E ancora, “*Love*” presenta numerosi nodi, che coprono tutti i vari significati del termine, dai significati più astratti all'attrazione fisica, e presenta anche, con un rapporto di antinomia, la parola “*Death*”»<sup>124</sup>. L'uso di *WordNet* con *ThinkMap* dimostra come sia possibile creare servizi avanzati sulla base di contenuti organizzati semanticamente.

### 3) *Intelligenza collettiva.*

«L'architettura di internet, e il World Wide Web, così come i progetti di software open source come Linux, Apache, e Perl, è tale che gli utenti che perseguono i propri interessi “egoistici” costruiscono un valore collettivo come conseguenza automatica. Ognuno di questi progetti ha un piccolo nucleo centrale, meccanismi di estensione ben definiti e un approccio che consente a chiunque di aggiungere qualsiasi componente ben funzionante, facendo crescere gli strati più esterni di quello che Larry Wall, il creatore di Perl, definisce “la cipolla”. In altre parole, queste tecnologie dimostrano gli effetti della rete, semplicemente attraverso il modo in cui sono state progettate»<sup>125</sup>. Una conseguenza di questo fatto è la nascita di reti sociali tra persone che condividono gli stessi interessi, i cui esempi più noti sono Myspace e, soprattutto, Facebook, il cui numero di utenti ha raggiunto, nell'aprile 2009, i duecento milioni, e il cui successo è dato dalla facilità con cui è possibile trovare e collegarsi ad altri utenti già presenti nel “mondo piccolo” network.

Un altro esempio illustre di intelligenza collettiva è Wikipedia, un'enciclopedia online basata sull'idea che ciascuna voce possa essere aggiunta e modificata da chiunque e che applica alla creazione di contenuti il detto di Eric Raymond (coniato originariamente nel contesto del software open source), secondo cui «con molti occhi puntati addosso, ogni bug diventa una bazzecola»<sup>126</sup>. L'impostazione di Wikipedia si sta diffondendo anche in ambito scientifico e letterario, dove il tradizionale sistema delle recensioni su carta è affiancato da veri e propri servizi on-line come quello di Internet Bookshop, Amazon.com e Anobii, che oltre alle recensioni consente agli utenti di inserire moltissime informazioni (data di inizio e fine lettura,

---

120 Ivi.

121 <<http://www.visualthesaurus.com/>>.

122 <<http://www.thinkmap.com/>>.

123 <<http://wordnet.princeton.edu/>>.

124 F. Meschini, *eContent: tradizionale, semantico o 2.0?*, cit.

125 T. O'Reilly, *Che cos'è web 2.0*, cit.

126 Ivi.

commenti come nelle note a margine, un voto che va da uno a quattro) e che contiene oltre undici milioni di libri.<sup>127</sup> Vale infine la pena ricordare strumenti per la creazione di bibliografie collaborative come Citeulike e Connotea<sup>128</sup>, che permettono di pubblicare sul web liste bibliografiche e di navigare nelle bibliografie altrui tramite folksonomie, e servizi più evoluti per la ricerca scientifica come Zotero<sup>129</sup>, un'estensione del browser firefox che avvicina sempre di più il desktop virtuale del ricercatore alla sua scrivania tradizionale, arricchita da funzioni eccezionali. Un tool particolare, ad esempio, permette di annotare i testi con note a margine, proprio come si fa con i documenti cartacei. Dei vari "online reference managers" forse Zotero è quello che può maggiormente rispondere ai bisogni degli umanisti che lavorano prevalentemente sui testi (fonti primarie e secondarie). Zotero permette di gestire i documenti di archivi online (come ad esempio JSTOR), dei word processor (sia online, tipo Google doc, sia offline) e dei siti commerciali (Amazon, Google scholar) oltre a vere e proprie "note" aggiunte dal lettore.

#### 4) *Decentralizzazione: ogni client è anche un server (P2P).*

Molti sistemi del web2.0 sono pensati e progettati per incoraggiare la partecipazione diretta degli utenti in almeno tre modi, indicati da Dan Bricklin come segue: «Il primo, dimostrato da Yahoo!, è di pagare le persone perché lo facciano. Il secondo, che prende ispirazione dalla comunità open source, è di cercare volontari che realizzino lo stesso compito. L'Open Directory Project, un concorrente open source di Yahoo, ne è il risultato. Ma Napster ha dimostrato un terzo modo. Avendo come default la possibilità di mettere a disposizione automaticamente qualsiasi pezzo musicale che viene scaricato, Napster ha consentito a ogni utente di contribuire all'aumento del valore del database condiviso. Questo stesso approccio è stato seguito da tutti gli altri servizi di condivisione di file P2P.»<sup>130</sup> I protocolli peer to peer trasformano i client (componenti che accedono ai servizi o alle risorse) in server (componenti che forniscono servizi). In pratica, sezioni particolari della memoria dei calcolatori degli utenti divengono accessibili agli altri utenti.

#### 4) *Accesso aperto.*

In tale contesto, le licenze software e il controllo delle API divengono irrilevanti in quanto il software non ha più bisogno di essere distribuito ma solo utilizzato, e il suo valore è proporzionale alla scala e al dinamismo dei dati che esso aiuta a gestire. L'accesso aperto permea così tutti gli strati della rete, dai protocolli ai contenuti, ed è un principio che informa la rete ad ogni livello, definendo tanto gli aspetti tecnici tanto la filosofia e gli aspetti socio-culturali del cosiddetto software sociale.

