



Corso di ***Progettazione*** ***di sistemi multimediali***

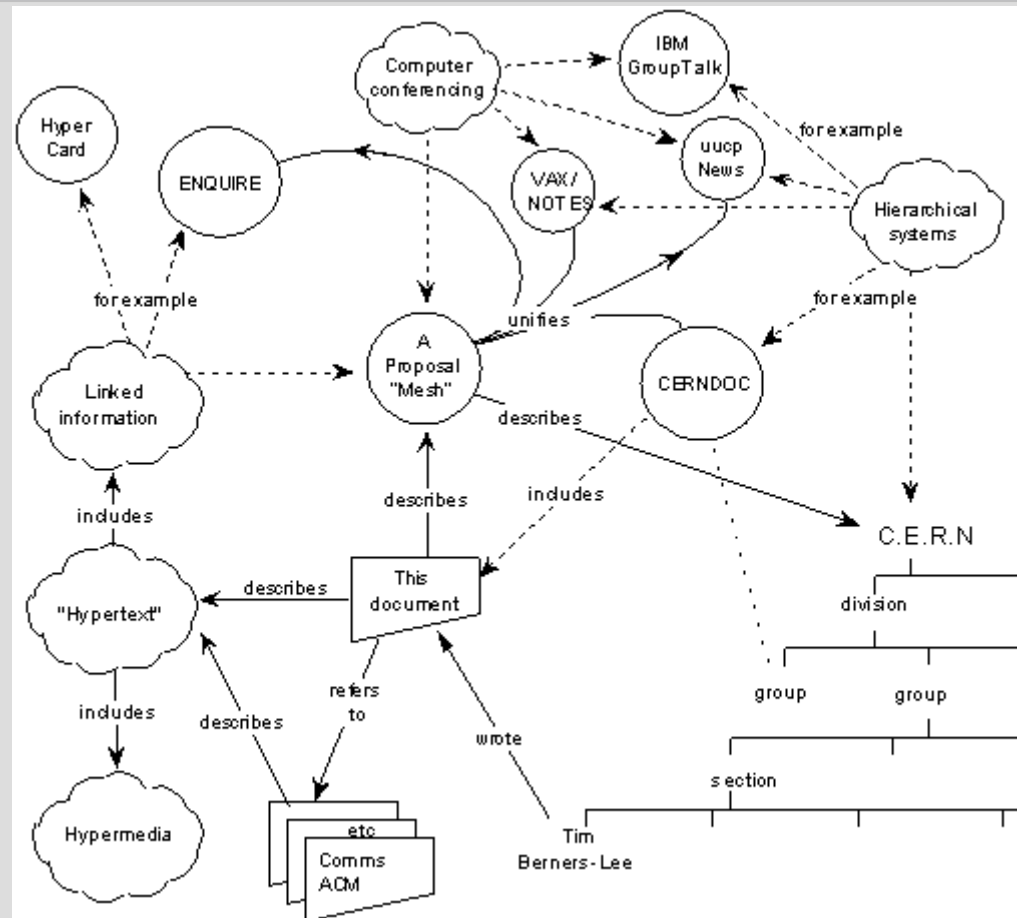
prof. Pierluigi Feliciati

a.a.2013/14

Modulo 3

i linguaggi, gli strati informativi e la gestione del web :
ipertesto, HTML, XML, XHTML, i metadati descrittivi,
strutturali e gestionali, web 2.0, web semantico, le
applicazioni

L'immagine classica del Web



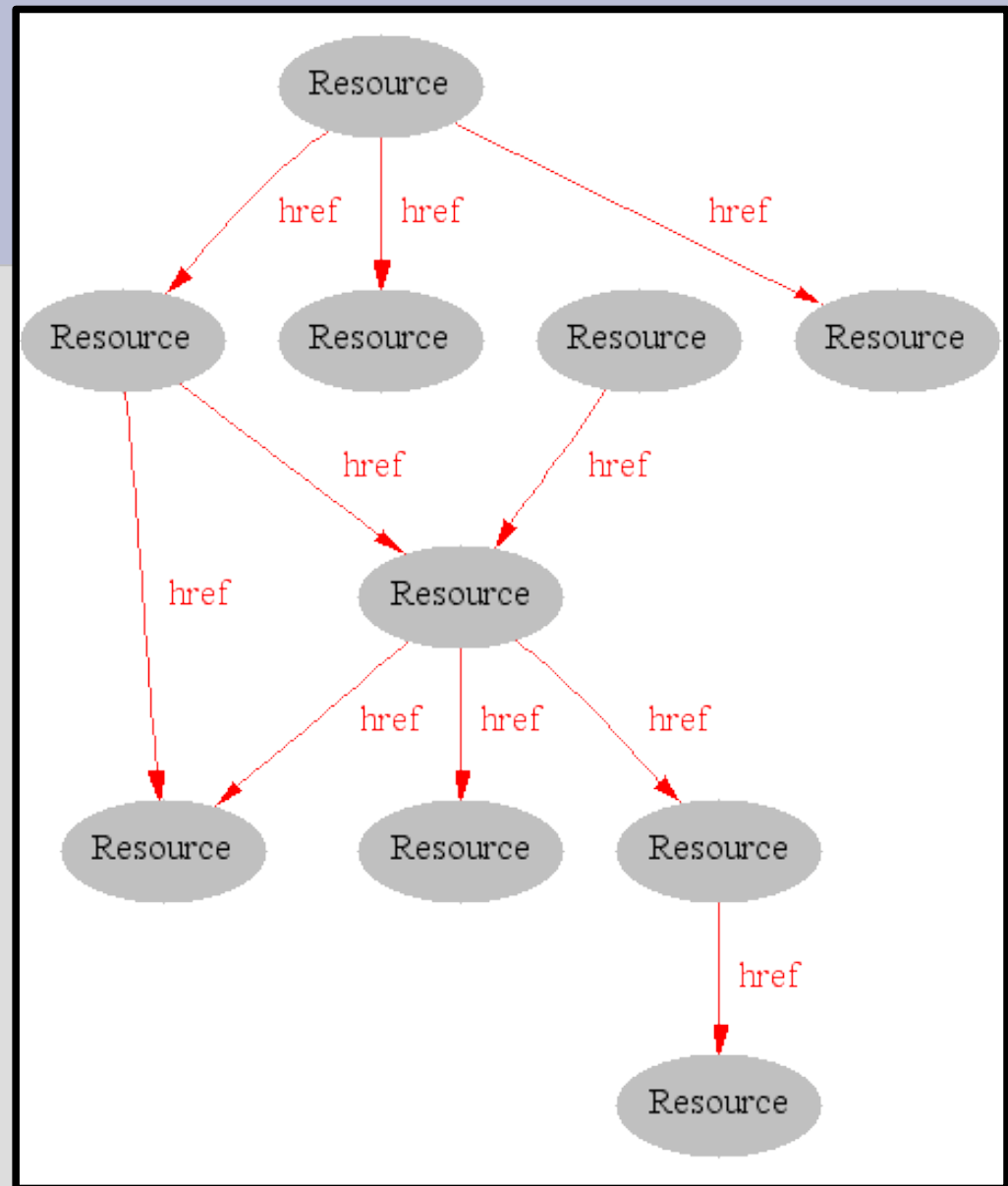
Information Management: A Proposal

by **Tim Berners-Lee (CERN 1990)** <<http://www.w3.org/History/1989/proposal.html>>

L'immagine classica del Web

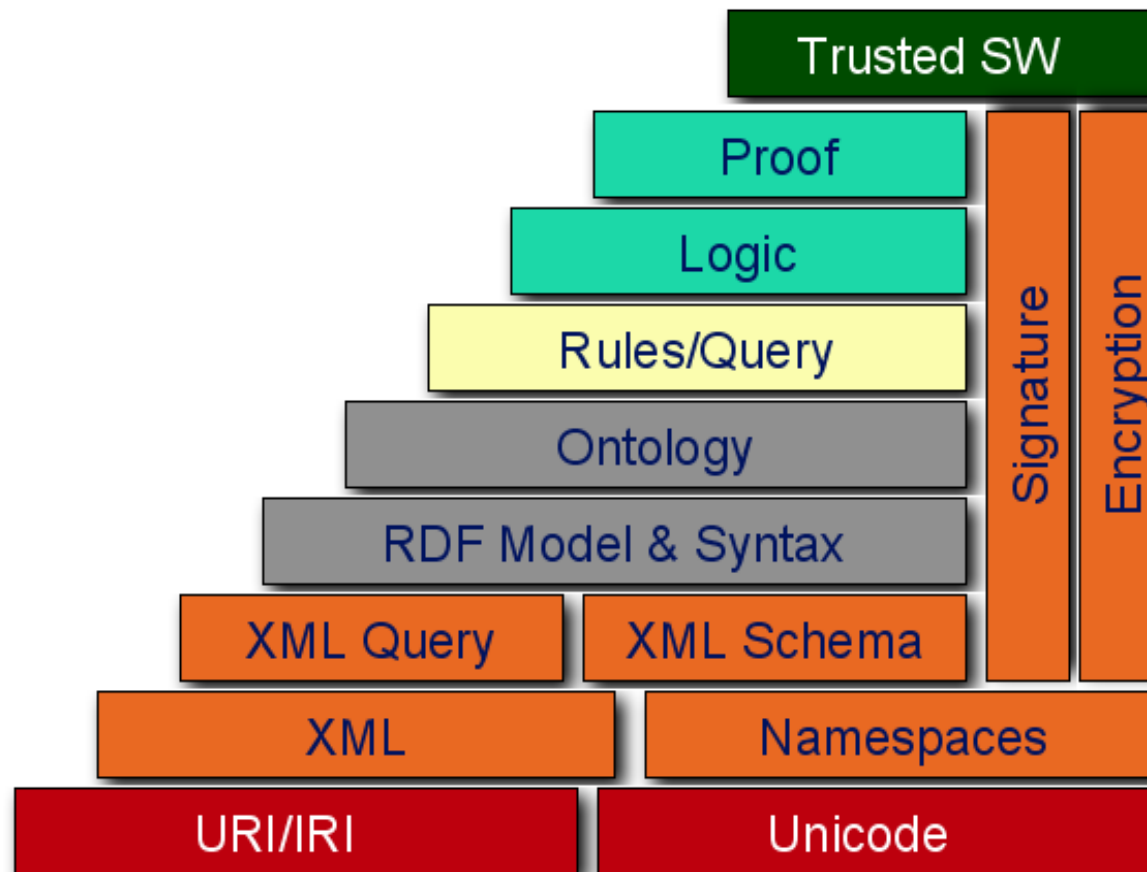
- Tim Berners Lee nel 1980 al CERN di Ginevra realizzò, per uso personale, il primo software per immagazzinare informazioni usando associazioni casuali: *Enquire*. Tale prototipo, mai pubblicato, formerà la base concettuale per il futuro sviluppo del WWW.
- Nel 1989 propose un progetto globale sull'ipertesto, *Information Management: A Proposal*, base del *World Wide Web*
- Berners-Lee e Robert Cailliau nel 1990 misero poi a punto il protocollo HTTP e una prima specifica di HTML, sulla base dei quali sono stati realizzati un primo programma *browser/editor* ipertestuale per il WWW, utilizzato dal CERN nel 1991.

