

LA STORIA DI INTERNET

Negli anni novanta Internet si afferma ed esplose definitivamente, sia come nuova forma di comunicazione, sia come fenomeno sociale e culturale.

Internet non è solo un mezzo ed uno strumento di comunicazione, ma un vero e proprio MEDIA di MEDIA, che offre con i suoi contenuti multimediali (YOU TUBE, YOU STREAM, etc.) un'occasione di svago e divertimento, con i suoi contenuti più classici (WIKIPEDIA, Websites, etc.) uno strumento di lavoro e di studio e con i suoi servizi di comunicazione (SKYPE, VOIP, INSTANT MESSAGING, etc.) il nuovo strumento di comunicazione integrata.

Una definizione corretta, ma soprattutto stabile, per Internet non esiste, del resto per molti anni si è detto, in modo provocatorio, "Internet non esiste", ma ormai le cose sono cambiate e quasi tutti i gesti della nostra vita quotidiana conducono in qualche modo alla grande rete.

Con una terminologia molto generica e poco approfondita potremmo affermare che Internet è il collegamento attraverso una qualsiasi rete di comunicazione di tutti i computer (mini, personal, palmari, smartpone, etc.) del mondo che adottano un certo linguaggio, ma se volessimo essere un po' più precisi dovremmo dire tutti i dispositivi elettronici collegati e collegabili, tuttavia la realtà delle cose è molto più complessa, ma per adesso accontentiamoci di questa definizione e cerchiamo di capire come si è arrivati ad Internet così come la conosciamo oggi.

Nel 1968 il National Physical Laboratory realizzò una prima rete telematica, basata su principi rivoluzionari rispetto a quello che era il tradizionale modo di pensare alle reti di telecomunicazioni.

L' *ARPA* (*Advanced Research Project Agency*) Agenzia per i progetti avanzati del dipartimento di Difesa decise di sponsorizzare il progetto, per tutte le applicazioni militari e strategiche che esso comportava.

Nell'anno successivo l'University of California a Los Angeles, lo Standford Research Institute, l'University of California a Santa Barbara e l'University of Utah aprirono quattro nodi per questa rete che fu chiamata *ARPANET*, in onore del principale finanziatore.

Ufficialmente *ARPANET* rappresenta la prima rete per lo scambio di dati con grande estensione geografica, ma il personal computer non esisteva ancora, gli strumenti di calcolo erano delle grosse macchine, e non è difficile comprendere che *ARPANET*, progetto di ricerca del Dipartimento della Difesa, nasce in realtà con l'obiettivo di diventare una rete indistruttibile, capace di continuare a lavorare anche in condizioni estreme, come ad esempio una guerra nucleare.

Nel momento in cui *ARPANET* inizia ad essere utilizzata, i centri di calcolo comunicano tra loro come previsto, ma i ricercatori avendo per le mani un nuovo e potente canale di comunicazione, si scambiano testi e messaggi elettronici con estrema facilità e andando al di là degli scopi e dei compiti del loro lavoro inventano di fatto la posta elettronica.

Durante gli anni Settanta si collegarono ad *ARPANET* tutte le reti universitarie e quelle di ricerca, vennero messi a punto dei protocolli di rete, un insieme di regole predefinite alle quali i diversi calcolatori dovevano attenersi per parlare fra loro e per comprendersi.

TCP/IP (*Transmission Control Protocol / Internet Protocol*) è il nome della versione definitiva di questo protocollo, e che ancora oggi rimane alla base delle comunicazioni via Internet, esso è in sostanza un

linguaggio comune fornito a tutti gli utenti della rete per comunicare, o, in altre parole, un metodo per scambiare informazioni e comprenderle.

- Nel 1983 termina la convivenza - sullo stesso mezzo di trasmissione - tra le strutture di ricerca universitaria e quelle di ricerca militare, la sezione militare di ARPANET si separa e nascono due nuove entità: MilNet, la parte militare della rete, e NSFNET (National Science Foundation Network), parte universitaria e di ricerca, alla quale si erano ormai collegate moltissime università anche fuori degli USA e praticamente in tutto il mondo.

