

GINO RONCAGLIA

La rete medium globale

Internet e il web come strumenti di comunicazione

La rete Internet rappresenta lo strumento di comunicazione (e il *medium* di massa) che forse più di ogni altro caratterizza questi primi anni del nuovo millennio. Per capirne le caratteristiche attuali e delinearne le principali prospettive di sviluppo, può essere utile partire da alcune brevi considerazioni sulle modalità della comunicazione via rete e sulla sua evoluzione fino a oggi.

La maggior parte dei *mass media* tradizionali (cinema, televisione, radio, stampa ecc.) è basata su una comunicazione prevalentemente verticale o unidirezionale: il messaggio proviene da un unico mittente, o da un numero limitato di mittenti, mentre i destinatari sono molti, e non esiste di norma la possibilità di inversione del ruolo. Il processo comunicativo avviene dunque in una sola direzione: il mittente produce il messaggio, i (molti) destinatari non possono fare altro che riceverlo e decodificarlo.

Una modalità diversa caratterizza i *media* orizzontali, come la comunicazione epistolare o il telefono. In questo caso esiste una pluralità di mittenti e destinatari che possono scambiarsi i ruoli. Nei *media* orizzontali tradizionali, tuttavia, ogni singolo processo comunicativo coinvolge pochi utenti (di norma, solo due). La comunicazione è dunque bidirezionale, e assume la forma del dialogo: entrambi i protagonisti dell'interazione possono divenire mittenti e possono produrre messaggi.

I *media* reticolari, come Internet, sono un'evoluzione di quelli orizzontali. Anche in questo caso esistono molti emittenti e molti destinatari, che possono scambiarsi di ruolo. Ma ciascun agente è in grado di comunicare contemporaneamente con molti altri. Nella comunicazione reticolare si realizza quindi una interazione allargata, che coinvolge un gran numero di utenti.

La comunicazione reticolare non è propria solo delle reti telematiche: si può sostenere infatti che

forme di comunicazione reticolare siano anticipate da reti di relazioni sociali come quelle sviluppatesi nel 18° sec. nell'ambito delle *sociétés savantes* o come quelle che caratterizzano alcune forme di partecipazione politica (secondo alcuni, fin dalla democrazia ateniese). Robert Darnton ha efficacemente sottolineato – facendo particolare riferimento alla Francia del 18° sec. – come l'insieme degli strumenti di comunicazione utilizzati da una società istituisca sempre reti e strutture complesse di scambio informativo, anche indipendentemente dalle caratteristiche specifiche dei singoli *media* e dei canali di volta in volta utilizzati. Per molti versi, la comunicazione epistolare già in età classica, e, in età più recenti, quella telegrafica e telefonica, hanno sviluppato caratteristiche che andavano ben oltre il livello della comunicazione bidirezionale fra due interlocutori.

Ma la scala globale e aperta sulla quale Internet ha proiettato la comunicazione reticolare, e la pluralità di codici comunicativi il cui uso è reso possibile dalla codifica digitale dell'informazione (che permette di scambiare, attraverso gli stessi canali, informazione non solo testuale ma anche visiva e sonora, e di integrare codici comunicativi diversi in oggetti informativi multimediali), costituiscono un fenomeno del tutto nuovo. Come vedremo più in dettaglio fra breve, la *convergenza al digitale*, ovvero la tendenza a trasferire in formato digitale tipologie sempre più differenziate di informazione primaria e a utilizzare dispositivi digitali per la sua produzione e riproduzione, si sta trasformando all'inizio del terzo millennio in una vera e propria *convergenza verso la rete*, marcata dall'uso di Internet, e in particolare del web, come strumento principe per la produzione, la conservazione, la circolazione, il reperimento e la fruizione di molteplici tipologie di contenuti informativi.

Le radici di questo sviluppo vanno ricercate nella trasformazione di Internet in una rete di interconnessione globale aperta ai privati e, in prospettiva, ai singoli cittadini. La nascita della *Internet society*, nel 1992, marca questa trasformazione, fornendo a Internet un punto di riferimento attraverso una organizzazione internazionale, non governativa, che vede la partecipazione di una vasta base di organismi e di individui impegnati nella definizione di standard pubblici e condivisi.

Negli stessi anni, Tim Berners-Lee (uno dei padri del web), ricercatore al CERN (*European Organization for Nuclear Research*) di Ginevra, lavora

alla creazione di sistemi di rappresentazione ipertestuale dell'informazione utilizzabili via rete, immaginando un'architettura *client-server* per lo scambio di pagine informative ipertestuali e proponendo l'uso di un linguaggio semplice e immediato per l'inserimento di rimandi ipertestuali in documenti di testo: l'HTML (*Hyper-Text Markup Language*). Contemporaneamente, presso il National center for supercomputer applications della University of Illinois, Marc Andreessen e il suo gruppo di ricerca sviluppano un programma *browser* in grado di visualizzare e navigare in maniera efficace i documenti HTML e i rimandi ipertestuali contenuti al loro interno. Il programma si chiama *Mosaic*, ed è il primo *browser* a conoscere una vasta diffusione. La disponibilità di un linguaggio per creare facilmente documenti ipertestuali in rete e di un *browser* per navigarli in maniera semplice e immediata segnano la nascita dell'applicazione che determinerà il successo globale della rete Internet, anche a livello commerciale: il *world wide web*.

Gli anni Novanta vedono dunque prima la nascita e quindi la rapidissima espansione del web, e la trasformazione di Internet in una rete globale aperta sia all'uso commerciale sia all'accesso di un numero sempre maggiore di utenti individuali. All'inizio del nuovo millennio, Internet ha ormai circa 300 milioni di utenti (diventati circa un miliardo e mezzo nel corso del 2008), con un tasso di crescita più rapido di quello di qualunque altro mezzo di comunicazione di massa (compresi telefono e televisione), e rappresenta una realtà economica e commerciale in fortissima ascesa (con decine di nuove società – le cosiddette *dot.com*, dal suffisso '.com' compreso nell'indirizzo web della maggior parte delle società private presenti in rete – capitalizzate in borsa per cifre altissime).