Web 1.0



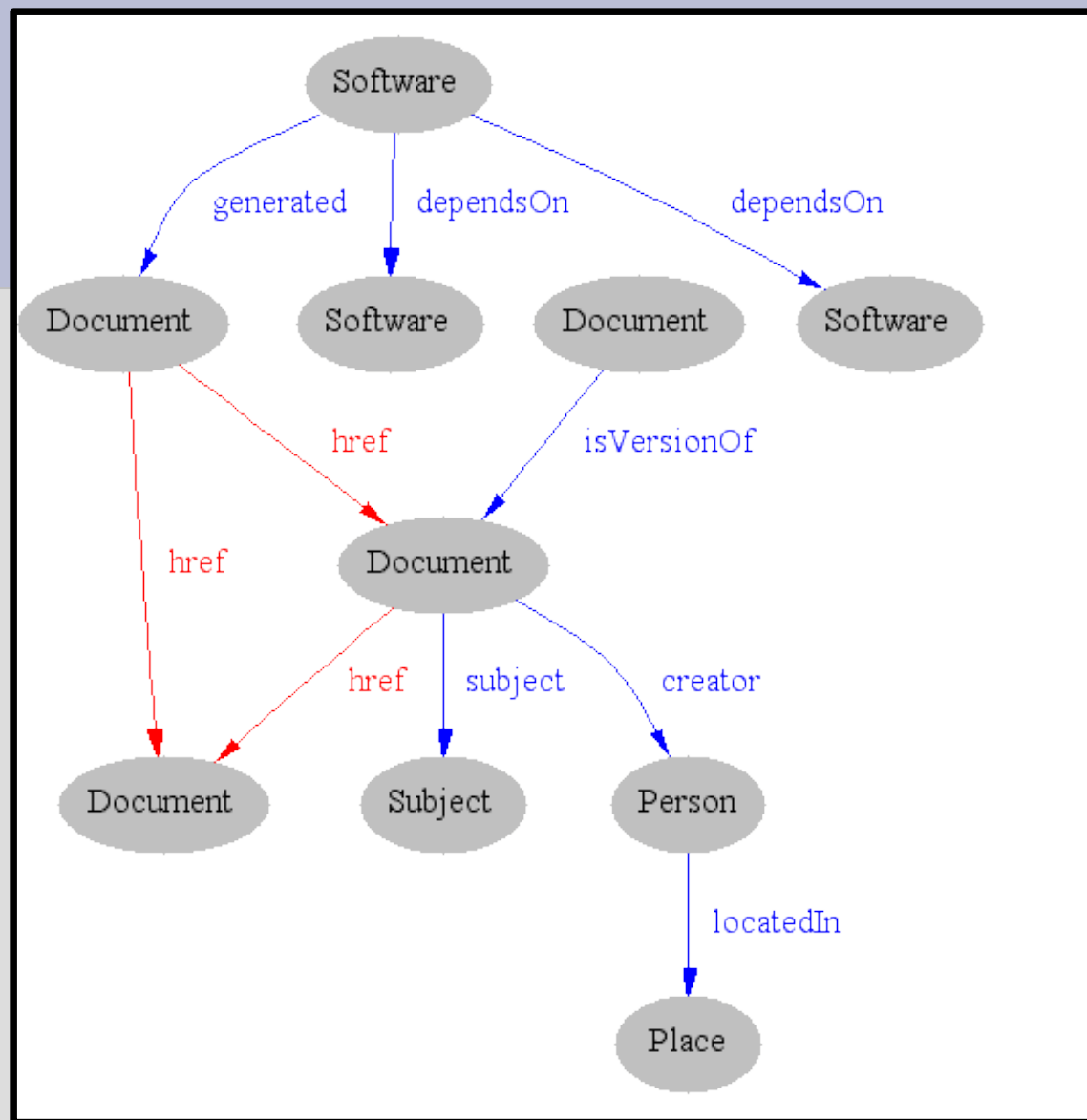
L'ipertesto applicato nel WWW:
risorse collegate tra di loro tramite link, con logica ipertestuale

La prossima piramide del Web



Semantic Web - XML2000 by *Tim Berners-Lee*
<<http://www.w3.org/2000/Talks/1206-xml2k-tbl/>>

Web semantico



Il web del futuro, fatto di milioni di risorse collegate con logica semantica, ovvero con informazioni sul **tipo di risorse e di relazioni**

La filosofia della grande rete

Le origini dell'ipertesto

Agostino Ramelli, ingegnere italiano nato nel 1531, ideò la **ruota dei libri**, un leggio multiplo rotante, per consentire l'agevole lettura contemporanea di più testi e che si può considerare una *prima forma di ipertesto*.

C'è chi fa risalire il concetto di ipertesto a Vannevar Bush che nel 1945 scrisse un articolo intitolato *As We May Think* nel quale descrive un sistema di informazione interconnesso chiamato **Memex**, mai realizzato e considerato profetico per la rete web.

<http://www.theatlantic.com/doc/194507/bush>

La filosofia della grande rete

Iper testo - storia

Nel 1929 il regista sovietico Sergej Michajlovič **Ejzenštejn** si lamentò di non potere raccogliere in un volume i suoi saggi in modo che fossero *percepiti tutti insieme simultaneamente per comparare ciascun saggio direttamente con gli altri, di passare dall'uno all'altro avanti e indietro*. “L'unico tipo di libro in grado di soddisfare queste condizioni dovrebbe avere la forma di una sfera”.

Mentre nel campo tecnico l'ipertesto si è diffuso, vedi il web, nel campo letterario è rimasto confinato ad ambiti prevalentemente sperimentali. La narrativa stampata si è finora basata sul sistema sequenziale, lineare, di pagine di un libro stampato.

La filosofia della grande rete

Iper testo - storia

I due **inventori** riconosciuti **dell'ipertesto** sono Ted Nelson e Douglas Engelbart: il primo inventò il termine *hypertext* nel 1963, indicando così qualsiasi sistema di scrittura non lineare che utilizzasse l'informatica. Aveva fondato il progetto *Xanadu* con l'intento di creare una rete di computer collegati e dotati di un'interfaccia semplice.

Ted Nelson si è infatti occupato soprattutto di uso dei software; una sua famosa massima recita:

« Un'interfaccia utente dovrebbe essere così semplice da poter essere compresa, in caso di emergenza, nel giro di dieci secondi da un principiante

La filosofia della grande rete

Iper testo - storia

Nel 1980 il programmatore Bill Atkinson realizzò per la Apple *HyperCard*, un'applicazione software che gestiva in maniera semplice grandi quantità di informazioni sotto forma di testo o di immagini, dotato di un avanzato linguaggio di programmazione ipertestuale, *HyperTalk*.

Malgrado fosse solo per la piattaforma MacOS, divenne uno dei più diffusi sistemi di realizzazione di ipertesti prima del World Wide Web. Fu ritirato dalle vendite nel marzo del 2004.

Ward Cunningham lo usò per abbozzare ciò che avrebbe sviluppato successivamente sotto il nome di *wiki*.

La filosofia della grande rete

Abbiamo già accennato a **Tim Berners-Lee** e alla sua invenzione del **World Wide Web** per rispondere alla necessità della comunità scientifica di un sistema di condivisione delle informazioni tra diverse università ed istituti di tutto il mondo.

All'inizio del 1993 all'Università dell'Illinois fu sviluppata e resa pubblica la prima versione del *browser* **Mosaic**, uno strumento leggero di navigazione multimediale per Internet.

Il traffico Web esplose, passando da soli 500 web server noti nel 1993 ad oltre 10.000 nel 1994 dopo la pubblicazione di Mosaic per Windows e Mac.

Oggi sono circa 95 milioni i domini web.

La filosofia della grande rete

Iper testo e ipermedia

L'ipertesto come lo intendiamo oggi è un insieme di testi pagine leggibili con l'ausilio di un'interfaccia elettronica, in modo non sequenziale, collegati tramite **collegamenti ipertestuali (link o hyperlink)**, che consentono una rete raggiata o variamente incrociata di informazioni, organizzate secondo diversi criteri, ad esempio paritetici o gerarchici, in modo da permettere più percorsi di lettura.

Con l'inserimento e collegamento di contenuti **di media diversi (multimedia)** in forma di ipertesto, si è cominciato a usare il termine **ipermedia**, anche se con **“testo”** ci si riferisce comunemente ad una narrazione multimediale

La filosofia della grande rete

Vantaggi dell'ipertesto

Visione globale: La suddivisione dell'informazione in unita' minimali, a senso compiuto e indipendenti l'una dall'altra, presentata con il display delle relazioni fra le stesse, agevola la visione di insieme.

Facilità di consultazione: La possibilita' di seguire i riferimenti in tempi brevi, non appesantisce la lettura del testo, rendendo piu' agevole l'approfondimento dell'argomento trattato.

Aggiornabilità: L'estrema facilita' di inserimento di nuovi dati informativi rende gli ipertesti lo strumento comunicativo piu' adatto ai settori ad alto tasso di innovazione, come quello tecnologico.

La filosofia della grande rete

Vantaggi dell'ipertesto

Lettura personalizzata e destinazione multipla:

La strutturazione degli ipertesti rende possibile una lettura non lineare dei testi, offrendo all'utente modi diversificati di consultazione dell'informazione, al contrario di quanto avviene nei testi cartacei.

