

Introduzione e breve storia del World Wide Web

Angelo Di Iorio
Università di Bologna



Cos'è il WWW?

- Il World Wide Web è un **sistema ipertestuale di documenti multimediali**
- Distribuito e scalato su Internet
- **Contenuti** testuali e non – sempre più **dati** – pensati per essere "consumati" sia da esseri umani che da applicazioni
- Una **knowledge-base** su cui costruire servizi sofisticati per gli utenti finali

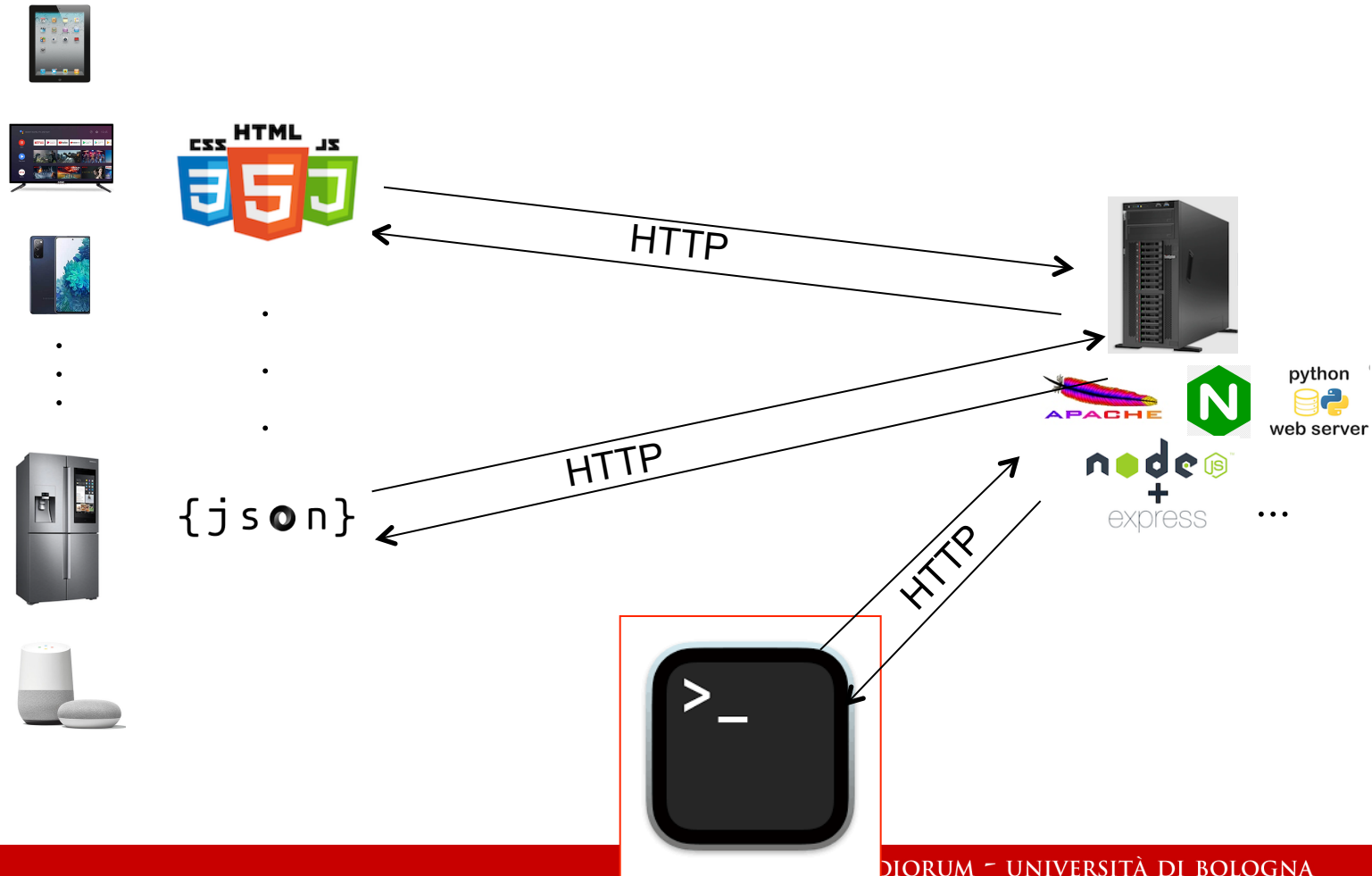


Architettura client-server

- Il World Wide Web si basa su un'architettura **client-server** :
 - Il **client** chiede le risorse al server
 - Il **server** è in attesa di richieste e, ricevuta una richiesta di risorsa, la restituisce al client
 - altre entità intermedie possono intervenire in questo flusso di comunicazione (es. proxy)
- Il **browser** è un client Web in grado di visualizzare i documenti e interagire con l'utente durante la navigazione
- Il client non ha necessariamente un'interfaccia grafica ma
 - può interfacce diverse, es. **screen reader** o **interfacce vocali**
 - non avere un'interfaccia utente, ma essere a sua volta un'applicazione che processa e combina i dati ricevuti