L'utilizzo del protocollo di trasmissione TCP/IP favorì la connessione a NSFNET delle reti che lo desideravano, e questo nuovo insieme di reti cominciava già ad essere chiamato Internet, di questo nucleo storico possiamo citare i seguenti enti e centri di ricerca: John von Neumann National supercomputer Center, Princetown, Newjersey; San Diego Supercomputer Center, University of California; National Center for Supercomputing Applications, University of Illinois; Cornell National Supercomputer Facility, Cornell University; Pittsburg Supercomputer Centeir, gestito da Westinghouse Electric Corporation, Carnegie Mellon University e University of Pittsburg; The Scientific Computing Division del National Center for Atmospheric Research, Boulder, Colorado.

Grandi investimenti consentirono a questa rete, che già era Internet, di ampliare la sua estensione e capillarità, migliorando qualità e velocità di trasmissione, portandola a diventare quello che oggi è sotto gli occhi di tutti.

- Di recente la *NSF* ha smesso di finanziare la spina dorsale di Internet che ora è gestita da un lato dalle università di tutto il mondo, e dall'altro da grosse imprese private che portano nei vari continenti collegamenti ad altissima velocità, chiedendo contributi agli utenti che vogliono collegarsi, in generale i contributi richiesti per la connessione sono forfetari e non a traffico come nella telefonia tradizionale. In quasi tutti i paesi del mondo ci sono degli organi di controllo, degli enti, che hanno il compito di sovrintendere al corretto utilizzo della rete che per sua natura tende ad essere sregolata ed anarchica, in Italia abbiamo il *GARR* (*Gruppo di Armonizzazione delle Reti di Ricerca*)

Dire che oggi Internet è la rete delle reti che legano i calcolatori di tutto il mondo fra di loro è corretto, ma incompleto, certamente ci sono migliaia di reti connesse fra loro, ma oggi Internet è anche una realtà di servizi, risorse, affari, enti e persone, un'enorme banca dati di quasi tutto lo scibile.

Per capire la portata del fenomeno Internet cominceremo con un paio di aneddoti e proseguiremo poi con delle cifre e dei dati molto interessanti.

Il 2 gennaio 1969 ci fu l'appalto di gara che assegnava l'incarico per la realizzazione del primo embrione della rete di cui si è parlato in precedenza, e le più grandi aziende informatiche non partecipano alla gara, secondo i loro esperti il compito da svolgere era praticamente impossibile.

Nel 1943 il presidente di una grande azienda informatica dichiarò: "il mercato mondiale può assorbire al massimo quattro o cinque computer", nel 1997 ci sono almeno 300 milioni di computer nel mondo e di questi si stima che almeno un quarto siano collegati ad Internet.

Il **World Wide Web** (letteralmente "rete di grandezza mondiale"), abbreviato **Web**, sigla **WWW**, è uno dei principali servizi di **Internet** che permette di navigare e usufruire di un insieme vastissimo di contenuti amatoriali (**multimediali** e non) collegati tra loro attraverso legami (**link**), e di ulteriori servizi accessibili a tutti o ad una parte selezionata degli utenti di Internet. Questa facile reperibilità di **informazioni** è resa possibile oltre che dai **protocolli di rete** anche dalla presenza, diffusione, facilità d'uso ed efficienza dei **motori di ricerca** e dei **web browser** in un modello di architettura di rete definito **client-server**.

Caratteristica principale della *rete Web* è che i **nodi** che la compongono sono tra loro collegati tramite i cosiddetti *link* (collegamenti ipertestuali), formando un enorme **ipertesto** e i suoi servizi possono essere resi disponibili dagli stessi utenti di Internet.[3] Per quanto riguarda i contenuti, quindi, il *Web* possiede la straordinaria peculiarità di offrire a chiunque la possibilità di diventare **editore** e, con una spesa estremamente esigua, di raggiungere un pubblico potenzialmente vastissimo distribuito in tutto il mondo.[4]

Il *Web* è stato inizialmente implementato da **Tim Berners-Lee** mentre era ricercatore al **CERN**, sulla base di sue idee e di un suo collega, **Robert Cailliau**, e oggi gli standard su cui è basato, in continua evoluzione, sono mantenuti dal *World Wide Web Consortium* (W3C: www.w3c.org).