Da tali premesse gli umanisti e gli scienziati sociali dovrebbero partire per prendere finalmente parte alla discussione e alle decisioni, di natura tanto tecnica quanto politica, su almeno due questioni che sono oggi al centro del dibattito della scienza del Web.

La prima riguarda la cosiddetta governance del Web. Se infatti la struttura decentralizzata della rete non facilita l'imposizione di standard, la creazione di un'infrastruttura attenta alle policy dovrà essere affrontata nello sviluppo degli strati più alti del Semantic Web, studiando il modo in cui rendere possibile l'implementazione di regole per diritti d'accesso, misure di sicurezza e rispetto della privacy degli utenti<sup>131</sup>.

127 Si veda L. Lipperini, "La critica fai da te. Dai libri al web, ecco il popolo dei recensori", la Repubblica, 2 giugno 2009.

128 <<http://www.citeulike.org>>; <<http://www.connotea.org>>.

129 <<http://www.zotero.org/>>

130 D. Bricklin, *The Cornucopia of the Commons*, cit in T. O'Reilly, *Che cos'è web 2.0*, cit.

131 D. J. Weitzner, J. Hendler, T. Berners-Lee, and D. Connolly, "Creating a Policy-Aware Web: Discretionary, rule-based access for the World Wide Web," in E. Ferrari B. Thuraisingham (a cura di), *Web and Information Security*, Hershey PA: Idea

La seconda questione, particolarmente rilevante per la ricerca scientifica sul Web, è relativa al problema della qualità delle informazioni accessibili in rete. Come possiamo fare in modo che il Web contenga buona scienza, piuttosto che mere superstizioni? Come trovare l'equilibrio migliore tra il libero scambio di opinioni e la selezione di informazioni secondo criteri di qualità?

Sul Web esistono siti, detti autorità (*authority*), che contengono molti link in entrata, vale a dire che sono ritenuti affidabili e pertanto ricevono molti link da altre pagine; e siti detti connettori (*hub*), che hanno un alto numero di link in uscita e che sono ritenuti attendibili in quanto puntano a informazione valida. Il problema della qualità dei contenuti è legato alla loro ricezione da parte dei lettori, una questione che dipende da criteri soggettivi e che rientra nell'ultimo strato del "layer cake" semantico, la "fiducia" (*trust*). È intuitivo che si tratta di un fattore importante per lo sviluppo del Web da molti punti di vista: i contenuti scientifici sono ritenuti attendibili in quanto contengono risultati validi; l'autore di un sito ha la fiducia dei lettori che lo conoscono; e lo stesso vale per i servizi che hanno dimostrato di essere affidabili (per esempio, per quanto riguarda la gestione dei dati delle carte di credito o dei dati personali). Perciò, la fiducia è un elemento essenziale nell'architettura del Web. E tuttavia, sul piano teorico, è un tema difficile e sfuggente, sia perché l'architettura del Web consente l'anonimato e rende estremamente facile la copia, sia a causa dei molteplici contesti in cui avvengono le interazioni sul Web. È stato osservato che spesso non è facile distinguere tra l'attendibilità e le sue cause<sup>132</sup>; inoltre, spesso l'affidabilità di un sistema e la fiducia in un individuo sono state assimilate; infine, le definizioni di fiducia differiscono in modo sostanziale, rendendo difficile la costruzione di una teoria adeguata ad affrontare i problemi tecnici che si presentano agli sviluppatori che si propongono di implementare applicazioni per la selezione dell'informazione sulla base di criteri univoci, un problema che si scontra tra l'altro con la difficoltà di fissare un tantum un concetto che è tutt'altro che statico, bensì variabile nel tempo e nello spazio, e che non ha soltanto una valenza privata ma un importante impatto pubblico. Tuttavia, è essenziale considerare con un'impostazione differente il problema dell'affidabilità di un sistema e quello della fiducia in un individuo. Il primo problema riguarda questioni relative alla Web governance come la sicurezza e la privacy cui si è accennato sopra. Il secondo problema è più delicato in quanto chiama in causa la reputazione di un individuo, un elemento essenziale nella costruzione della fiducia e però più soggettivo e difficilmente definibile in termini universali (la buona reputazione di una persona ad esempio non garantisce che questa si comporti in futuro come ha fatto in passato). Lo studio di tecniche adeguate per comprendere come si forma la reputazione richiede dunque di seguire e approfondire più linee di ricerca, che coinvolgano tanto tecnici quanto filosofi e scienziati sociali<sup>133</sup>; una questione tra le altre, che mostra come i problemi epistemologici e morali siano parte indispensabile nell'agenda della scienza del Web.

---

Group Inc, 2005 <<http://www.mindswap.org/users/hendler/2004/PAW.html>>; T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., pp. 99-100.

132 S. Grabner-Kräuter, E. A. Kaluscha, "Empirical research in on-line trust: A review and critical assessment," *International Journal of Human-Computer Studies*, vol. 58, pp. 783-812, 2003.

133 T. Berners-Lee, W. Hall, J.A. Hendler, K. O'Hara, N. Shadbolt and D.J. Weitzner, *A Framework for Web Science*, cit., pp. 89-94.