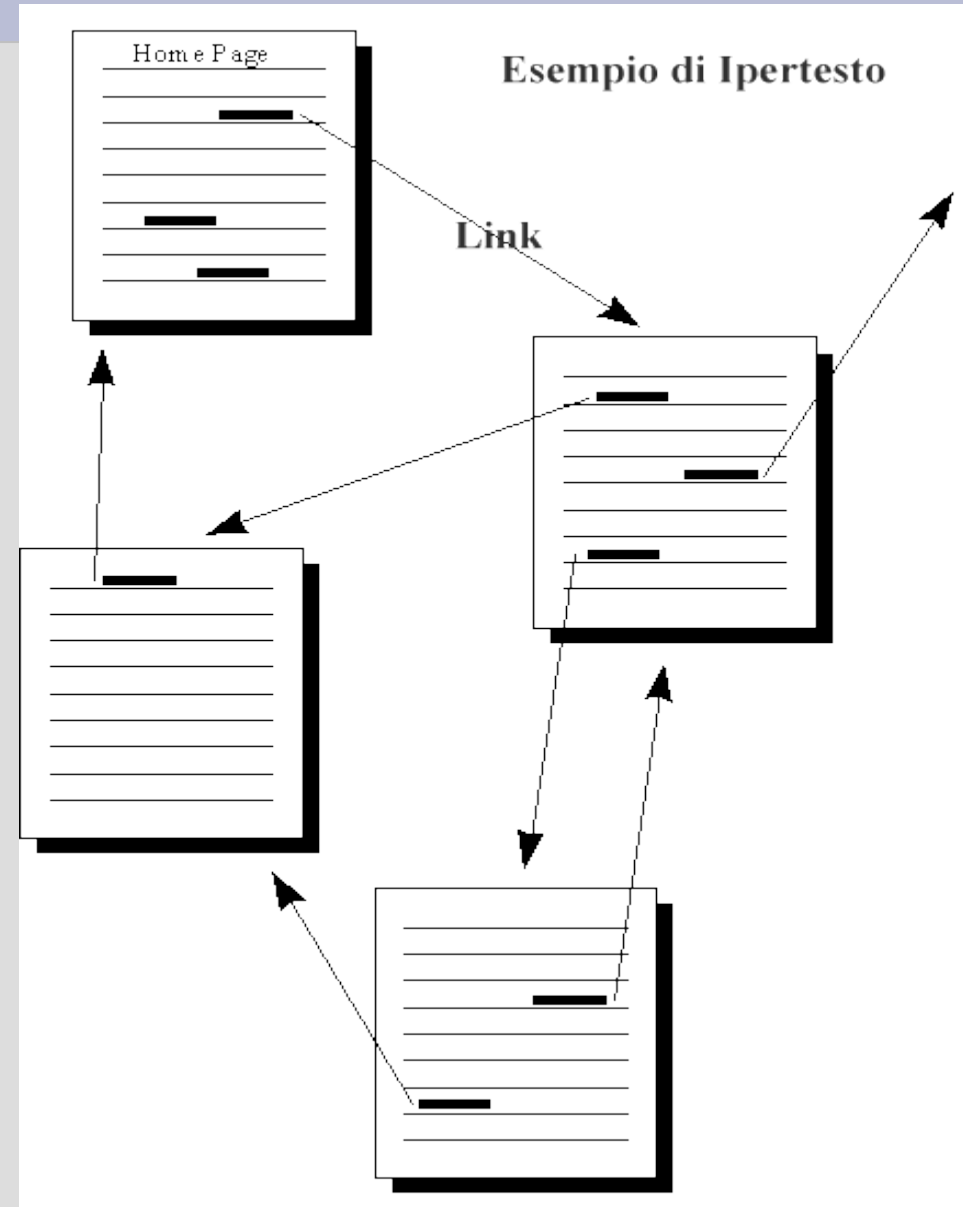
Le diverse modalità di lettura aumentano il ventaglio di utenza possibile: qualunque fruitore può visitare l'iperdocumento secondo una specifica chiave di consultazione dettata da particolari esigenze informative.

La filosofia della grande rete

Il documento contiene dei puntatori detti **link** evidenziati in vario modo, che consentono di passare dal documento ai documenti correlati per mezzo di **un'azione attiva**.

Un ipertesto e' pertanto costituito da un insieme di documenti fra loro correlati tramite questi puntatori.

Le informazioni "nascoste" dal link ipertestuale possono essere contenute in un'altra parte dello stesso documento, in un altro documento sullo stesso host oppure in un documento che si trova in un altro host.



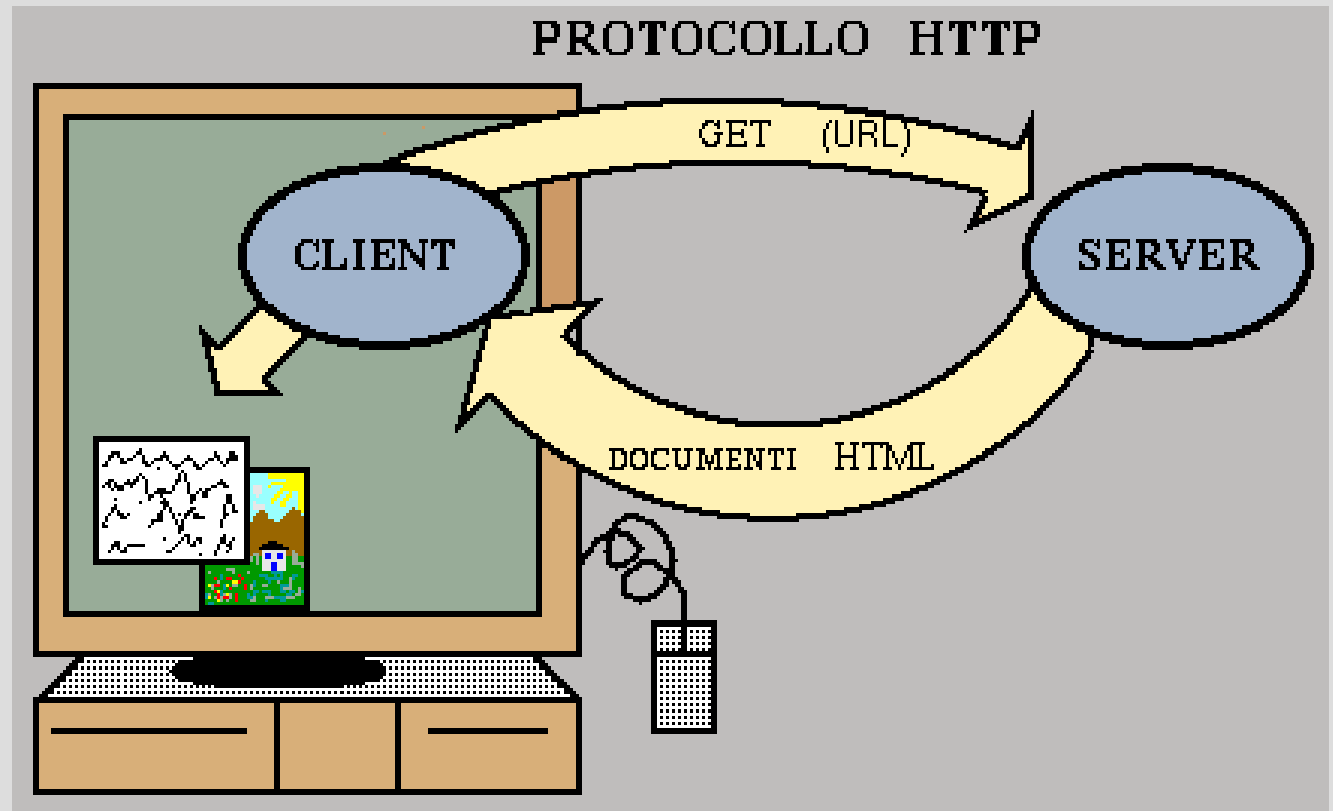
server e client nel Web

Un **server web** è un servizio o tipo di server, e per estensione il computer su cui è in esecuzione, che si occupa di fornire, tramite software dedicato e su richiesta dell'utente (**client**), file di qualsiasi tipo, tra cui pagine web. Le informazioni inviate dal server web all'utente viaggiano sulla rete Internet.



HTTP, browser, client/server

L'HTTP funziona su un meccanismo richiesta/risposta (client/server): il client tramite il *browser* esegue una richiesta (GET) e il server restituisce la risposta (*files*). Nell'uso comune il client corrisponde al browser ed il server al sito web.



Internet e Web

Quindi, il World Wide Web, con il protocollo HTTP, sono solo **uno dei servizi** che viaggiano sulla rete Internet.

Questa rete di computer, che ha origine alla fine degli anni '60, è basata sui suoi protocolli Transmission Control Protocol (TCP) e Internet Protocol (IP). L'Italia fu il terzo Paese in Europa a connettersi in rete, dopo Norvegia e Inghilterra, nel 1986.

Internet è una rete logica di enorme complessità appoggiata a strutture fisiche e collegamenti di vario tipo, fibre ottiche, cavi coassiali, collegamenti satellitari, doppino telefonico, collegamenti su radiofrequenza (WiFi), su ponti radio... Ogni dispositivo connesso a Internet si chiama *host*.

Internet

Generalmente Internet è definita **la rete delle reti**, perché è costituita da tutte le reti, private, pubbliche, aziendali, universitarie, commerciali, connesse tra di loro. Queste reti locali esistevano anche prima di Internet, ad es. nei centri di ricerca e nei dipartimenti universitari.

Il grande risultato di Internet è stato quello di creare uno standard *de facto* che interoperasse tra le reti eterogenee, consentendo ai più diversi enti e agenti di **scambiarsi dati** mediante un protocollo comune (TCP/IP), relativamente **indipendente** da hardware, da sistemi operativi, dai linguaggi di comunicazione degli apparati di comunicazione, da qualche anno anche smartphones.

cosa sono i siti web?

Un **sito web** o (erroneamente) **sito internet** è un insieme di pagine web, ovvero una struttura ipertestuale di documenti accessibili con un browser tramite World Wide Web su rete Internet.

I **siti web statici** rendono disponibili contenuti di sola lettura (fruizione *unidirezionale*).

I **siti web dinamici** presentano invece contenuti redatti e gestiti **dinamicamente** (per esempio con un database) e forniscono contenuti e servizi anche molto complessi. I siti web dinamici sono caratterizzati da una potenzialmente alta **interazione fra sito e utente**.

cosa sono i siti web?

L'accesso ai siti web

Le pagine di un sito web sono accessibili tramite una **radice comune** (detta **nome di dominio**, per esempio www.unimc.it), seguita da una serie di **sottocartelle** e dal **nome della pagina/file**. Il nome completo di ogni pagina è l'indirizzo web o, più tecnicamente, lo **URI (o URL)**. L'*home page* di un sito è la prima pagina che si ottiene digitando il solo nome di dominio (di solito il file è home o index o altri...).