La nascita del *Web* risale al 6 agosto 1991, giorno in cui Berners-Lee mise *on-line* su Internet il primo sito *Web*.^[5] Inizialmente utilizzato solo dalla **comunità scientifica**, il 30 aprile 1993 il CERN decide di rendere pubblica la tecnologia alla base del *Web*. A tale decisione fa seguito un immediato e ampio successo del *Web* in virtù della possibilità offerta a chiunque di diventare **editore**, della sua efficienza e, non ultima, della sua semplicità. Con il successo del *Web* ha inizio la crescita esponenziale e inarrestabile di Internet ancora oggi in atto, nonché la cosiddetta "era del *Web*". L'indirizzo della pagina (descrizione del progetto CERN) era

Il *Web* è uno spazio **elettronico** e **digitale** di Internet destinato alla pubblicazione di contenuti multimediali (testi, immagini, audio, video, ipertesti, ipermedia, ecc.) nonché uno strumento per implementare particolari servizi come ad esempio il **download** di software (programmi, dati, applicazioni, videogiochi, ecc.). Tale spazio elettronico e tali servizi sono resi disponibili attraverso particolari **computer** di Internet chiamati **server web**.

Chiunque disponga di un computer, di un **accesso ad Internet**, degli opportuni programmi e del cosiddetto **spazio web**, porzione di memoria di un server web destinata alla memorizzazione di contenuti web e all'implementazione di servizi web, può, nel rispetto delle leggi vigenti nel **Paese** in cui risiede il server web, pubblicare contenuti multimediali sul *Web* e fornire particolari servizi attraverso il *Web*. I contenuti del *Web* sono infatti costantemente *on-line* quindi costantemente fruibili da chiunque disponga di un computer, di un accesso a Internet, e degli opportuni programmi (in particolare del cosiddetto **browser web**, il programma che permette, come si dice in gergo, di "navigare" nel *Web*, cioè di fruire dei contenuti e dei servizi del *Web*.)

Non tutti i contenuti e i servizi del *Web* sono però disponibili a chiunque in quanto il proprietario dello spazio web, o chi ne ha delega di utilizzo, può renderli disponibili solo a determinati utenti, gratuitamente o a pagamento, utilizzando il sistema degli **account**.

I contenuti principali del *Web* sono costituiti da **testo** e **grafica** rappresentati in un insieme ristretto di standard definito dal **W3C**. Tali contenuti sono quelli che tutti i browser web devono essere in grado di fruire autonomamente, cioè senza software aggiuntivo.

I contenuti pubblicati sul *Web* possono essere però di qualunque tipo e in qualunque standard. Alcuni di questi contenuti sono pubblicati per essere fruiti attraverso il **browser web** e, non essendo in uno degli standard appartenenti all'insieme definito dal W3C, per poterli fruire attraverso il browser web questo deve essere integrato con i cosiddetti **plug-in**, software che integrano le funzionalità di un programma i quali, per quanto riguarda il browser web, normalmente sono scaricabili dal *Web*. Il resto dei contenuti del *Web* è utilizzabile con programmi autonomi. Ad esempio si può trattare di un **file eseguibile** sul **sistema operativo** che si sta utilizzando o di un documento di testo in formato **Microsoft Word**.

I contenuti del *Web* sono organizzati nei cosiddetti **siti web** a loro volta strutturati nelle cosiddette **pagine**

*web*le quali si presentano come composizioni di testo e/o grafica visualizzate sullo schermo del computer dal browser web. Le pagine web, anche appartenenti a siti diversi, sono collegate fra loro in modo non sequenziale attraverso i cosiddetti *link* (anche chiamati *collegamenti*), parti di testo e/o grafica di una pagina web che permettono di accedere ad un'altra pagina web, di scaricare particolari contenuti, o di accedere a particolari funzionalità, cliccandoci sopra con il *mouse*, creando così un *ipertesto*.

Tutti i siti web, sono identificati dal cosiddetto *indirizzo web*, una sequenza di caratteri univoca chiamata in termini tecnici *URL* che ne permette la rintracciabilità nel Web.