Client e Server Web





Protocolli e standard fondamentali

- Alla base di WWW ci sono i tre protocolli/linguaggi:
 - **URI**: standard per identificare in maniera generale risorse di rete e per poterle specificare all'interno di documenti ipertestuali
 - **HTTP**: protocollo di comunicazione stateless e client-server per l'accesso a risorse ipertestuali
 - **HTML**: linguaggio per la marcatura di documenti ipertestuali basato su SGML (e XML) che permette di descrivere la struttura di un documento e le sue componenti, inclusi oggetti multimediali e link, e di visualizzare i contenuti sul browser



...ma molti altri al centro della scena

- A questi si aggiungono molti altri standard, diventati sempre più centrali nello sviluppo di applicazioni Web
 - **CSS**: linguaggio di presentazione per decidere gli aspetti tipografici e di formattazione
 - **JavaScript** e **TypeScript**: linguaggio di programmazione client-side per aggiungere comportamenti dinamici
 - **JSON**: formato "leggero" di interscambio dati basato su Javascript
 - ...
- Ma è sempre stato così? Come si sono evoluti e si evolvono questi standard? Quali "forze" entrano in gioco?



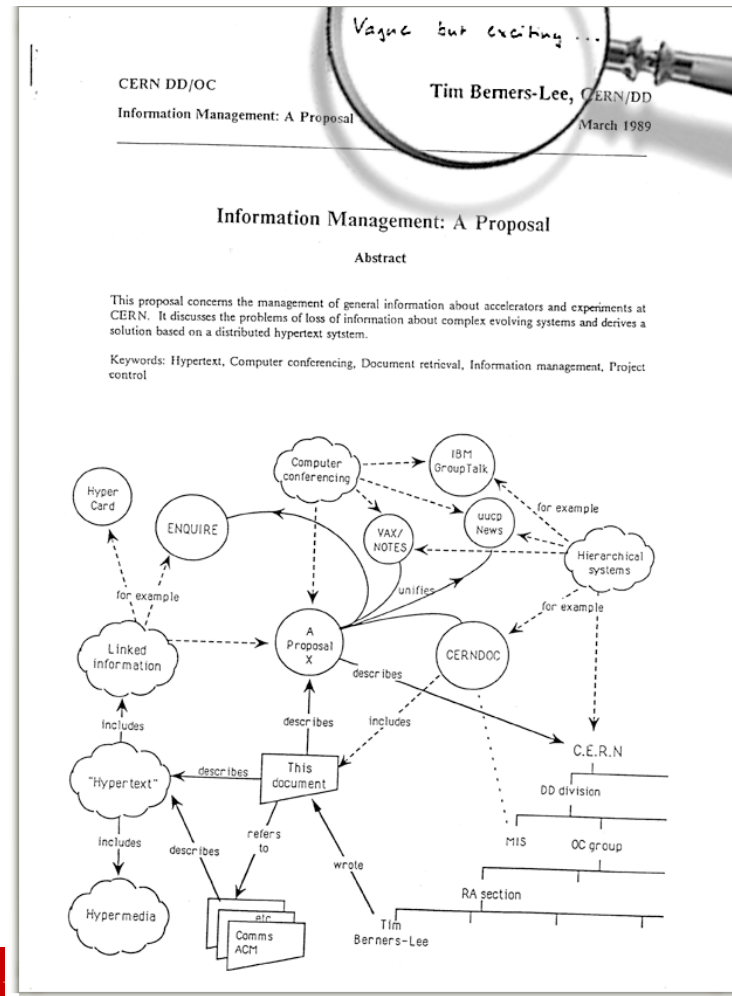
La nascita del World Wide Web

- Il World Wide Web nasce nel 1989 da un'idea di Tim Berners-Lee, Robert Cailliau e altri ricercatori del CERN, il centro di ricerca in fisica nucleare di Ginevra
- L'obiettivo era costruire una piattaforma per la diffusione e ricerca di articoli e informazioni tra i ricercatori
- Tim Berners Lee e gli altri identificarono tre pilastri su cui fondare la loro architettura:
 - **Internet**, ossia una rete di calcolatori collegati tra loro
 - **Iper testi**, ossia documenti collegati tra loro
 - **SGML**, un metalinguaggio per descrivere formati di interscambio di dati e documenti, derivato dalla proposta GML di IBM