Per esempio:

<http://www.w3c.org/Consortium/Offices/role.html>

www.w3c.org/ è la radice, o nome di dominio
[/Consortium/Offices/](http://www.w3c.org/Consortium/Offices/) sono le sottocartelle, separate dal simbolo "/"
[role.html](http://www.w3c.org/Consortium/Offices/role.html) è il nome della pagina

cosa sono i siti web?

Solitamente le pagine di un sito risiedono tutte **sullo stesso server web**, e la ramificazione in sottocartelle dell'indirizzo corrisponde ad una uguale ramificazione nell'*hard disk* di una macchina o server.

Ma, grazie alla tecnologia ipermediale, nella stessa pagina possono essere **inclusi/visualizzati anche contenuti residenti altrove**, richiamati dai link attraverso gli opportuni tag di HTML.

Questo è possibile grazie al **principio di univocità degli indirizzi Web**, per cui **qualsiasi oggetto nella rete può essere richiamato tramite il suo indirizzo univoco.**

Gli indirizzi del Web: URL

Un **Uniform Resource Locator** o **URL** è una sequenza di caratteri che identifica univocamente **l'indirizzo mnemonico di una risorsa in Internet**, tipicamente presente su un server, come ad esempio un documento, un'immagine, un video.

Il **dominio DNS** è un **indirizzo IP** (sequenza di 4 numeri da 0 a 255, divisi da punti) convertito in un **nome DNS**, una serie di stringhe separate da punti. La parte più importante è la prima **partendo da destra**. Questa è detta dominio di **primo livello** (o TLD, Top Level Domain), per esempio .org o .it. Un dominio di **secondo livello** consiste in due parti, per esempio wikipedia.org. Quando un dominio di secondo livello viene registrato, si autorizza a usare i nomi di dominio relativi ai successivi livelli.

L'ARCHITETTURA dei siti WEB

Per mettere a fuoco gli elementi in gioco nelle dinamiche comunicative (e nella qualità) delle pagine web, connesse tra di loro all'interno di *ambienti* informativi (i siti web), utilizzerò uno schema proposto da M. Filacchioni, *Un approccio moderno alle pagine web* (2006).

L'idea da cui parte è che esistono diversi aspetti, distinti, in una pagina Web:

- **Contenuto**
- **Struttura**
- **Presentazione**
- **Compotamento**

L'ARCHITETTURA dei siti WEB

il contenuto

che rappresenta **l'informazione che si vuole comunicare** all'utente (testuale, visuale, sonora, video..), che come è facile intuire è l'aspetto che maggiormente contribuisce a determinare il valore che l'utente attribuisce alla pagina

la struttura

intesa sia come **struttura logica del contenuto**, cioè l'organizzazione di questo in paragrafi, intestazioni, elenchi puntati e numerati, ecc. sia come **struttura di navigazione**, costituita dall'insieme di link che consentono all'utente di raggiungere le varie sezioni di un sito, eventuali form di ricerca e login, ecc. sia

L'ARCHITETTURA dei siti WEB

la presentazione

che rappresenta la modalità con cui i **contenuti strutturati vengono presentati all'utente**. La presentazione può essere di tipo visuale (grafica), testuale e non visuale, per gli utenti che usano tecnologie assistive ed alcuni browser alternativi (se ne parlerà avanti)

il comportamento

che consente di **alterare** la struttura e la presentazione in risposta ad **eventi generati dall'utente** (pressione di un tasto del mouse, movimento del puntatore, pressione di un tasto della tastiera, ecc.) o ad eventi e **condizioni indipendenti dall'utente**

L'ARCHITETTURA dei siti WEB

I primi tre aspetti (**contenuto**, **struttura**, **presentazione**) definiscono una pagina come **entità statica**, l'ultimo (**comportamento**) ne determina la **dinamicità**.

Premesso che **contenuto ed un minimo di struttura logica sono sempre presenti**, non tutte le pagine hanno tutti e quattro gli aspetti.

Molte pagine, infatti, non presentano nessuna struttura di navigazione, altre non presentano nessun comportamento ed altre ancora, per la verità piuttosto rare, non presentano alcuna presentazione.

L'ARCHITETTURA dei siti WEB

Questi vari aspetti coinvolgono figure professionali diverse:

il **redattore** che è responsabile del contenuto e della struttura logica

l'**esperto di architettura dell'informazione** che è responsabile dell'organizzazione dell'intero sito e quindi della struttura di navigazione, ma anche della struttura logica delle varie tipologie di pagine in esso presenti

il **designer** che è responsabile della presentazione

lo **sviluppatore** che è responsabile del comportamento

Spesso accade, comunque, soprattutto nei progetti Web di piccole e medie dimensioni, che **più ruoli collassino su una stessa persona.**

I'ARCHITETTURA del WEB

Il vecchio modo di concepire le pagine Web

A partire dall'affermazione del Web come fenomeno di massa, avvenuta a pochi anni dalla sua creazione grazie alla diffusione dei **browser grafici** (*Mosaic* dell'NCSA, *Netscape Navigator* della Netscape Communications – padre di Firefox/Mozilla - ed *Internet Explorer* di Microsoft) e successivamente degli **editor HTML visuali** (*Netscape Composer*, *FrontPage* di Microsoft, *DreamWeaver* di Macromedia, ecc.), sono state prodotte **centinaia di milioni di pagine Web**.

L'ARCHITETTURA dei siti WEB

La stragrande maggioranza di queste pagine erano (ed in parte sono tuttora) caratterizzate da un **mescolamento degli aspetti** che le caratterizzano, da una **struttura logica minimale** (in molti casi quasi inesistente) e da una **presentazione predominante** (utilizzata anche per colmare la carenza di struttura logica).

Si noti comunque che **il mescolamento del contenuto e della struttura è inevitabile**, essendo questi aspetti per loro natura fortemente dipendenti.

L'ARCHITETTURA dei siti WEB

Il vecchio modo presenta i seguenti **svantaggi**:

- non rende indipendenti le figure professionali associate ai vari aspetti delle pagine
- rende difficile la modifica dei singoli aspetti delle pagine successivamente alla loro creazione, complicando molto:
 - l'aggiornamento dei contenuti
 - il restyling complessivo dei siti
 - il miglioramento dell'usabilità
- costringe a duplicare i contenuti qualora si volessero realizzare presentazioni alternative degli stessi
- pregiudica l'accessibilità
- rende difficile l'estrazione delle informazioni contenute nelle pagine da parte di procedure automatiche

L'ARCHITETTURA dei siti WEB

Questi svantaggi dipendono da:

una esagerata **esasperazione della presentazione grafica** da parte dei committenti dei siti e dei designer (che soprattutto in passato non avevano una formazione specifica sul Web)

le **estensioni degli standard** da parte dei produttori dei browser

la lentezza con cui i produttori hanno implementato gli standard

il fatto che gli editor HTML visuali spesso **non incoraggiano l'uso degli standard**

L'ARCHITETTURA dei siti WEB

L'ARCHITETTURA A STRATI

Grazie ad un maggior **supporto degli standard da parte dei browser** e ad una generale presa di coscienza dei problemi accennati, negli ultimi anni si sta lentamente affermando un nuovo modo di concepire il Web.

Questo nuovo approccio propone **un'architettura** in cui **gli aspetti delle pagine sono nettamente separati** (dal punto di vista **logico e fisico**), in modo tale che ad ognuno di essi corrisponda uno strato, un livello indipendente

fatta eccezione per il contenuto e la struttura ai quali corrisponde un unico strato

L'ARCHITETTURA dei siti WEB

Lo strato di comportamento poggia sia su quello di presentazione, che su quello di contenuto e struttura in quanto agisce su entrambi

Lo strato di presentazione poggia unicamente su quello di contenuto e struttura.

Ogni strato agisce esclusivamente su quello sottostante (quelli sottostanti nel caso del comportamento) e deve totalmente ignorare la presenza di quelli sovrastanti.

I'ARCHITETTURA del WEB

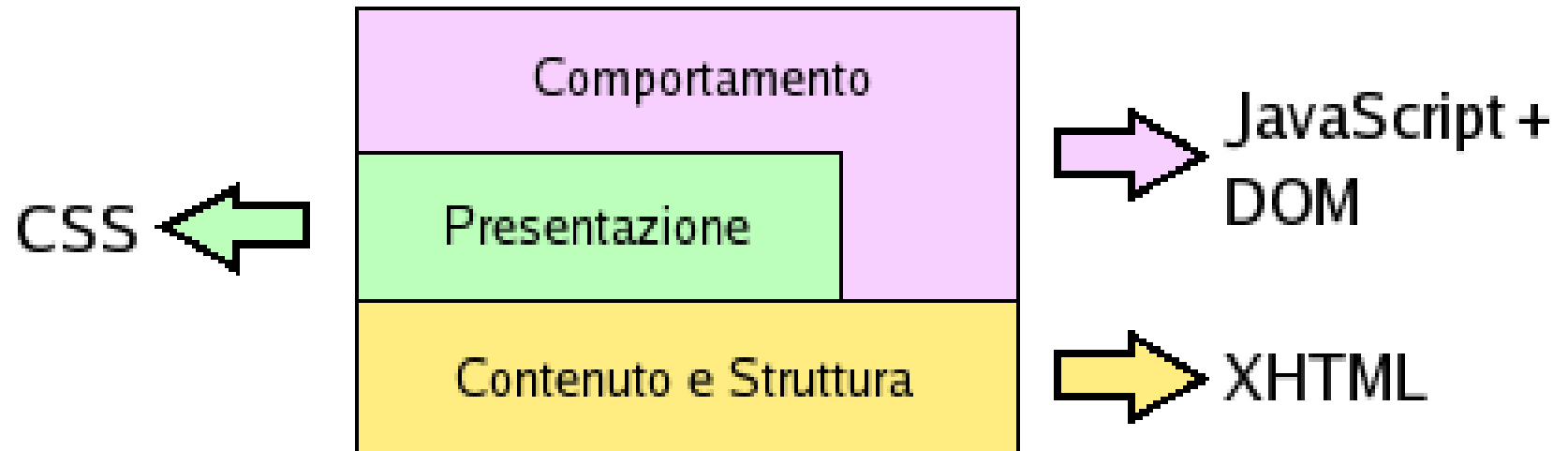
Lo strato di **contenuto+struttura** deve essere il più **semplice e lineare** possibile ed esaltare il significato delle informazioni.

Gli accorgimenti necessari per livellare le personalizzazioni dei browser degli utenti , invece, devono riguardare unicamente gli strati di **presentazione e comportamento**.