Non è previsto un indice aggiornato in tempo reale dei contenuti del Web, quindi nel corso degli anni sono nati ed hanno riscosso notevole successo i cosiddetti *motori di ricerca*, siti web da cui è possibile ricercare contenuti nel Web in modo automatico sulla base di *parole chiave* inserite dall'utente, e i cosiddetti *portali web*, siti web da cui è possibile accedere ad ampie quantità di contenuti del Web selezionati dai redattori del portale web attraverso l'utilizzo di motori di ricerca o su segnalazione dei redattori dei siti web.

Oltre alla pubblicazione di contenuti multimediali il Web permette di offrire servizi particolari implementabili dagli stessi utenti del Web. I servizi implementabili sono innumerevoli, in pratica limitati solo dalla *velocità* della *linea di telecomunicazioni* con cui l'utente e chi fornisce il servizio sono collegati e dalla *potenza di calcolo* dei loro computer. Di seguito quindi sono elencati solo quelli contraddistinti da una denominazione generica:

- *download*: la distribuzione di *software*;
- *web mail*: la gestione della casella di *posta elettronica* attraverso il Web;
- *streaming*: la distribuzione di *audio/video* in tempo reale;
- *web TV*: la *televisione* fruita attraverso il Web;
- *web radio*: la *radio* fruita attraverso il Web;
- *web chat*: la comunicazione testuale in tempo reale tra più utenti di Internet, tramite pagine web;

Il Web è *implementato* attraverso un insieme di *standard*, i principali dei quali sono i seguenti:

- *HTML* (e suoi derivati): il *linguaggio di markup* con cui sono scritte e descritte le pagine web;
- *HTTP*: il *protocollo di rete* appartenente al *livello di applicazione* del *modello ISO/OSI* su cui è basato il Web;
- *URL*: lo schema di identificazione, e quindi di rintracciabilità, dei contenuti e dei servizi del Web.

La peculiarità dei contenuti del Web è quella di non essere memorizzati su un unico computer ma di essere *distribuiti* su più computer, caratteristica da cui discende efficienza in quanto non vincolati ad una particolare localizzazione fisica. Tale peculiarità è realizzata dal protocollo di rete HTTP il quale permette di vedere i contenuti del Web come un unico insieme di contenuti anche se fisicamente risiedono su una moltitudine di computer di *Internet* sparsi per il *planeta*.

La visione di una *pagina web* inizia digitandone l'*URL* nell'apposito campo del *browser web* oppure cliccando su *uncollegamento ipertestuale* presente in una pagina web precedentemente visualizzata o in altra risorsa come ad esempio un'*e-mail*. Il browser web a quel punto dietro le quinte inizia una serie di messaggi di comunicazione con il *web server* che ospita quella pagina con lo scopo di visualizzarla sul terminale utente.

Per prima cosa la porzione di server-name dell'*URL* è risolta in un *indirizzo IP* usando il database globale e distribuito conosciuto come *Domain Name System* (in sigla *DNS*). Questo indirizzo IP è necessario per inviare e ricevere *pacchetti* dal *server web*.

A questo punto il browser richiede le informazioni inviando una richiesta a quell'indirizzo. In caso di una tipica pagina web, il testo **HTML** di una pagina è richiesto per primo ed immediatamente **interpretato** dal browser web che, successivamente, richiede eventuali immagini o file che serviranno per formare la pagina definitiva.

Una volta ricevuti i file richiesti dal web server, il browser **formatta** la pagina sullo schermo seguendo le specifiche **HTML**, **CSS**, o di altri linguaggi web. Ogni immagine e le altre risorse sono incorporate per produrre la pagina web che l'utente vedrà.

La prima proposta di un sistema **ipertestuale** si può far risalire agli studi di **Vannevar Bush**, poi pubblicati nell'articolo *As We May Think* (in italiano "Come potremmo pensare") del **1945**.

La data di nascita del **World Wide Web** viene comunemente indicata nel 6 agosto **1991**, giorno in cui l'**informatico** inglese **Tim Berners-Lee** pubblicò il primo sito web dando così vita al fenomeno "WWW" (detto anche "della tripla W").