È da questa realtà che partirà dunque la nostra analisi, volta a indagare lo sviluppo e le prospettive di Internet nei primi anni del terzo millennio.

Il web come *medium* pervasivo

Si è già accennato alla progressiva trasformazione della convergenza al digitale in vera e propria convergenza verso la rete. Si tratta di una trasformazione di grande portata, della quale è bene capire meglio le caratteristiche.

Quello passato è stato definito, a ragione, come il secolo dei *media*. In effetti il Novecento è stato caratterizzato da un processo di moltiplicazione e differenziazione dei *media* che non ha precedenti: un

processo iniziato in realtà nella seconda metà dell'Ottocento, con la diffusione della fotografia e l'invenzione del telefono, del fonografo e del cinema, e proseguito lungo tutto il 20° sec., con l'introduzione della radio, del televisore, di audio e videocassette, del fax, della telefonia mobile, e, a partire dalla fine degli anni Settanta, con la progressiva diffusione dei *media* digitali e delle reti telematiche.

I *media* digitali tuttavia presentano – come si è già ricordato – una caratteristica del tutto nuova rispetto al panorama mediatico precedente: il fatto di basarsi tutti sugli stessi meccanismi di codifica digitale, che trasforma l'informazione – indipendentemente dalla sua natura testuale, visiva o sonora – in una serie di stati binari, rappresentabili attraverso 0 e 1 ed elaborati da un microprocessore. Ciò permette, attraverso il computer, un livello di integrazione fra codici diversi totalmente nuovo. Tradizionalmente, i supporti utilizzati per i testi scritti – nel mondo 'gutenberghiano' in primo luogo la carta stampata (libri, giornali, riviste ecc.) – erano diversi, nelle tecnologie impiegate e nei modi di fruizione, dai supporti utilizzati per i suoni (pensiamo per es. a dischi e audiocassette), da alcuni di quelli usati per le immagini (negativo fotografico, diapositive ecc.) o per i filmati (pellicola cinematografica, videocassetta ecc.). L'universo comunicativo del 'secolo dei *media*' era popolato non solo da una pluralità di supporti, ma da *media* tecnologicamente assai diversi l'uno dall'altro (si pensi alle differenze esistenti fra stampa, radio, televisione, cinema ecc.), che richiedevano strumenti di produzione e di fruizione diversi e portavano dunque alla costruzione di mercati e di stili comunicativi fortemente differenziati. L'integrazione di queste diverse risorse informative, sia dal punto di vista tecnologico, sia da quello comunicativo, sia da quello sociale, poteva risultare – e di fatto spesso risultava – assai difficile, se non del tutto impossibile.

Con il digitale questa situazione è cambiata radicalmente. Per la prima volta uno stesso oggetto informativo può integrare non solo testo e immagini statiche, ma anche suoni e video. E per la prima volta questo oggetto informativo può staccarsi da un particolare supporto e circolare liberamente attraverso le reti telematiche. Inoltre, grazie alle capacità che ha il computer di gestire informazioni in formato digitale, i contenuti informativi possono essere fruiti in maniera altamente interattiva.

La convergenza al digitale rende dunque possibile forme di comunicazione nuove, le cui caratteristiche e potenzialità stanno appena iniziando a essere

esplorate. Pur senza determinarne meccanicamente le forme, la convergenza tecnologica sembra portare dunque a una convergenza degli strumenti comunicativi, dei format, dei mercati.

Non stupirà quindi che lo sviluppo prepotente del processo di convergenza al digitale abbia portato alla fine del secolo scorso all'idea che l'ultimo strumento di comunicazione nato nel panorama dei *media*, il computer, si apprestasse non solo a 'divorare' i propri fratelli maggiori, ma anche a proporsi come piattaforma unica per la fruizione di qualunque tipo di contenuto informativo, sostituendo la televisione nella funzione di vero e proprio totem mediatico della postmodernità.

Tuttavia la lezione degli ultimi anni sembra andare in una direzione almeno parzialmente diversa. All'enorme forza centripeta della rivoluzione digitale corrisponde infatti una moltiplicazione altrettanto prepotente delle interfacce e delle situazioni di fruizione dell'informazione. Se è vero che la maggior parte dell'informazione ci arriva ormai in formato digitale, le interfacce e gli strumenti che utilizziamo per riceverla continuano a moltiplicarsi, per meglio rispondere alle esigenze di situazioni di fruizione diverse. Possiamo così, per es., guardare un film nella sala di un cinema, ma anche – se preferiamo – nel salotto di casa attraverso un impianto di *home cinema*, alla scrivania sul computer, o in mobilità attraverso un lettore multimediale portatile. Inoltre, anche se all'interno dei *media* digitali troviamo microprocessori e memorie basati su principi di funzionamento largamente comuni, una macchina fotografica digitale, un lettore MP3, un navigatore satellitare per auto, un telefono cellulare, un *decoder* satellitare, un lettore per DVD (*Digital Versatile Disc*) o per dischi Blu-Ray, una *console* per videogiochi sono comunque strumenti tecnologici diversi: la loro ibridazione porta spesso alla moltiplicazione e non all'unificazione degli strumenti che utilizziamo.

La differenziazione delle interfacce avviene tuttavia sulla base di una duplice base comune: da un lato, come si è visto, il ricorso al digitale per la codifica e l'elaborazione dell'informazione, dall'altro l'uso di Internet come piattaforma comune di scambio e condivisione. Sempre più frequentemente, infatti, Internet viene percepita, più che come un singolo *medium* con caratteristiche comunicative specifiche e uniformi, come una piattaforma di integrazione e di scambio per contenuti e format diversi, che consente al panorama vasto e differenziato di dispositivi digitali di dialogare fra loro scambiandosi informazione.