Le principali tecnologie utilizzate dai vari strati sono:

- l'**XHTML** per lo strato di contenuto e struttura
- i fogli di stile **CSS** per lo strato di presentazione
- **JavaScript, DOM, Flash, ...** per lo strato di comportamento

l'architettura del web a strati



cosa sono i siti web?

i siti web e HTML

Il linguaggio più diffuso con cui i siti web sono costruiti è l'**HTML** (*Hyper Text Markup Language*, linguaggio di marcatura ipertestuale) e/o i suoi derivati, che viene interpretato/decodificato da particolari software chiamati **web browser** (tra i più celebri Internet Explorer, Chrome e Firefox).

Il termine *markup* (**marcatura**) deriva dall'ambiente tipografico dove si usava marcare con annotazioni le parti del testo che andavano evidenziate o corrette, allo scopo di segnalarle al compositore o al dattilografo.

cosa sono i siti web?

HTML

Nei documenti HTML, un tratto di testo può essere contrassegnato inserendo delle **etichette**, **tag**, che ne descrivono la funzione, il colore, il link, o altre caratteristiche.

L'HTML non è un linguaggio di programmazione ma di **strutturazione e impaginazione, formattazione** (layout) del contenuto, testuale e non, di una pagina web.

Punto HTML (.html) o punto HTM (.htm) è l'estensione comune per riconoscere i documenti in questo formato.

cosa sono i siti web?

i siti web e HTML

Un'importante caratteristica di HTML è che è stato concepito per definire il **contenuto logico** e non **l'aspetto finale** del documento.

I dispositivi (hardware e software) che possono accedere ad un documento HTML in rete sono molteplici, non sempre dotati di potenti capacità grafiche: **non esiste alcuna garanzia che un documento html venga visualizzato in ugual modo su tutti i dispositivi.**

Se da una parte questo ha imposto in passato dei forti limiti agli sviluppatori di pagine Web, ha dall'altro garantito la massima diffusione di Internet ed evitato che diventasse un medium di élite.

cosa sono i siti web?

La struttura di HTML

Ogni file HTML deve necessariamente comprendere:

- la dichiarazione della grammatica in uso (con rinvio allo schema DTD)
- Un'intestazione (**header**) per i metadati
- Un corpo (**body**) per i contenuti da visualizzare nella pagina

Documento HTML

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 .....>
```

DTD

```
<html>
```

```
<head>
```

Header
(informazioni di servizio)

```
</head>
```

```
<body>
```

Body
(contenuti visualizzati dal browser)

```
</body>
```

```
</html>
```

Il linguaggio dei linguaggi XML

XML (*eXtensible Markup Language*) è un **meta-linguaggio** di markup, ovvero un linguaggio marcatore che definisce un meccanismo sintattico che consente di estendere o controllare il significato di altri linguaggi.

Il W3C, in seguito alla cd. *guerra dei browser* (ovvero la corsa di fine anni '90 alle estensioni proprietarie all'HTML ufficiale), fu costretto a scegliere quali caratteristiche standardizzare e quali lasciare fuori dalle specifiche ufficiali dell'HTML. Fu in questo contesto che si sviluppò un linguaggio che desse **maggior libertà nella definizione dei tag**, pur rimanendo **uno standard**.

Il linguaggio dei linguaggi XML

La **semplicità**, **adattabilità** e **piena compatibilità** di XML con gli altri linguaggi di marcatura, la quasi totale mancanza di caratteristiche opzionali, la **leggibilità umana** e la **facilità per le macchine** di elaborarlo ne hanno decretato il successo.

L'unico difetto è la poca economicità in termini di spazio di occupazione, dovuta alla ridondanza di elementi + dati in ciascun file.

Lo sviluppo del linguaggio DTD prima, XSD poi, l'adozione dei *NameSpace* (per la definizione delle strutture dati e il richiamo di diverse regole in uno stesso file) e degli *stylesheet* (per la visualizzazione) hanno sostenuto questo percorso.

Il linguaggio dei linguaggi XML

Ci sono diversi modi per **dichiarare uno schema di codifica**, vale a dire la grammatica che stabilisce **quali elementi e quali regole** su di essi sarà possibile **usare all'interno di un documento XML** e in che modo tali elementi si potranno relazionare fra di loro. Il modo storico per dichiarare una codifica è la **Document Type Definition** (o **DTD**), una sintassi ereditata direttamente dal linguaggio SGML, da cui l'XML deriva.

Recentemente a questo metodo si sono aggiunti gli **XML Schema** (o **XSD**), espressi in vari modi e dimostratosi più efficace delle DTD. Lo schema più usato è quello del W3C <http://www.w3.org/TR/xmlschema-0/>.

Il linguaggio dei linguaggi XML

Le regole sintattiche di Xml sono pochissime, ma rigide:

Ogni documento deve presentare un **prologo** per la dichiarazione della specifica XML che si usa, la codifica testuale adottata e il richiamo esplicito alla DTD o *XML Schema* o *namespace* che si usa (nell'esempio, lo schema Dublin Core):

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
```

```
<xsd:schema  
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">  
  <xmlns:dc="http://purl.org/dc/elements/1.1/">
```

Il linguaggio dei linguaggi XML

Il contenuto di un documento XML deve corrispondere a una struttura di **tag** (o **elementi**) non sovrapposti, quindi i tag che si aprono per primi debbono chiudersi per primi.

Per elemento/tag (o marcatore) si intende una coppia `<parola>(apertura) </parola> (chiusura)` dove parola=nome del tag

<DOCUMENTO>

`<TITOLO>Titolo del documento</TITOLO>`

`<TESTO>`

`<PARAGRAFO>testo del primo paragrafo</PARAGRAFO>`

`<PARAGRAFO>testo del secondo paragrafo</PARAGRAFO>`

`</TESTO>`

</DOCUMENTO>

Da HTML a XHTML

L'**XHTML** non è altro che la **reformulazione dell'HTML in XML**, cioè in un **meta-linguaggio di descrizione**.

Le pagine XHTML, dal 2000, hanno una **struttura più individuabile** di quella delle pagine HTML e sono **estendibili**, in quanto l'XHTML può essere combinato con altri linguaggi derivati dall'XML, come MathML (Mathematical Markup Language) per le formule matematiche e SVG (Scalable Vector Graphics), per la descrizione di immagini vettoriali statiche e animate.

L'XHTML non tollera estensioni come l'HTML, ad esempio impedendo l'uso di tag di formattazione.

Da HTML a XHTML

Attualmente, l'XHTML è fermo alla versione 1.2, mentre da un paio di anni, dopo 10 anni di HTML 4.1, si è rilasciato HTML 5.

Le novità introdotte dall'HTML5 sono migliorare il **disaccoppiamento tra struttura**, caratteristiche di **resa** (tipo di carattere, colori, eccetera), definite dalle direttive di stile, e **contenuti** di una pagina web, definiti dal testo vero e proprio.

Inoltre l'HTML5 in sintonia con il web 2.0, supporta la **memorizzazione locale di grosse quantità di dati**, per consentire l'utilizzo di applicazioni sincroniche basate su web (Gmail o altri servizi analoghi) anche in assenza di collegamento a Internet.

Da HTML a XHTML

Un documento XHTML si basa su XML, quindi l'apertura deve citare sia la sintassi XML che quella XHTML:

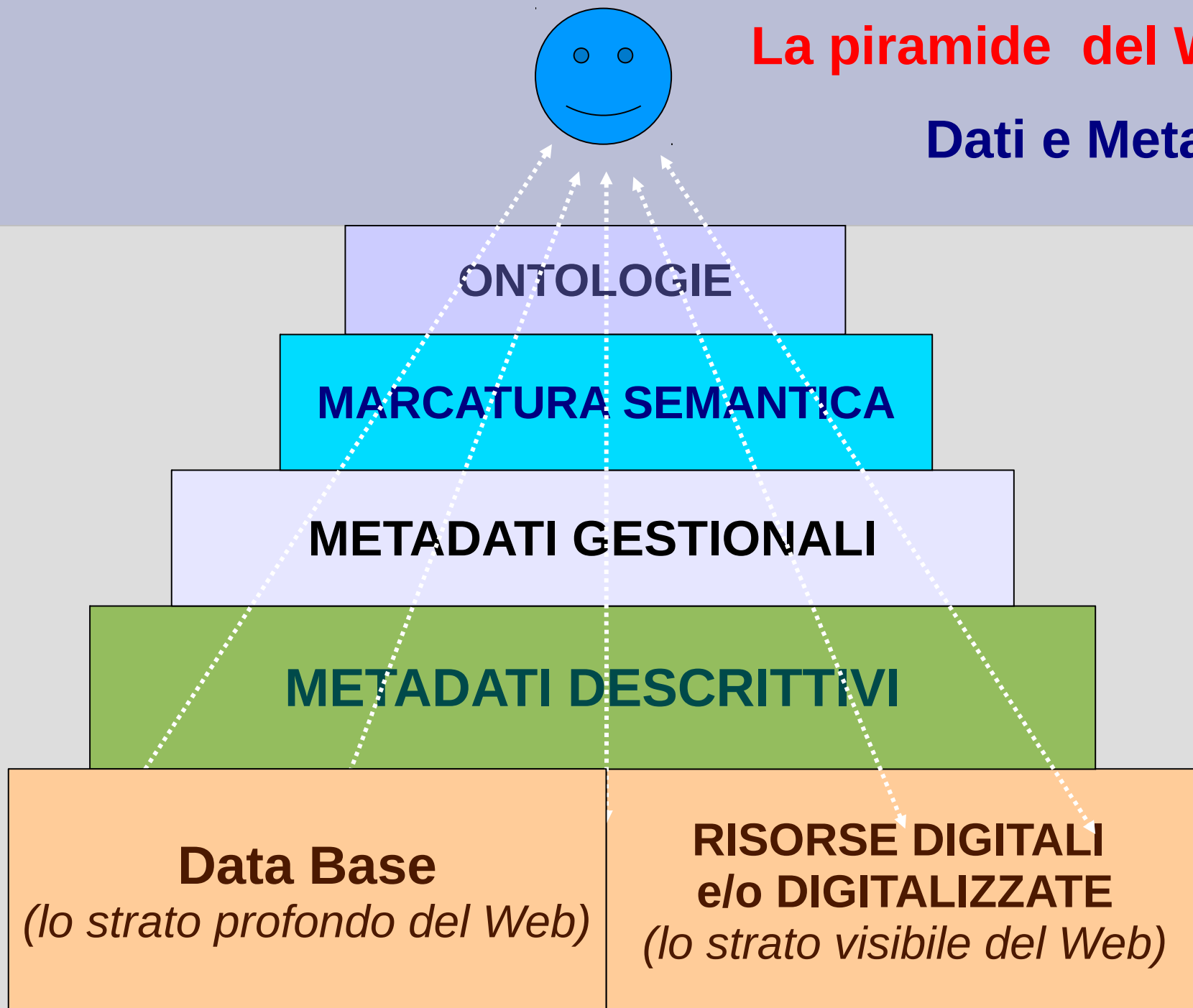
```
<?xml version="1.0" encoding="iso-8859-1"?>  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"  
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd"?>
```