L'idea del World Wide Web era nata due anni prima, nel **1989**, presso il **CERN** (Conseil Européen pour la Recherche Nucléaire) di **Ginevra**, il più importante **laboratorio** di fisica **europeo**. Il ricercatore inglese fu colpito da come alcuni colleghi italiani usavano trasmettere informazioni tramite linea telefonica da un piano all'altro dell'istituto visualizzando informazioni tramite video. Il 12 marzo **1989** Tim Berners-Lee presentò infatti al proprio supervisore il **documento** *Information Management: a Proposal*, una cui copia è esposta presso il CERN, che fu valutato «vago ma interessante». Alla sua base vi era il progetto dello stesso Berners-Lee e di un suo collega, **Robert Cailliau**, di elaborare un software per la condivisione di **documentazione scientifica** in **formato elettronico** indipendentemente dalla **piattaforma informatica** utilizzata, con il fine di migliorare la **comunicazione**, e quindi la **cooperazione**, tra i ricercatori dell'istituto. A lato della creazione del software, iniziò anche la definizione di standard e protocolli per scambiare documenti su **reti di calcolatori**: il linguaggio **HTML** e il **protocollo di rete HTTP**.

Questi standard e protocolli supportavano inizialmente la sola gestione di pagine **HTML statiche**, vale a dire file ipertestuali -preparati precedentemente- visualizzabili e, soprattutto, navigabili utilizzando opportune applicazioni (**browser web**).

Dopo i primi anni in cui era stato usato solo dalla comunità scientifica, il 30 aprile **1993** il CERN decise di mettere il WWW a disposizione del pubblico rinunciando ad ogni **diritto d'autore**. La semplicità della **tecnologia** decretò un immediato successo: in pochi anni il WWW divenne la modalità più diffusa al mondo per inviare e ricevere dati su Internet, facendo nascere quella che oggi è nota come "**era del web**".

Per superare le limitazioni del progetto iniziale, furono subito definiti strumenti capaci di generare pagine **HTML dinamiche** (ad es. utilizzando dati estratti da un **database**). La prima soluzione di questo tipo furono le **CGI** (*Common Gateway Interface*). Attraverso una CGI è possibile richiedere ad un **web server** di invocare un'**applicazione** esterna e presentare il risultato come una qualsiasi pagina HTML. Questa soluzione, sebbene molto semplice da realizzare, presenta numerose limitazioni di progetto (l'applicativo esterno viene eseguito ad ogni richiesta utente e non è prevista alcuna **ottimizzazione**, non vi è alcuna gestione dello **stato della sessione**, etc.).

Per dare al web una maggiore interattività e dinamicità sono state perseguite due strade. Da un lato sono state aumentate le funzionalità dei **browser** attraverso un'evoluzione del linguaggio HTML e la possibilità d'interpretazione di **linguaggi di scripting** (come il **JavaScript**). Dall'altro, si è migliorata la qualità di elaborazione dei server attraverso una nuova generazione di linguaggi integrati con il web server

(come JSP, PHP, ASP, etc.), trasformando pertanto i web server negli attuali **application server**.

La diffusione di queste soluzioni ha consentito di avviare l'utilizzo del web come **piattaforma applicativa** che oggi trova la sua massima espressione nei **Web Service**, alla cui realizzazione e diffusione sta lavorando l'intera industria mondiale del software per la gestione d'azienda, dai grandi nomi commerciali (come **SAP** e **Oracle**) fino alle comunità **Open Source**. L'utilizzo dei web-service all'interno dell'architettura di integrazione **SOA** permetterà anche alle piccole imprese di gestire senza grandi sforzi i propri **processi aziendali**.

Scopo dei Web Service è di limitare il più possibile le attività di **implementazione**, consentendo di accedere a servizi software resi disponibili in rete, assemblarli secondo le proprie necessità e pagarli soltanto per il loro utilizzo effettivo, metodologia individuata nella terminologia anglosassone come *pay per use, on demand software, just in time software, on tap software*, etc.

È chiaro, quindi, che i web-service e il loro successo hanno ed avranno un legame strutturale ed intrinseco con i processi aziendali che dovranno supportare nell'ambito di una nuova organizzazione basata sui processi.

Nonostante tutte queste evoluzioni, il web rimane, ancora e soprattutto, una gigantesca **biblioteca** di pagine HTML **statiche** *[senza fonte]* on-line. Però, se da un lato lo standard HTML con la sua semplicità ha contribuito all'affermazione del web, dall'altro ha la grossa limitazione di occuparsi solo ed esclusivamente della **formattazione** dei **documenti**, tralasciando del tutto la struttura e il **significato** del contenuto.