Questa convergenza verso la rete tende a renderci tutti utenti di Internet, non solo per utilizzarne le pur funzionalità comunicative per così dire 'tradizionali' (posta elettronica, navigazione su web ecc.) ma anche per coordinare e raccordare altre forme di uso dell'informazione: per scaricare – legalmente o no – la musica da ascoltare sul lettore MP3 o i film da visualizzare attraverso un lettore multimediale, per acquistare libri (su carta, ma anche in formato elettronico, per la lettura su dispositivi dedicati), per utilizzare la telefonia e videotelefonia via rete e così via. Tendono a comparire dunque forme 'trasversali' di uso della rete, non più legate esclusivamente all'uso di un computer tradizionale (un settore in rapidissima espansione è così, ad esempio, quello dei collegamenti alla rete in mobilità, attraverso dispositivi portatili).

Non vi è dubbio che, anche per questa sua funzione di piattaforma trasversale di scambio dell'informazione, nel novero dei cosiddetti nuovi *media* la rete sia quello che negli ultimi anni ha conosciuto la crescita più accentuata: il settimo rapporto CENSIS (*CEN*tro *Stu*di *Investimenti Sociali*) sulla comunicazione mostra come tra il 2001 e il 2008 gli utenti di Internet nel nostro paese siano più che raddoppiati, con una diffusione ormai assolutamente pervasiva soprattutto per quanto riguarda le giovani generazioni: la percentuale dei giovani che usano Internet in Italia ha ormai decisamente superato la soglia dell'80% (83% nel 2007). Ma ancor più significativo è l'aumento fra i giovani degli utenti abituali di Internet, passati in quattro anni dal 39,8% al 73,8%. A questa tendenza corrisponde, ancora una volta soprattutto nelle fasce di età più giovani, una diminuzione degli utenti abituali della televisione, calati di oltre dieci punti percentuali in soli tre anni.

Tendenze analoghe sono riscontrabili negli altri Paesi dell'Occidente industrializzato, rispetto ai quali l'Italia sconta anzi un certo ritardo. E i tassi di crescita della rete nel resto del mondo – in particolare nelle maggiori economie asiatiche – pur senza cancellare il cosiddetto *digital divide* (il fossato di competenze culturali e sviluppo tecnologico che separa anche in questo campo le economie più forti da quelle più deboli), sembrano avviarsi verso un aumento tanto più rilevante in quanto avviene in una situazione per altri versi assai meno favorevole.

Crescita, organizzazione e reperimento dell'informazione in rete

All'impetuoso ampliamento nella base di utenti si affianca un ancor più rilevante aumento della quantità e della tipologia di contenuti disponibili in rete. Un ampliamento legato in particolare a tre fenomeni convergenti: la progressiva digitalizzazione del patrimonio informativo preesistente; il trasferimento in rete della produzione e della gestione di gran parte della nuova informazione, sia in ambito privato sia da parte delle pubbliche amministrazioni; la crescita dei contenuti provenienti da singoli utenti (il cosiddetto UGC, *User Generated Content*, che rappresenta una delle tendenze più interessanti del nuovo web e sul quale avremo occasione di tornare in seguito).

Come è facile capire, la convergenza di questi tre fenomeni pone enormi problemi di organizzazione e di reperimento dell'informazione disponibile in rete e in particolare di quella accessibile attraverso il web. Per comprendere le strategie utilizzate per rispondere a questi problemi, occorre soffermarsi su una considerazione importante: sempre più spesso i contenuti sono inseriti in rete non in forma occasionale e disorganizzata ma attraverso l'uso di strumenti specifici di gestione (i cosiddetti CMS, *Content Management System*), o nell'ambito di basi strutturate di dati. Il problema di garantire l'organizzazione e la reperibilità di tali contenuti dipende dunque da due questioni strettamente interrelate: la disponibilità all'interno degli strumenti di gestione utilizzati di adeguate funzionalità di indicizzazione e recupero delle informazioni, e la capacità degli utenti di comprendere e utilizzare tali funzionalità.

L'esperienza sembra indicare che, come avveniva anche prima della rivoluzione digitale nelle discipline legate all'organizzazione e al reperimento dell'informazione, e in particolare in ambito bibliotecario e archivistico, il concetto chiave è in questo caso quello di *metainformazione*: gli strumenti utilizzati per l'inserimento in rete e la gestione dei contenuti informativi devono permettere l'inserimento e il recupero di metainformazione organizzata e pertinente, in forma di *metadati* standardizzati.

Ma cosa sono i metadati? Etimologicamente, il termine richiama l'idea di 'dati di secondo livello', ossia dati utilizzati per descrivere e classificare altri dati. L'esempio tipico, e quello più comunemente utilizzato per spiegare il concetto, è il catalogo di una biblioteca. Gli scaffali di una biblioteca raccolgono libri (i nostri dati di primo livello, o 'informazione primaria'), mentre il catalogo raccoglie schede che *descrivono* i libri, sulla base di caratteristiche prefissate: il nome

dell'autore, il titolo, il luogo e l'anno di pubblicazione, il numero di pagine, la collocazione negli scaffali della biblioteca (o in un sistema di classificazione astratto) e così via. Le schede di un catalogo contengono dunque dati di secondo livello, *metadati* appunto, che descrivono l'informazione primaria e aiutano a gestirla e reperirla. Dal canto suo, la biblioteca comprende *sia* i libri *sia* il catalogo: una buona raccolta di risorse informative, infatti, comprende sempre anche informazione secondaria, metadati, che permettono di descrivere, organizzare e reperire l'informazione primaria. Una biblioteca, come sappiamo, offre al proprio utente anche molte altre cose, e in particolare un insieme assai ampio di strumenti e servizi che aiutano a lavorare sia sull'informazione primaria sia sui metadati: come vedremo, anche questo aspetto di servizio – o meglio, di articolazione di servizi – è rilevante nel campo della gestione delle informazioni in rete.