La proposta originale

- Il primo commento all'idea di Tim Berners Lee: "Vague but exciting..."
- Anche gli informatici furono inizialmente freddi.
- Nel 1991, alla conferenza sugli ipertesti, Berners-Lee e Cailliau mostrarono con poco successo il loro primo prototipo, accolto invece con entusiasmo dai fisici





Digressione

World Wide Web non è sinonimo di Internet

The screenshot shows a news article on the R.it website. The page header includes the R.it logo and the word 'Tecnologia'. A navigation bar contains links for News, App, Social Network, Mobile, Videogiochi, Sicurezza, Prodotti, Interattivi, and Video. A red banner advertises 'Vodafone ADSL solo online a un prezzo speciale' with a 'Scopri' button. Below the banner are social media sharing icons for Facebook (139), Twitter (33), Google+ (16), LinkedIn (6), and Indoon. The main headline reads: 'Internet compie 25 anni. Tim Berners-Lee: "Il web deve essere un diritto di tutti"'. A sub-headline states: 'Il www, nato come rete militare e poi come risorsa accademica, esisteva in forme via via più evolute fin dagli anni '60. Il 12 marzo 1989 Tim Berners-Lee depositò al Cern la proposta di creare l'infrastruttura digitale'. To the right of the article is a blue box for 'UN PROGETTO DI la Repubblica AF AFFARI & FINANZA', mentioning collaborations with CONAD and nielsen.

[Disclaimer: non è un attacco a Repubblica! 😊]



Il WWW su larga scala e la guerra dei browser

- Negli anni 1992-1994 il World Wide Web si afferma su scala mondiale grazie al browser Mosaic e Netscape
- Dal 1994 al 1999 sia Netscape che Microsoft introducono piccoli "miglioramenti" su HTML per aumentare l'appeal del proprio browser
- Con Explorer 3.0 (pre-installato con Windows) e poi 4.0 Microsoft vince definitivamente e dominerà la scena per i successivi cinque anni
 - Microsoft sarà anche condannata per questa operazione commerciale e "abuso di posizione dominante"
- Da Netscape nascerà poi Firefox e da lì la comunità di Mozilla



World Wide Web Consortium

- Per mitigare l'effetto di queste guerre commerciali e sbilanciamenti nello sviluppo del WWW, Berners-Lee e Cailliau fondano il W3C
- Il Consorzio prevede la partecipazione di aziende, enti di ricerca, università, istituzioni pubbliche e private. A fine 2020 conta 460 membri.
- Il W3C si è sempre battuto per garantire accessibilità al Web da parte di tutti e libertà d'espressione
- Ha avuto un'importanza fondamentale durante la guerra dei browser nel costruire un tavolo comune di discussione e integrazione delle caratteristiche di HTML introdotte dagli uni e dagli altri
- Successivamente, il W3C ha diretto lo sviluppo dei più importanti standard del Web tra cui HTTP, XML, CSS, WAI, etc.