Questo pone notevoli difficoltà nel **reperimento** e **riutilizzo** delle **informazioni**. Per rendersi conto di questo è sufficiente eseguire una ricerca utilizzando uno dei molti motori disponibili in rete e ci si accorgerà che, delle migliaia di documenti risultanti dalla **query**, spesso solo una piccola percentuale è d'interesse per la ricerca che s'intendeva fare. Ad esempio, per un qualsiasi motore di ricerca, non esiste alcuna differenza fra il termine *Rossi* nel contesto *Il Sig. Rossi* ed il termine *rossi* nel contesto *capelli rossi*, rendendo la ricerca molto difficile.

La risposta a questo problema è venuta, ancora una volta, dal fisico inglese **Tim Berners-Lee**, che, abbandonato il **CERN**, ha fondato il consorzio **W3C** che ha assunto il ruolo di governo nello sviluppo di standard e protocolli legati al web. Egli nel **1998** ha definito lo standard **XML** (eXtensible Markup Language), un metalinguaggio derivante dall'**SGML**, che consente la creazione di nuovi linguaggi di marcatura (ad es. lo stesso HTML è stato ridefinito in XML come XHTML). Sua caratteristica innovativa è la possibilità di aggiungere informazioni semantiche sui contenuti attraverso la definizione di opportuni **tag**.

I principali obiettivi di XML, dichiarati nella prima specifica ufficiale^[7] (ottobre **1998**), sono pochi ed espliciti: utilizzo del **linguaggio** su Internet, facilità di creazione dei documenti, supporto di più **applicazioni**, chiarezza e comprensibilità. Con queste semplici caratteristiche l'XML fornisce un modo comune di **rappresentare i dati**, cosicché i programmi software siano in grado di eseguire meglio ricerche, visualizzare e manipolare informazioni nascoste nell'oscurità contestuale.

È per questo che, nonostante la sua giovane età, l'XML è alla base di tutte le nuove specifiche tecnologiche distribuite dal **W3C** ed è stato adottato come standard di rappresentazione dati da tutta l'industria informatica *[non chiaro]* (dai *file di configurazione* delle applicazioni alla definizione di *formati di interscambio dei dati*). *[senza fonte]*

Le specifiche XML hanno però una lacuna molto importante: non definiscono alcun meccanismo univoco e condiviso per specificare **relazioni tra informazioni** espresse sul web per una loro **elaborazione automatica** (ad es. più documenti che parlano dello stesso argomento, persona, organizzazione, oggetto), rendendo molto

difficile la **condivisione delle informazioni**.

Anche in questo caso la soluzione al problema è venuta dal W3C di Berners-Lee, attraverso la formalizzazione del **web semantico**. Il W3C considera l'ideale evoluzione del web dal *machine-representable* al *machine-understandable*. L'idea è di generare documenti che possano non solo essere letti e apprezzati da esseri umani, ma anche accessibili e interpretabili da *agenti* automatici per la ricerca di contenuti.

A tale scopo sono stati definiti alcuni linguaggi, quali **Resource Description Framework (RDF)** e **Web Ontology Language (OWL)**, entrambi basati su XML, che consentono di esprimere le relazioni tra le informazioni rifacendosi alla **logica dei predicati** mutuata dall'**intelligenza artificiale**.^[8] Questi standard sono già disponibili, ma continuano ad essere ulteriormente sviluppati insieme a formalismi e strumenti per dotare il web di **capacità di inferenza**.

Quello appena esposto è un processo solo apparentemente tecnico, ma ben visibile nella sua portata, che ha come obiettivo l'approdo all'**intelligenza condivisa** del web che promette, a breve, l'uso più efficiente dei siti internet e, a più lungo termine, una autentica trasformazione nella natura del software e dei servizi.

Tanto interesse per queste tecnologie è da ravvisare nel fatto che tutti (utenti, produttori di software e di servizi piccoli e grandi) hanno da avvantaggiarsi dalla diffusione piena di questi standard **[senza fonte]**. La formazione nel corpo del web di una vasta **rete "semantica"** è, infatti, la condizione chiave per il decollo di un nuovo modo di intendere ed usare il web.