Abbiamo dunque a che fare con metadati quando, volendo descrivere o gestire una risorsa informativa o un certo insieme di risorse informative, ne individuiamo e selezioniamo alcune caratteristiche al fine di costruirne un modello semplificato, funzionale ai nostri scopi e facilmente condivisibile con altri. Questo avviene, di norma, individuando all'interno dell'universo informativo di partenza (l'informazione primaria) proprietà, tratti comuni e differenze che ci aiutino a organizzarlo. Selezionare un certo insieme di metadati (e scartarne molti altri possibili) dipende sempre dai nostri scopi e implica la scelta di un particolare modello, di una particolare rappresentazione dell'informazione primaria. In un certo senso, equivale a scegliere una certa 'visione del mondo', definendo le categorie (e le relazioni) in base alle quali vogliamo organizzare l'universo informativo che ci interessa. Per questo a proposito di metadati si parla spesso di *ontologie*: un'ontologia ci dice cosa è reale per noi, quali sono le caratteristiche che scegliamo di prendere in considerazione nel nostro dominio di riferimento.

Una distinzione abbastanza diffusa è quella fra metadati *descrittivi*, *gestionali-amministrativi* e *strutturali*. Come suggeriscono i nomi, i metadati descrittivi (per es., il titolo di un libro) hanno lo scopo fondamentale di descrivere l'informazione primaria; quelli amministrativo-gestionali (per es., il responsabile dell'acquisto di un libro) hanno lo scopo principale di semplificarne la gestione, e quelli strutturali (per es., la suddivisione in sezioni di un testo) aiutano a strutturarla. Ma questa classificazione

non è né l'unica possibile, né necessariamente esaustiva o sempre soddisfacente: proprio come la scelta dei metadati stessi, la scelta di particolari classificazioni o organizzazioni dei metadati è legata al tipo di informazione primaria di cui ci stiamo occupando, ai nostri obiettivi, alla nostra 'visione del mondo'.

Così, per es., in certi contesti possono avere rilievo metadati *valutativi* (come le stelle attribuite a un film dalla recensione di un giornale) mentre in altri contesti può essere preferibile evitarli. E in effetti fra le varie distinzioni proposte in quest'ambito non manca quella fra metadati *oggettivi* (ricavabili direttamente dall'analisi dell'informazione primaria, in maniera tendenzialmente indipendente dalle opinioni o valutazioni di chi classifica) e metadati *soggettivi* (come appunto i metadati valutativi, in cui l'elemento di valutazione soggettiva è preponderante). Così come non manca una distinzione relativa alla responsabilità della metadatozione, fra metadati *primari* (forniti dall'autore della risorsa informativa), *secondari* (forniti dagli utenti) e *terziari* (forniti da servizi o organizzazioni indipendenti). O ancora la distinzione fra metadati inclusi nella risorsa informativa (*embedded metadata*) e metadati esterni alla risorsa informativa (*external metadata*).

Gli strumenti di rete hanno progressivamente sviluppato, nei primi anni del nuovo millennio, una migliore capacità di gestione dei metadati e in questo settore sono nati numerosi standard, legati a tipologie diverse di informazione primaria. Occorre sottolineare che nell'era di Internet e della rivoluzione digitale lo scopo dei metadati non è solo quello di facilitare il lavoro degli utenti, ma anche e forse soprattutto di agevolare quello dei programmi ('agenti software') che gli utenti usano per gestire e reperire informazione. È importante dunque che gli standard utilizzati in questo campo siano interoperabili e formalmente rigorosi. Un notevole passo avanti si è avuto al riguardo con la diffusione di XML (*Extensible Markup Language*), un (meta)linguaggio che permette la definizione e descrizione rigorosa di insiemi di metadati utilizzabili per accompagnare, attraverso un meccanismo di marcatura, l'informazione primaria. La maggior parte delle ontologie utilizzate per descrivere informazione in rete si avvale proprio di XML per definire in modo rigoroso e formale i propri metadati e le loro caratteristiche.

Partendo da queste basi, Berners-Lee ha prefigurato la nascita di un vero e proprio 'web semantico', nel quale, attraverso l'uso di sistemi

rigorosi e standardizzati di metadati, agenti software 'intelligenti' siano in grado di svolgere in forma automatizzata, dialogando fra loro e con gli utenti, molta parte del lavoro di organizzazione e reperimento dell'informazione, rispondendo così in maniera efficace ai problemi posti dalla crescita esponenziale della quantità di informazione disponibile in rete e dal moltiplicarsi delle sue tipologie (Berners-Lee 1999).

Se la prospettiva di un web semantico realmente sviluppato è ancora lontana, non vi è dubbio che la tendenza alla standardizzazione e all'interoperabilità degli strumenti e dei parametri descrittivi abbia ricevuto, con la crescita di Internet, un impulso notevole. I primi passi in questa direzione sono stati compiuti garantendo l'interoperabilità dei cataloghi ad accesso pubblico dei sistemi bibliotecari (OPAC, *Online Public Access Catalogue*), ma negli ultimi anni la costruzione di depositi (*repositories*) interoperabili di informazione strutturata e accompagnata da metadati standardizzati è riscontrabile in molti settori. Fra le tipologie di *repositories* più diffuse e importanti (a loro volta variamente articolate) ricordiamo le *digital libraries* (biblioteche digitali), gli *open archives* (archivi aperti), le piattaforme per la raccolta, gestione e condivisione di contenuti sonori o visivi (immagini e filmati), i sistemi per la gestione e l'accesso a basi dati di riviste e libri elettronici realizzati da molte case editrici, i LCMS (*Learning Content Management Systems*), destinati alla gestione e distribuzione di contenuti e risorse di apprendimento. Spesso i confini fra queste diverse tipologie sono labili e dipendono, più che dallo specifico strumento informatico utilizzato, dall'uso che ne viene fatto o da considerazioni di mercato. In tutti questi casi, comunque, la gestione dei metadati gioca un ruolo fondamentale e una delle caratteristiche distintive dei vari *repositories* è proprio legata ai sistemi di metadati che essi riconoscono e consentono di utilizzare.