La **presentazione** è resa necessariamente con regole di stile, incluse in ogni documento xhtml (<style type="text/css") oppure, meglio, definite in un file CSS esterno e richiamate sempre in apertura:

```
<link rel="stylesheet" href="/css/default.css"  
  type="text/css" />
```

La piramide del Web:

Dati e Metadati



Dati e metadati

Le funzioni dei metadati

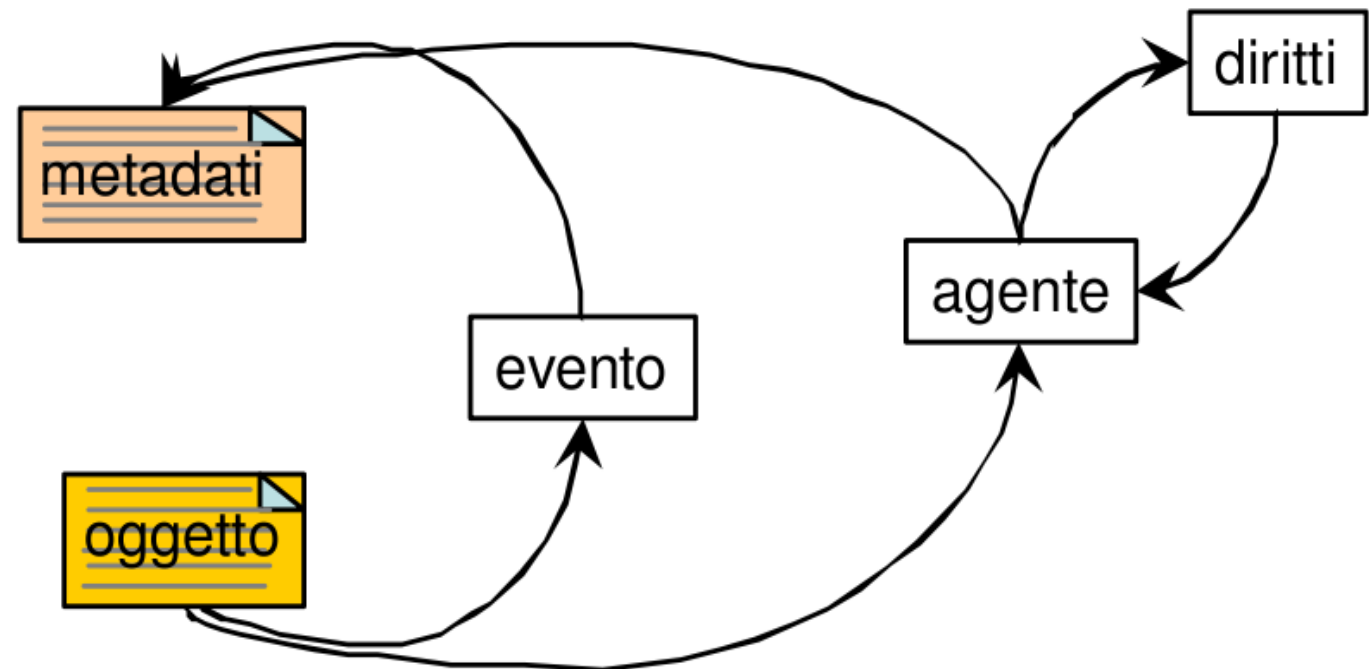
sono consentire il raggiungimento di alcuni obiettivi:

- **Ricerca**, individuare l'esistenza di un documento;
- **Localizzazione**, rintracciare una particolare occorrenza del documento;
- **Selezione**, analizzando, valutando e filtrando in una serie di documenti;
- **Gestione**, gestire le raccolte di documenti grazie all'intermediazione di banche dati, repository e cataloghi;
- **Disponibilità**, ottenere informazioni sull'effettiva disponibilità del documento
- **Conservazione**, gestire l'insieme complesso di attività che garantiscono nel tempo la fruizione

Uno schema per dati e metadati

Nel ciclo di vita di un oggetto digitale

Metadati = Biografia dell'oggetto digitale



Tipologie di metadati: 1 - Metadati descrittivi

servono per l'**identificazione ed il recupero** degli oggetti digitali; sono costituiti da **descrizioni** delle risorse, nate in formato digitale o digitalizzate; possono risiedere **all'esterno** dell'archivio digitale oppure essere **inseriti nelle risorse** o ancora **annidati in altri metadati**:

1. *Handle*, PURL (Persistent Uniform Resource Locator). DOI...
3. Dublin Core e MODS (metadati descrittivi)
4. I *collection description Schemas* (già citati)
5. I Meta Tag di HTML (descrittori all'interno della risorsa)
6. Encoded Archival Description (EAD) e EA Context (EAC)
7.

Dati e metadati

Tipologie di metadati: 2 - Metadati gestionali e di conservazione

Memorizzano le **informazioni tecniche sulla creazione e la manutenzione delle risorse** nel sistema di gestione dell'archivio digitale e sono necessari per una corretta esecuzione delle relative attività, di sicurezza e non.

Data la labilità dell'informazione elettronica, questi metadati assumono una **grande importanza ai fini della conservazione permanente degli oggetti digitali**: essi possono documentare i processi tecnici associati alla conservazione permanente, fornire informazioni sulle condizioni e i diritti di accesso agli oggetti digitali, certificare l'autenticità e l'integrità del contenuto, documentare la catena di custodia degli oggetti, identificarli in maniera univoca. Esempi sono gli standard **METS, MAG e PREMIS**

Dati e metadati

Tipologie di metadati: 3 - Metadati strutturali

Ne abbiamo già parlato: collegano le varie componenti delle risorse per un'adeguata e completa fruizione, che avviene attraverso la mappatura di schemi di metadati diversi.

Questi metadati, per “tenere insieme” i file, forniscono dati di **identificazione e localizzazione**, come il codice identificativo, l'indirizzo del file sul server, l'archivio digitale di appartenenza e il suo indirizzo di rete. Tipicamente sono espressi in XML, e sono frequentemente inclusi nei profili di applicazione degli standard di metadati gestionali, ad esempio MAG e METS.

Vediamo meglio, per chiarire, il profilo italiano MAG

Dati e metadati: MAG

i metadati amministrativi-gestionali (MAG)

MAG è un *profilo applicativo* (AP) espresso in XML che offre le specifiche formali per la fase di **raccolta e di trasferimento dei metadati e dei dati digitali** nei rispettivi archivi. Esso è indipendente da specifici hardware e software e comprende al suo interno alcuni elementi descrittivi.

Il set MAG, più che per particolari tipologie di documenti fonte (ad es. periodici, musica a stampa, grafica, ecc.), viene definito per tipologie di oggetti digitali (es. immagini statiche, testi prodotti con tipologia OCR, suono, audiovisivi, ecc.).

Dati e metadati: MAG

Lo **schema dell'AP** (metadigit.xsd) **è composto da diverse sezioni**, alcune generali e altre specifiche del contenuto digitale:

<gen> informazioni generali sul progetto e sul tipo di digitalizzazione

<bib> metadati descrittivi sull'oggetto analogico

<stru> metadati strutturali

**** metadati specifici relativi alle immagini fisse

<ocr> metadati specifici relativi al riconoscimento ottico del testo

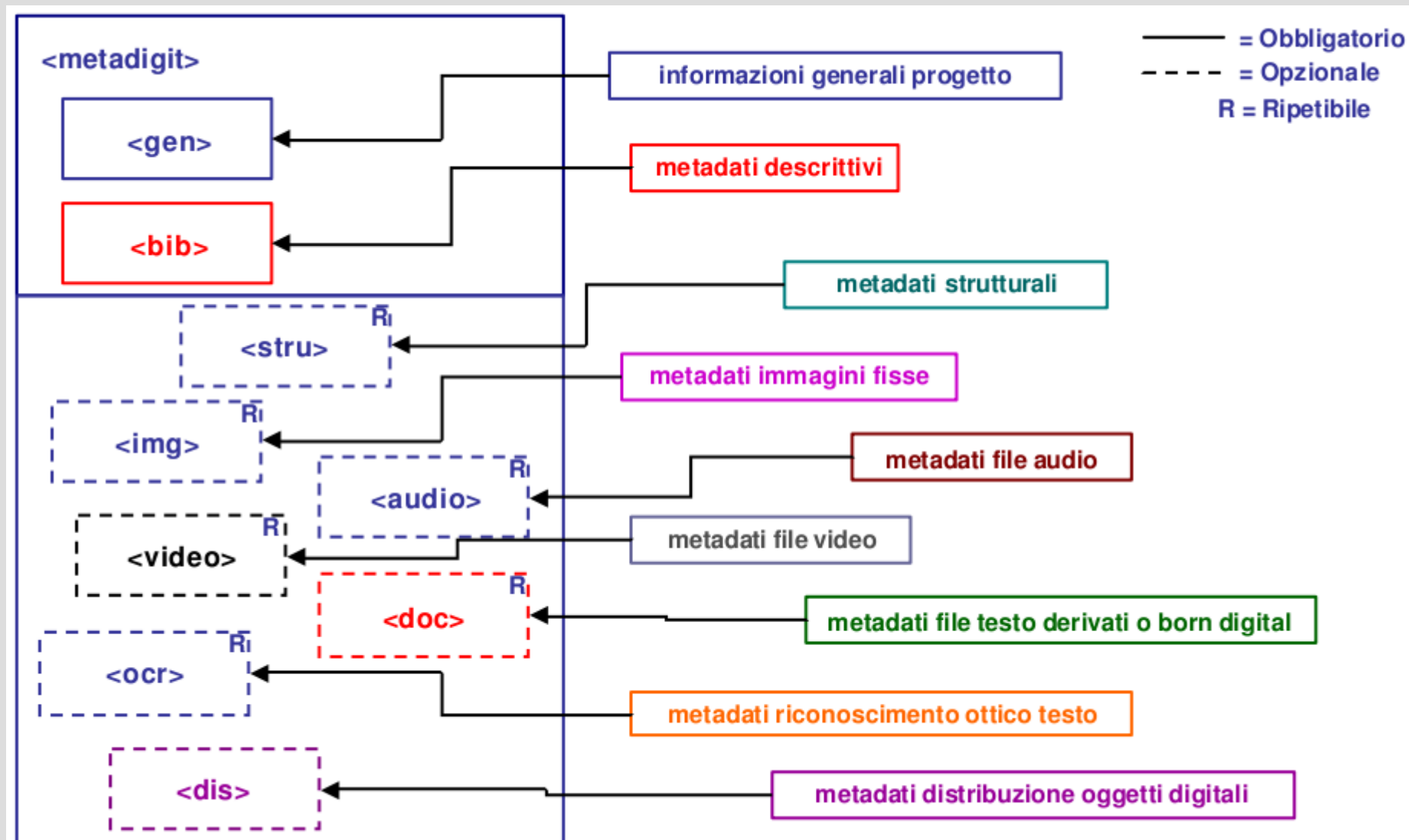
<doc> sezione utilizzata per descrivere un file testo (es. pdf o rtf)

<audio> metadati specifici per file audio

<video> metadati specifici per file video

<dis> metadati specifici per la distribuzione di oggetti bibliografici

Dati e metadati: MAG



Il Web 2.0



Web 2.0 è un termine utilizzato per indicare uno stato dell'evoluzione del World Wide Web, rispetto alla condizione precedente.