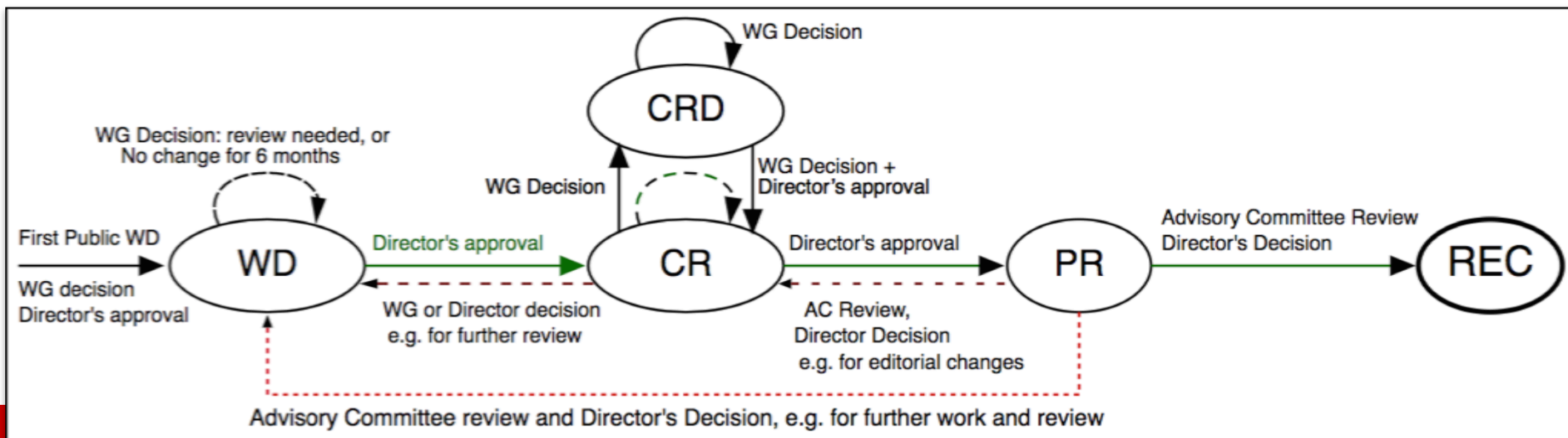
Standard W3C e consenso

- Il W3C promuove processi **rigorosi e democratici** di standardizzazione che puntano ad un largo consenso da parte della comunità
- Partendo dall'interesse della comunità su un dato argomento, il W3C forma gruppi di lavoro di diverso tipo:
 - **Community Group** e **Interest Group**, pensati per lo scambio di idee e proposte
 - **Working group**, che hanno il compito di produrre documentazione e standard
- Questi gruppi sono lavorano alla redazione di documenti di standard, attraverso fasi ben definite, fino alla loro approvazione come **Candidate Recommendation**



Standardizzazione W3C

- Le fasi del processo di standardizzazione: *working draft*, *candidate recommendation*, *proposed recommendation*, *recommendation*
- Necessarie diverse approvazioni, sia del working group che dei direttori, e di revisione pubblica
 - Un punto importante è la verifica del supporto delle feature proposte in almeno due implementazioni software





W3C e HTML

- Un processo così rigoroso va di pari passo con lo sviluppo di standard rigorosi
- Il caso di HTML è significativo:
 - HTML4 è la versione standard dopo la prima guerra dei browser
 - Nel 2001 il W3C approva XHTML 1.0 che usa regole rigorose di buona forma e validazione senza introdurre nuovi elementi
 - Nel 2002 iniziano i lavori su XHTML 2.0 che introduce nuovi elementi (es. SECTION per strutture gerarchiche) ed è progettato da esperti di linguaggi di marcatura ma **senza il supporto dei produttori di browser, che invece spingono per maggiore interattività e "programmabilità"**
- A questo punto il W3C perde il ruolo di centralità nello sviluppo di HTML e il suo ruolo cambia, così come cambia ruolo stesso di HTML



WHATWG e HTML5

- Ci sono due forze contrastanti:
 - i produttori di browser spingono per lo sviluppo di nuove versioni di HTML, con enfasi su aspetti di interattività e programmazione
 - il W3C che spinge per specifiche rigorose e basate su XML
- Il W3C non accetta le proposte dei produttori di browser che formano e finanziano un gruppo separato, il Web Hypertext Application Technology Working Group (WHATWG)
- Inizialmente fondato da Apple, Mozilla Foundation e Opera Group, oggi WHATCG è guidato da Apple, Google, Mozilla Foundation e Microsoft



HTML Living Standard

- Il “nuovo” HTML è, per scelta del WG, una specifica perennemente in sviluppo (“*living standard*”):

<https://html.spec.whatwg.org/>

- Questa instabilità diventa una feature:

*"The WHATWG HTML spec can now be considered a "**living standard**". It's more mature than any version of the HTML specification to date, so it made no sense for us to keep referring to it as merely a draft."*

- Cambia completamente il modello di sviluppo del linguaggio che si allontana sensibilmente dall’approccio sistematico e democratico (ma non privo di difetti) di evoluzione degli altri standard W3C
- **E’ la vittoria indiscussa dei produttori di browser: sia per le nuove caratteristiche del linguaggio sia per il modo in cui è sviluppato**



Due versioni di HTML?