Per rendere disponibili e ricercabili i propri contenuti, un *repository* deve dunque rendere disponibili e ricercabili i propri metadati. Ed è bene ricordare che, anche se frequentemente pensiamo ai metadati come a una sorta di involucro che circonda l'informazione primaria, non è affatto detto che questi debbano sempre essere utilizzati 'insieme' ai relativi contenuti. Proprio come le schede del catalogo di una biblioteca possono essere utilizzate anche indipendentemente dai libri che descrivono (e magari essere raccolte in pubblicazioni autonome, utili a controllare anche a distanza il patrimonio di una

biblioteca o a recuperare specifiche informazioni bibliografiche), anche i metadati in formato elettronico possono avere una vita autonoma, circolare ed essere distribuiti per disseminare le informazioni relative ai contenuti di un certo *repository* e facilitarne il reperimento.

Intuitivamente saremmo portati a pensare che il modo migliore per semplificare il reperimento e il recupero di grandi quantità di contenuti informativi sia quello di non disperderli, realizzando *repositories* più larghi e centralizzati possibile, abbastanza ricchi e completi da non richiedere all'utente eccessive peregrinazioni da un *repository* all'altro per trovare il contenuto che lo interessa.

A ben vedere, però, per garantire semplicità ed efficienza non occorre necessariamente utilizzare pochi grandi *repositories*: è anche possibile (e in molti casi è più semplice ed efficace) realizzare più depositi, magari settoriali o corrispondenti a istituzioni diverse, che però condividano l'uso di sistemi standard di metadati e di recupero dei contenuti, e rendano disponibili i loro metadati per forme di indicizzazione integrata e cooperativa. In sostanza, non occorre centralizzare i contenuti: basta standardizzare i metadati e centralizzare i servizi per la loro ricerca e indicizzazione (tenendo anche presente che in genere i metadati 'pesano' in termini di bit molto meno delle relative risorse informative primarie e possono dunque viaggiare con maggiore facilità).

È chiaro che questo modello richiede l'adozione generalizzata di standard e protocolli condivisi, sia per quel che riguarda i sistemi di metadati sia per quanto concerne le procedure per facilitarne la circolazione e la consultazione da parte di servizi esterni, sia ancora per quel che riguarda la richiesta da parte dell'utente e l'invio dei contenuti da parte del *repository*.

Un meccanismo di questo tipo si baserà dunque su una pluralità di *repositories* che fungano da *data provider*, che cioè rendano disponibili i contenuti e che *espongano* i relativi metadati in un formato standard e riconoscibile (basato in genere sull'uso di XML). Alla pluralità di *data providers* corrisponderanno pochi *service providers*, che saranno in grado di raccogliere automaticamente (*harvesting*) i metadati esposti dai vari *data providers*, di integrarli e di offrire servizi centralizzati di interrogazione e ricerca (ma anche, per es., servizi statistici o di altro tipo).

Questo modello, nato nel mondo degli *open archives* (la tipologia di *repository* utilizzata sempre più spesso da istituzioni accademiche e di ricerca per

raccogliere e distribuire in forma aperta i prodotti delle attività di ricerca: articoli, *pre-print*, *post-print* ecc.), è evidentemente applicabile a moltissimi altri tipi di *repository*, e l'uso del protocollo OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*), un insieme di convenzioni che disciplinano l'esposizione dei metadati e il loro *harvesting*, si sta estendendo dal mondo *open archive* a *repositories* di altro genere.

All'evoluzione degli strumenti di gestione dell'informazione si accompagna quella degli strumenti di ricerca in rete. I motori di ricerca (e in particolare il più importante e utilizzato fra essi, *Google*) tentano così di allargare la base dell'informazione indicizzata, aggiungendo ai siti web tradizionali – spesso assai eterogenei per affidabilità e aggiornamento – risorse informative più qualificate, a cominciare proprio da quelle accessibili attraverso *repository*. È il caso di progetti quali *Google Scholar*, che permette di svolgere ricerche limitate a siti e contenuti particolarmente autorevoli e aggiunge ai risultati la possibilità di verificare citazioni e riferimenti incrociati; *Google Book Search*, che si concentra sui contenuti provenienti dal progetto di digitalizzazione bibliotecaria avviato da Google in collaborazione con numerose biblioteche e case editrici inizialmente soprattutto nordamericane; o *Google Health*, che si propone come strumento interattivo per la consultazione di informazioni relative alla sfera della salute e della sanità.

La posizione predominante di Google nel campo della ricerca in rete ha suscitato comprensibili preoccupazioni, legate alla concentrazione nelle mani di una sola azienda di un enorme potere informativo, alla prevalente attenzione verso le risorse informative in inglese e al carattere 'chiuso' di alcune delle procedure utilizzate, in particolare nell'ambito della digitalizzazione bibliotecaria. Ed è proprio in quest'ultimo campo che numerosi progetti nazionali e sovranazionali (compreso quello della *European Library*) si propongono di affiancare, integrare o sostituire, in forma più aperta e con una migliore validazione delle metodologie, il lavoro di Google.

Va infine ricordato come alla diffusione di contenuti digitali in rete, e al miglioramento degli strumenti per la loro gestione e per il loro reperimento, si accompagnino notevoli problemi legati alla gestione dei relativi diritti (Digital right management, DRM), a cominciare dal diritto di proprietà intellettuale e dal copyright. Di fatto, non disponiamo ancora di modelli e di strumenti realmente adeguati a una gestione dei

diritti in grado di rispondere alle sfide dell'epoca della riproducibilità digitale, bilanciando e garantendo adeguatamente gli interessi degli autori, degli editori e degli utenti. In questa situazione, a soluzioni basate su forme di controllo rigido della duplicazione e dell'uso dei contenuti digitali (e su sistemi di sanzioni tanto severe quanto di fatto difficilmente applicabili) si sono affiancate soluzioni basate sulla circolazione aperta delle informazioni (*open access*, *open content*), in particolare – ma non unicamente – nel caso di contenuti di particolare rilievo scientifico o sociale.