Si tende a indicare come Web 2.0 l'insieme di **tutte quelle applicazioni online che permettono uno spiccato livello di interazione tra sito e utente** (blog, forum, chat, wiki, flickr, youtube, facebook, myspace, twitter, google+, linkedin, wordpress, foursquare, ecc.)

Il Web 2.0

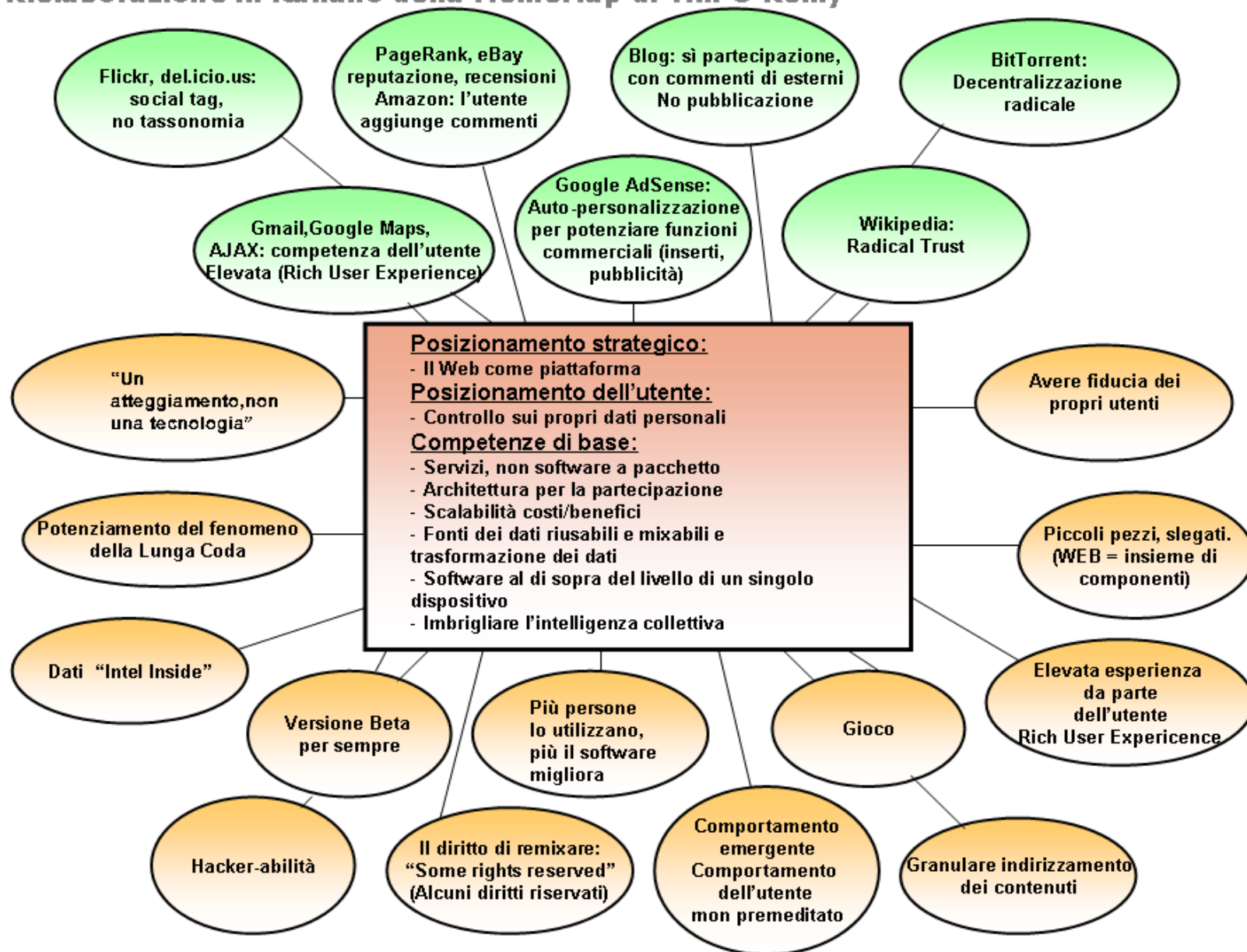


Il termine Web 2.0 è stato coniato dal guru Tim O' Reilly per porre l'accento sulle **differenze** rispetto al cd. **Web 1.0**, fino agli anni '90, composto prevalentemente da **siti web statici**, quasi senza possibilità di interazione con l'utente eccetto la normale navigazione tra le pagine, l'uso delle e-mail e dei motori di ricerca.

Per le applicazioni Web 2.0, spesso vengono usate **tecnologie di programmazione particolari**, come AJAX (Gmail usa largamente questa tecnica) o Adobe Flex.

Il Web 2.0

Rielaborazione in italiano della MemeMap di Tim O'Reilly



Il Web 2.0

Da un punto di vista strettamente tecnologico, il Web 2.0 è del tutto equivalente al Web 1.0, in quanto l'infrastruttura di rete continua ad essere costituita da **TCP/IP** e **HTTP** e l'**ipertesto** è ancora alla base delle relazioni tra i contenuti.

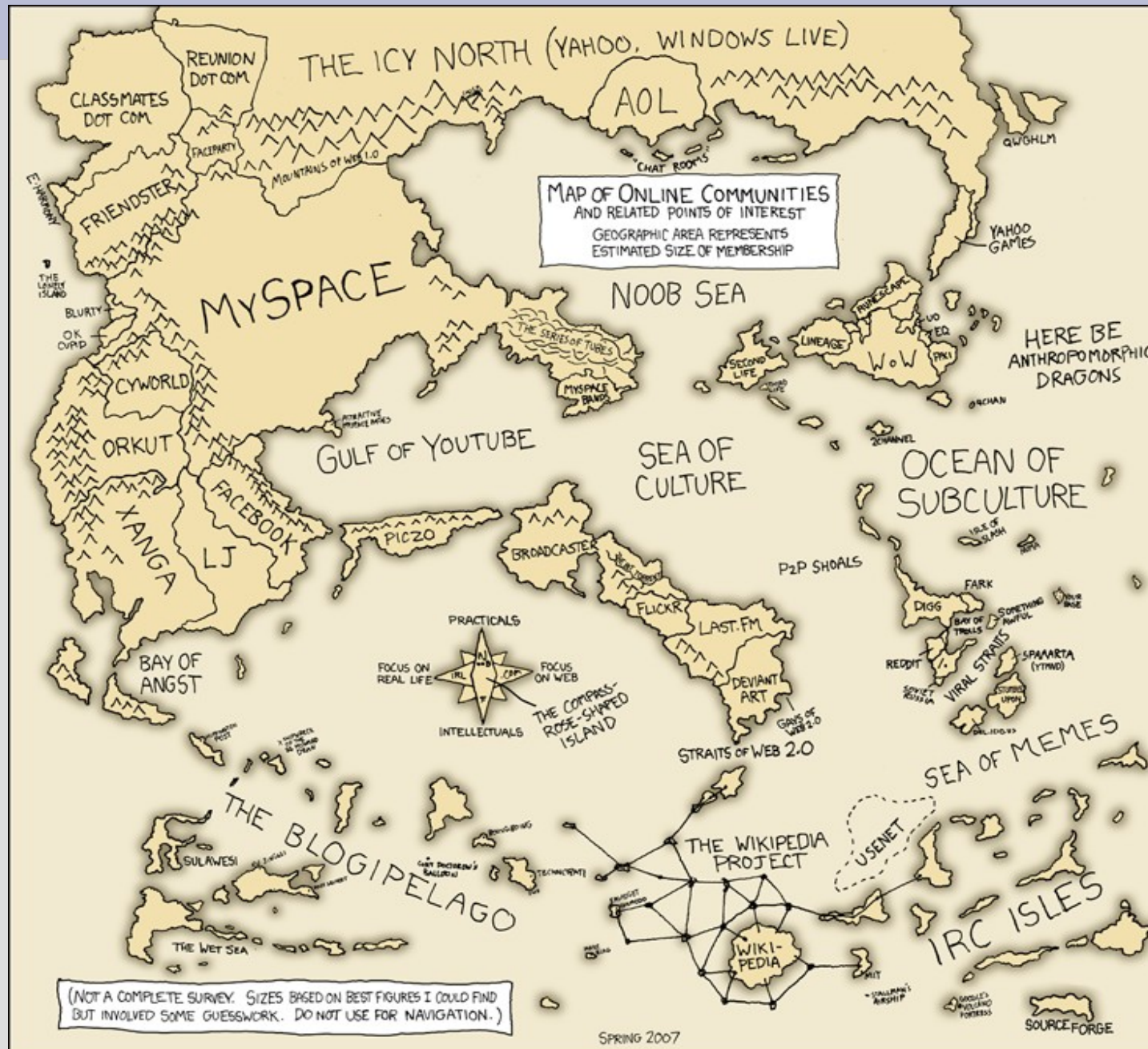
La differenza, più che altro, sta nell'**approccio con il quale gli utenti si rivolgono al Web**, che passa fondamentalmente dalla semplice consultazione alla possibilità di contribuire popolando e alimentando il Web con propri contenuti.