Gli strumenti del nuovo web e l'uso di Internet come spazio di interazione sociale

Al miglioramento degli strumenti legati alla gestione e al reperimento su web di informazione fortemente strutturata si è accompagnato negli ultimi anni anche un cambiamento degli strumenti utilizzati per l'interazione con gli utenti e per la raccolta e la gestione di contenuti informativi più eterogenei, spesso provenienti dagli utenti stessi. Questa evoluzione ha seguito il passaggio dal web pionieristico del primo decennio (a grandi linee, quello cresciuto fra il 1993 e il 2003) al cosiddetto web 2.0.

Da qualche anno a questa parte, tutti gli strumenti e i siti web che vogliono proporsi come di successo e innovativi si presentano come incarnazione del web 2.0 e a seconda delle situazioni e delle necessità declinano quest'espressione in un verso o nell'altro. Come molti termini alla moda, dunque, l'espressione 'web 2.0' ha, più che un significato preciso e ben definito, una connotazione abbastanza generale e in parte generica.

La sua introduzione si deve probabilmente a un'intuizione della dinamica casa editrice specializzata O'Reilly e in particolare del suo vicepresidente Dale Dougherty. L'esplosione della bolla speculativa nata attorno alle aziende legate al mondo Internet (enormemente sopravvalutate fra il 1999 e il 2000, le azioni delle *dot.com* perdono nel corso del 2001 buona parte del loro valore, trascinando nel crollo i valori dei listini azionari di mezzo mondo e in particolare del listino tecnologico statunitense NASDAQ, *National Association of Securities Dealers Automated Quotation*) non ha comportato affatto, notava Dougherty, una perdita di rilievo o una 'crisi' di Internet, ma si è accompagnata semmai a un suo ulteriore sviluppo, marcato dalla nascita di tendenze e di strumenti nuovi. È proprio a questo sviluppo che la O'Reilly ha dedicato nel 2004, in collaborazione con

MediaLive International, un incontro di studio, denominato *Web 2.0 Conference* (da allora l'evento si ripete con cadenza annuale). Alla base c'è l'idea che il web si sia ormai trasformato, da superficie sostanzialmente 'piatta' su cui appoggiare informazioni, in una sorta di piattaforma applicativa condivisa, all'interno della quale le informazioni possono essere non solo distribuite ma anche create ed elaborate collettivamente, eventualmente con l'aiuto di apposite *web applications* utilizzabili direttamente dall'interno del proprio programma di navigazione.

A fare emergere questa idea sono stati probabilmente e in primo luogo due sviluppi del 'panorama di rete' avvenuti fra il 2001 e il 2004. Da un lato la diffusione dei blog (abbreviazione di weblog): sorta di 'diari in rete' realizzati attraverso strumenti di gestione che rendono estremamente semplice la realizzazione e l'aggiornamento continuo del sito. Nel luglio 2004 i blog erano già oltre tre milioni, e prefiguravano non solo una molteplicità di punti di distribuzione individuali di informazione ma una vera e propria 'blogosfera', rete di contenuti caratterizzata dalla possibilità di una circolazione personalizzata e dinamica e di una rielaborazione parzialmente automatica dell'informazione. Dall'altro la nascita dei primi siti in cui la condivisione di risorse era esplicitamente orientata alla creazione di relazioni sociali fra gli utenti (il cosiddetto *social networking*, inaugurato dal sito *MySpace* nel 2003 e cresciuto attraverso siti come *Facebook*, la piattaforma di *social network* creata nel 2004 da Mark Zuckerberg e acquisita in seguito dalla Microsoft). Nel febbraio 2004 era nata anche *Flickr*, piattaforma per la condivisione di immagini che permetteva agli utenti di 'etichettare' liberamente le fotografie archiviate in rete, secondo il principio del *social tagging*. un'altra delle caratteristiche che, come vedremo, diventeranno tipiche del web 2.0.

Il web come piattaforma, dunque. Ma per quali contenuti? Può essere a questo punto utile cercare di riassumere, almeno indicativamente, quelle che sono state finora le aree e le idee più frequentemente collegate al 'nuovo web', e delineare le prospettive e i punti di vista che portano a privilegiare questa o quella concezione del web 2.0.

Una prima prospettiva è legata all'intervento più attivo degli utenti come produttori di contenuti: a determinarla è la crescita impetuosa del già ricordato UGC, il contenuto generato dagli utenti. Una rivoluzione nata proprio con la crescita della

blogosfera, e proseguita negli anni successivi con lo sviluppo di strumenti che facilitano l'immissione in rete e la condivisione anche di contenuti non testuali: i *podcast*, flussi informativi costituiti da contenuti audio, che gli utenti possono seguire attraverso appositi programmi e ascoltare anche in mobilità attraverso lettori MP3, e piattaforme quali la già ricordata *Flickr* per la condivisione in rete di immagini e *YouTube* per la condivisione di video.

Una caratteristica specifica che il contenuto generato dagli utenti assume nel nuovo web è la possibilità di una sua diffusione quasi virale, attraverso strumenti che consentono di incorporarlo (*embedding*) in siti diversi da quello di provenienza. A favorire questa tendenza è la diffusione di contenuti in forma di flussi organizzati (*feed*), generati automaticamente dalle piattaforme di pubblicazione in rete e suscettibili di essere raccolti e aggregati da specifiche applicazioni o da siti terzi.

Come abbiamo visto, con il moltiplicarsi dei produttori di contenuti in grado di inserire facilmente in rete informazione strutturata cresce naturalmente anche l'esigenza di descrivere e organizzare questa informazione per garantirne la reperibilità. E siamo qui a una seconda prospettiva di particolare rilievo nel parlare di web 2.0, quella legata alla classificazione, alla ricerca e al reperimento dell'informazione.

In questa prospettiva il web 2.0 nasce, se non in opposizione, almeno in una situazione di qualche tensione rispetto al progetto del cosiddetto web semantico di cui si è parlato nella sezione precedente. Come si è detto, anche il web semantico nasce dalla necessità di trovare strumenti per migliorare la classificazione e la reperibilità dell'informazione; ma il progetto di web semantico è strettamente legato all'idea che il lavoro di organizzazione e gestione dell'informazione debba essere basato su descrizioni fortemente standardizzate e formalizzate. I sistemi di classificazione dell'informazione alla base del web semantico sono – o dovrebbero essere – *ontologie formali*: griglie di classificazione elaborate da esperti del settore, espresse in maniera uniforme e rigorosa e associate all'informazione primaria attraverso l'uso di linguaggi e formalismi a loro volta rigidamente strutturali e ben definiti.

Al contrario, il web 2.0 mette gli utenti al centro anche del processo di classificazione. Al posto di sistemi classificatori predefiniti, gli utenti che immettono informazione in rete e quelli che la usano sono invitati ad aggiungere all'informazione primaria delle etichette descrittive (*tags*) totalmente libere, sulla

base della loro percezione della natura dell'informazione stessa e dei contesti di sua potenziale utilità: il cosiddetto *social tagging*.

Questa strada è davvero percorribile? E, se lo è, è anche preferibile rispetto ai meccanismi di classificazione più affidabili e 'professionali' che rappresentavano un po' il presupposto del web semantico? Per un verso, dovremmo rispondere sicuramente che è preferibile una classificazione rigorosa e controllata. Ma va comunque tenuto presente che – pur se assai meno rigorosa – anche la via del *social tagging*, più rapida e facile, sembra comunque funzionare abbastanza bene, almeno per certi scopi e relativamente a determinate tipologie di contenuti.

Sistemi di *social tagging* (si usa talvolta a questo riguardo anche il termine *folksonomy*), etichettatura libera e diffusa dei contenuti informativi, sono ormai assai diffusi in rete. Così, quasi tutti i sistemi per la gestione di blog mettono a disposizione degli autori la possibilità di etichettare e classificare liberamente i singoli messaggi. La stessa funzione è svolta, per quanto riguarda l'audio, dai sistemi *podcast* più diffusi; per le immagini dai sistemi come *Flickr* e per i video dai sistemi come *YouTube*. È possibile inoltre classificare in questo modo anche i siti web, creando le proprie liste di siti preferiti e associando a ciascuno di essi una o più etichette scelte in maniera completamente libera.

Certo, se queste classificazioni libere e poco strutturate restano del tutto prive di qualunque meccanismo di validazione, che assicuri un minimo di uniformità e affidabilità, il rischio della confusione e dell'arbitrio è fortissimo. Chi assicura che le etichette che io ho usato per classificare una certa informazione corrispondano a quelle che avrebbe utilizzato qualcun altro, e soprattutto a quelle che userebbe un altro utente al momento di ricercare i miei contenuti, o contenuti simili a quelli immessi da me? Il grosso vantaggio delle ontologie controllate è proprio la standardizzazione e l'affidabilità dei sistemi di classificazione, di norma elaborati da esperti.

Anche il *social tagging*, però, ha una freccia importante al proprio arco: la sua natura *sociale*. Le etichette con le quali si classificano i contenuti informativi sono sì libere, ma non vengono scelte nel vuoto: sono attribuite utilizzando piattaforme che hanno centinaia di migliaia di altri utenti, dunque in un contesto sociale assai ricco. E i sistemi di gestione delle informazioni possono confrontare tra loro i comportamenti classificatori, segnalare

automaticamente le etichette più frequenti per certe tipologie di contenuti (così come quelle più rare e anomale), contribuire insomma al *filtraggio collaborativo* dei metadati.

Esistono anche altri sistemi per migliorare il *social tagging*, e in generale per cercare di conciliare questo con le ontologie controllate, ma il filtraggio collaborativo è sicuramente il più importante e rappresenta la terza prospettiva che è opportuno tenere presente nel cercare di individuare le caratteristiche specifiche del nuovo web. In questo caso, il filtraggio collaborativo è utilizzato per selezionare e far emergere comportamenti classificatori, e dunque applicato ai metadati, ma è usato dal nuovo web in maniera generalizzata, anche con riferimento alle informazioni primarie. Il suo uso più frequente è legato al confronto fra i gusti degli utenti. Così per es. esistono siti che permettono agli utenti di indicare le loro preferenze in fatto di musica e musicisti (o film, interpreti e registi) e che automaticamente segnalano brani e autori a esse rispondenti e già indicati da altri utenti con gli stessi gusti. La principale libreria in rete, *Amazon* (che ha rappresentato una vera e propria palestra di sperimentazione per strumenti di questo genere) e molti altri siti fanno lo stesso per quanto riguarda i libri.

Come si è detto, il filtraggio collaborativo dell'informazione è basato sulla dimensione sociale della condivisione di risorse via web. Proprio l'accentuazione del rilievo di tale dimensione costituisce la quarta prospettiva sotto cui considerare il web 2.0: quella legata ai siti e agli strumenti di *social networking*. Caratteristica tipica degli strumenti di *social networking* è il tentativo di creare relazioni fra gli utenti, permettendo a ciascuno di evidenziare i propri interessi, i propri gusti e di collegare il proprio profilo a quello di amici e conoscenti (ma anche di eventuali sconosciuti che risultassero avere profilo e interessi tanto simili al proprio da renderli interessanti). In realtà, le 'reti sociali' che questi siti vogliono aiutare a creare hanno una duplice dimensione: reti di individui e reti di contenuti informativi legati agli individui che ne sono produttori o consumatori. Molto spesso esiste poi una terza dimensione, rappresentata non dai singoli utenti ma da gruppi o comunità di utenti legati da interessi comuni.