Il Web 2.0

Riuscite ad immaginare, oggi, un Web senza social networks, senza Google maps, senza YouTube, senza tagging delle foto, senza blogs, senza feed RSS e sharing dei contenuti, dei video, della musica, accessibile anche da dispositivi mobili?

Ecco, il web fino al 2003 circa era così: siti “istituzionali” e siti personali, statici, solo da fruire e diversi motori di ricerca che caricavano le pagine di messaggi pubblicitari in modo incomparabile con oggi.

Web 2.0: Map of online communities



oltre il sito web: il Web 3.0

il web semantico

Condividere la conoscenza sul web significa poter disporre di strumenti e tecnologie che consentano di *esprimere* i contenuti, *strutturarli* e *presentarli* in modo adeguato, rendendone esplicito il significato, la *semantica* e consentendo la fruizione dell'informazione a tutti, indipendentemente dal particolare retroterra culturale e dal contesto tecnologico.

Oggi i contenuti sono tantissimi, ricchi e vari in quanto a *medium*, ma l'unico criterio di ricerca, di individuazione, di selezione restano **le parole associate, più o meno strutturate**

oltre il sito web

La **ricerca di informazioni** è uno dei principali **punti deboli del web**, nonostante il gran numero di motori di ricerca esistenti, che sono poveri di semantica sia in fase di indicizzazione che in fase di ricerca. In fase di ricerca viene consentito di combinare le parole con operatori di contesto (“tutte le parole”, “una parola qualunque”, “nel titolo”), ma in definitiva il risultato scaturisce sempre da una ricerca sulla presenza di parole chiave e dall’identificazione dei documenti più affini alla domanda posta.

Un elemento rivoluzionario è che le conoscenze codificate nel semantic web sono **rappresentate in forma elaborabile**, e quindi possono essere **utilizzate da componenti automatizzati, denominati *agenti software***. Qualcosa di vicino all'intelligenza artificiale.

oltre il sito web

Il **Semantic Web** si basa su una architettura a livelli (cfr. Figura dell'inizio). L'intera comunità scientifica sta investendo molte energie nel settore del Semantic Web
<http://www.semanticweb.org/>

La sfida del semantic web, quindi, è fornire **un linguaggio per esprimere dati e regole per ragionare sui dati**, con l'esportazione sul Web di **qualsunque sistema di rappresentazione della conoscenza**.

XML (con *Name Space* e *xmlschema*) consente di dare ai documenti **una struttura arbitraria**.

RDF, uno schema XML, si usa per **esprimere il significato**, asserendo cioè che alcuni particolari elementi hanno delle proprietà (p.es. essere *autore-di*, *costruito-da*, *nato-il...*).

oltre il sito web: RDF

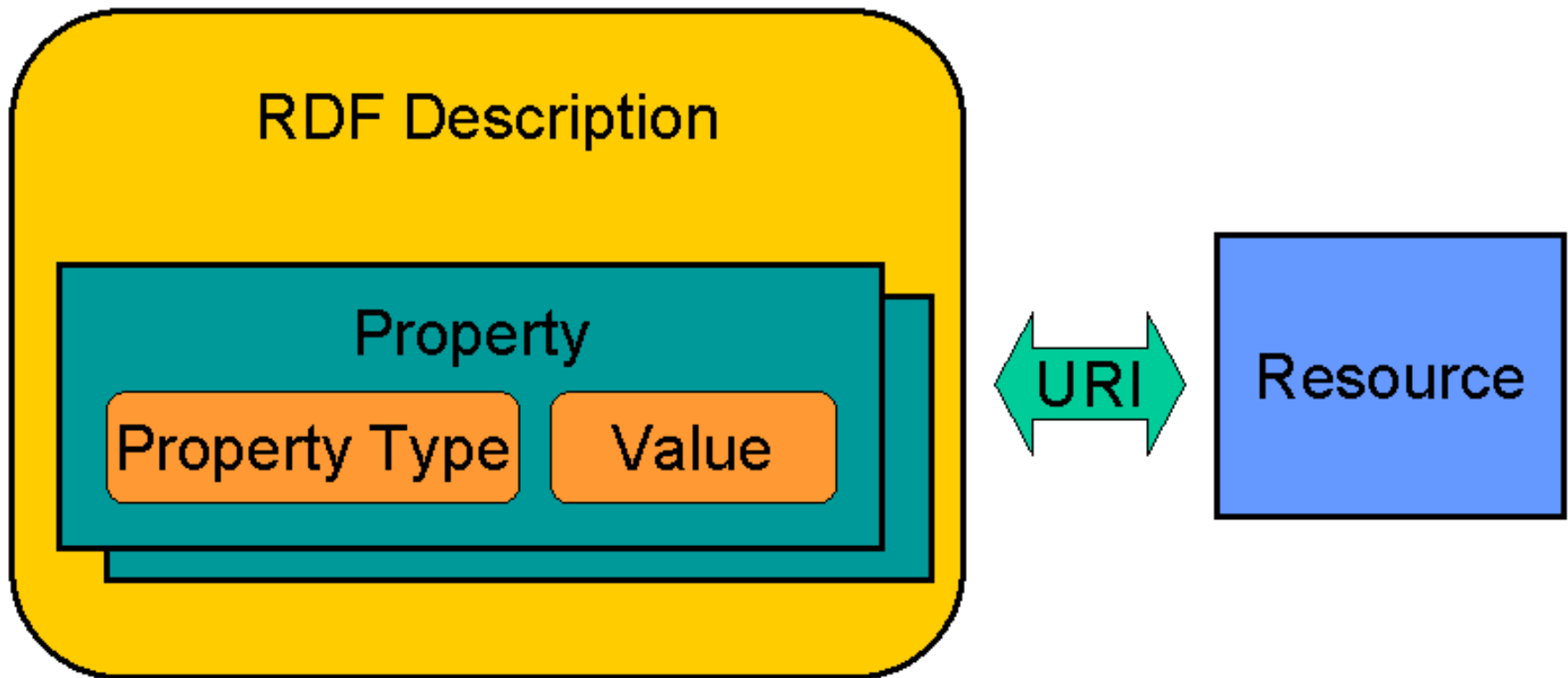
Resource Description Framework fornisce il fondamento per l'interoperabilità di metadati tra **differenti comunità di descrizione** delle risorse. Il problema è la molteplicità di standard, incompatibili per definizioni sintattiche e per schemi di metadati.

L'RDF Data Model si basa su **tre principi chiave**:

- Qualunque cosa **può essere identificata** da un Universal Resource Identifier (URI).
- *The least power*: utilizzare il linguaggio **meno espressivo** per definire qualunque cosa.
- Qualunque cosa **può dire qualunque cosa su qualunque cosa**.

oltre il sito web: RDF

Qualunque cosa descritta da RDF è detta **risorsa**, reperibile sul web e identificata da un URI.



oltre il sito web: RDF

Esempio di RDF/XML

Si deve “serializzare” l'affermazione: "Mario_Rossi"
"è autore di" "Rosso_di_sera_bel_tempo_si_spera"

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:au="http://description.org/schema/">
  <rdf:Description
    about="http://www.book.it/Rosso_di_sera_bel_tempo_si_spera/">
    <au:author>Mario_Rossi</au:author>
  </rdf:Description>
</rdf:RDF>
```

oltre il sito web: RDF

La stessa affermazione, adottando uno schema di metadati descrittivi in XML, sarebbe stata espressa più o meno così

```
<?xml version="1.0" encoding="UTF-8"?>
<resource>
<type>book</type>
<creator>Mario Rossi</creator>
<title>Rosso di sera bel tempo si spera</title>
<source>http://www.book.it/Rosso_di_sera_bel_tempo_si_spera/</
source>
</resource>
```

oltre il sito web: ontologie

Un terzo componente necessario per il web semantico, oltre a XML e RDF è l'**Ontology Vocabulary** (*livello ontologico*), inteso come il contenitore che definisce in modo formale le relazioni fra i termini.

Il linguaggio definito dal W3C per scrivere ontologie strutturate, in architettura web, è **OWL** (Ontology Web Language).

Le ontologie devono rappresentare solo **un dominio di conoscenza definito** ed è preferibile che siano connesse tra di loro tramite **core ontologies**, di livello più generale.

il web che cambia: le applicazioni

- Un discorso a parte meritano le **applicazioni**, o **apps**, dedicate in particolare ai dispositivi mobili (mobile devices: smartphones, tablets, ...) che stanno secondo alcuni addirittura uccidendo il web in senso proprio.
- In informatica con il termine applicazione si intende il software che, in esecuzione grazie all'hardware, rende possibile **un servizio o una serie di servizi o strumenti** utili e selezionabili a desiderio dell'utente.

il web che cambia: le applicazioni

Il web interattivo, interoperabile e multi-canale ha aperto spazi immensi di business per l'offerta di piccole applicazioni di storage o di produttività, residenti nella cosiddetta *cloud*, oppure che recuperano/rielaborano dati dal web per renderli disponibili più facilmente.

Nel primo caso, esistono centinaia di offerte, gratuite o a prezzi competitivi, di **memoria web ad accesso privato** per i nostri *file*, sia per sicurezza che così da averli a disposizione dovunque siamo e con qualunque dispositivo ci connettiamo.

il web che cambia: le applicazioni

Nel secondo, sono stati sviluppate *suite* di produttività ed altre applicazioni unicamente residenti nel web, gratuite o a pagamento. **Google Apps**, ad esempio, che va da Gmail a Google Documents a servizi personalizzati per le imprese, oppure **Microsoft Web App**, che va dai CMS web a strumenti per lo sviluppo di portali e di siti di ecommerce, ma anche applicazioni specializzate (utility o giochi) per gli smartphones e i tablet PC: **Samsung Apps, Android Apps, Apple Apps Store...**

Tra queste ultime, quelle che rielaborano dati web “aperti”, si pensi a servizi come Meteo, orari dei treni, news, mappe, etc.

Gli open data

A quest'ultimo proposito, vale la pena ricordare che molti dei dati su cui si basano le Apps sono ad accesso pubblico.

Da qualche anno si sta facendo in modo che i dati pubblici (delle pubbliche amministrazioni, prima di tutto) siano resi **aperti**, ovvero disponibili sul web, e che lo siano **in modo strutturato ed elaborabile** (nel solito XML...).

Si parla di *open government* e di *open data*: ha iniziato Obama negli USA con data.gov, hanno seguito vari governi europei, tra cui l'Italia: dati.gov.it, cui è collegato il progetto **Apps4Italy**,
<http://www.dati.gov.it/content/applicazioni-smartphone> .