Tutti i siti di *social networking* hanno fra i propri strumenti principali la 'pagina di profilo' del singolo utente; spesso, oltre a proporre una fotografia e alcuni

dati su di sé e sui propri interessi, nonché l'elenco dei propri 'amici' all'interno del sistema, queste pagine possono essere fortemente personalizzate attraverso l'aggiunta di immagini, musica, video, e attraverso *widget*, programmini utilizzabili via web che permettono, per es., di far sapere agli altri utenti quale brano musicale si sta ascoltando in quel momento, o che libro si sta leggendo. Le funzionalità dei *widgets* di questo genere sono moltissime, e lo sviluppo di *widgets* utili e interessanti (spesso demandato agli stessi utenti attraverso la distribuzione di appositi strumenti di programmazione semplificata) rappresenta uno dei campi di maggior interesse del mondo del *social networking*.

Per valutare il rilievo che il *social networking* sta assumendo nel nuovo web, basti considerare che i due siti più popolari in questa categoria, i già ricordati *MySpace* e *Facebook*, avevano complessivamente a inizio 2009 quasi 300 milioni di utenti. Va osservato, inoltre, che strumenti orientati al *social networking* si stanno diffondendo anche al di fuori dei siti 'specializzati'. Così, per es., siti come *Flickr* o *YouTube* permettono di creare pagine personali di autopresentazione degli utenti, e di costruire reti di amici e gruppi di interesse.

Nei prossimi anni i siti di *social network* potranno probabilmente integrarsi con un altro sviluppo interessante: l'interazione fra utenti all'interno di ambienti virtuali tridimensionali. L'esempio più noto è rappresentato, attualmente, dalla piattaforma *Second Life*.

Solo un cenno, in questa sede, alla quinta prospettiva rilevante per il web 2.0, quella legata ai presupposti tecnici per poter realizzare siti web in grado di interagire con gli utenti attraverso vere e proprie applicazioni di rete. Alla base vi è in primo luogo una tecnologia di programmazione, AJAX (*Asynchronous JavaScript and XML*), sulla quale sarebbe troppo complesso soffermarsi. Ma è importante sapere che, per i programmatori esperti, AJAX è senz'altro una delle dimensioni più importanti del web 2.0.

Una sesta prospettiva che esula dal campo degli strumenti e delle tecnologie, ma non è meno rilevante per comprendere la complessa connotazione del termine 'web 2.0', è legata al design. L'evoluzione dei nuovi strumenti web ha infatti coinciso con lo sviluppo di siti dal design semplificato e ripensato in funzione dell'immediatezza dell'interazione, caratterizzati da una forte riduzione del numero di elementi grafici e di contenuti informativi – in particolare testuali – presenti

in una stessa pagina (per es., sparisce spesso la distribuzione in colonne) e da un largo uso di colori pastellati, riflessi, e icone di grandi dimensioni e di forte impatto.

Infine, un'ultima prospettiva da tener presente è quella legata alla *georeferenziazione* delle risorse informative. L'uso dei sistemi di posizionamento satellitare (GPS, *Global Position System*), che permettono di determinare con grande precisione dove è stata scattata una fotografia, dove è stato ripreso un video, dove si svolge un certo evento, in quale strada si trova un negozio o dove si trova una persona in un determinato momento, costituisce il primo presupposto di questo sviluppo. A esso si associa la disponibilità in rete di mappe estremamente dettagliate, che possono essere 'incorporate' in altre applicazioni: anche in questo caso è stata Google, con la distribuzione di *Google Earth*, a fornire per prima uno strumento di facile uso e larga diffusione. L'insieme di questi fattori ha favorito la diffusione pervasiva in rete di forme di georeferenziazione legate a contenuti informativi di ogni genere. Uno sviluppo che porta a collegare in maniera assai più stretta che in passato i territori virtuali della geografia di rete e quelli reali di un globo che si rivela, anche per questa via, sempre più interconnesso.

Bibliografia

- T. BERNERS-LEE, *Weaving the web*, London 1999 (trad. it. *L'architettura del nuovo web*, Milano 2001).
- M. CASTELLS, *The rise of the network society. The information age: economy, society and culture*, 1, Malden-Oxford (Mass.) 2000² (trad. it. *La nascita della società in rete*, Milano 2008).
- L. MANOVICH, *The language of new media*, Cambridge (Mass.) 2002 (trad. it. Milano 2002).
- H. RHEINGOLD, *Smart mobs. The next social revolution*. Cambridge (Mass.) 2002 (trad. it. *Smart mobs. Tecnologie senza fili, la rivoluzione sociale prossima ventura*, Milano 2003).
- A.-L. BARABÁSI, *Linked: How everything is connected to everything else and what it means for business, science,*

and everyday life, New York 2003 (trad. it. *Link. La scienza delle reti*, Torino 2004).

M. CALVO, F. CIOTTI, G. RONCAGLIA et al., *Internet 2004. Manuale per l'uso della rete*, Roma-Bari 2003.

D. GILLMOR, *We the media. Grassroots journalism by the people, for the people*, Beijing-Sebastopol (Cal.) 2004.

C. GNOLI, V. MARINO, L. ROSATI, *Organizzare la conoscenza. Dalle biblioteche all'architettura dell'informazione per il Web*, Milano 2006.

G.P. LANDOW, *Hypertext 3.0. Critical theory and new media in an era of globalization*, Baltimore 2006.

C.L. BORGMAN, *Scholarship in the digital age. Information, infrastructure and the internet*, Cambridge (Mass.) 2007.

Si veda inoltre:

Architecture of the World Wide Web, vol. I, a cura del World Wide Web Consortium, <http://www.w3.org/TR/webarch/>, 2004 (14 ott. 2008).