- Nel 2007 il W3C dovette ammettere che queste modifiche avevano un impatto innegabile e riaprì il working group con tutti i membri del WHAT per creare una nuova versione di HTML
- Nel Working Group convivono quindi le due “anime” ma con scarsi risultati e nel 2011 i gruppi si dividono
- Lo sviluppo va avanti in parallelo: il W3C continua a standardizzare "snapshot" di "HTML Living Standard" e a dargli un'approvazione formale
- Nel 2019 W3C e WHATCG raggiungono un accordo e si impegnano a sviluppare un'unica versione di HTML, nell'ottica di Web come Open Platform



Cos'è il Web 2.0?

- Una delle buzzword più usate, forse la più usata, nella storia del Web
- Non un cambio di tecnologie ma un uso diverso delle tecnologie esistenti
- L'affermazione di strumenti di partecipazione attiva alla costruzione di contenuti per il web: blog, wiki, podcast, RSS, folksonomie, ecc.
- Il Web diventa quindi una piattaforma, fortemente orientata al mash-up di dati da sorgenti diverse, su cui **pubblicare contenuti con facilità ed elaborare dati**
- Le applicazioni Web diventano molto più veloci, complesse ed interattive



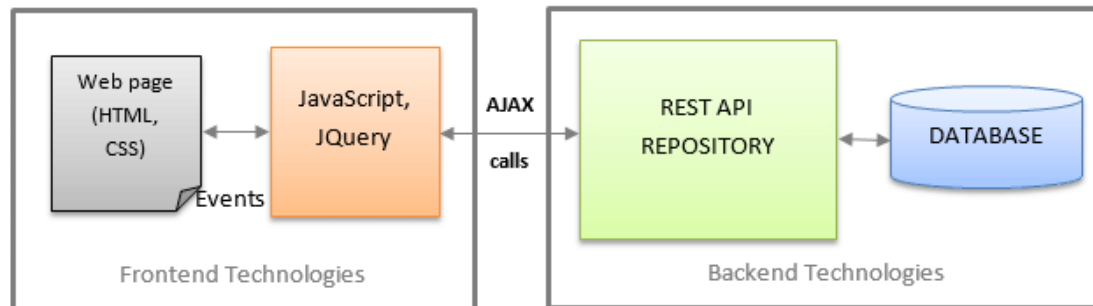
Web 2.0, REST e AJAX

- Alcuni fattori sono stati determinanti in questo cambiamento:
 - Enorme diffusione dei dispositivi mobili
 - Avanzamento tecnologico di questi dispositivi che ormai hanno elevate capacità di memoria ed elaborazione
 - Accesso a Internet diffuso e (ultra) veloce
- Per quanto riguarda le tecnologie Web, due aspetti ortogonali:
 - Affermazione del paradigma REST per sfruttare le caratteristiche di HTTP e creare applicazioni solide e scalabili
 - Sviluppo dei linguaggi di programmazione client-side e di AJAX, che permette richieste asincrone al server e pre-caricamento di dati



Rich Client ed API

- Le pagine Web diventano la base per applicazioni che aggregano contenuti, anche da sorgenti diverse e in modo sofisticato (Rich Client)
- Sono numerosi i servizi che producono SOLO i DATI; è il client ad occuparsi della loro elaborazione
- **Diventa sempre più importante fornire i propri dati in modo chiaro attraverso un'API (Application Programming Interface)**





API Web

- Così come per i linguaggi di programmazione, un'API Web definisce le modalità per interagire con un'applicazione
- In pratica, elenca le possibili richieste e risposte che l'applicazione è in grado di gestire
- Le applicazioni, sia server-side che client side, possono quindi raccogliere i dati per poi elaborarli invocando una o più API:
 - applicazioni diverse costruite sugli stessi dati
 - applicazioni di *mash-up* che combinano dati in formati diversi e da sorgenti diverse

Maggior dettagli
in lez. API REST





Dal Web of Content al Web of Data

- I contenuti Web sono "letti" non solo da esseri umani ma sempre più da agenti software
- Questo era possibile anche in passato ma molto più complicato – bisognava estrarre informazioni e fare *scraping*, cosa ancora necessaria oggi in assenza di API
- Il Web diventa quindi una piattaforma di **dati intercollegati** e non solo di contenuti ipertestuali
- Da link semplici si passa a **link complessi e di diverso tipo** in grado quindi di esprimere relazioni tra oggetti e costruire una **base di conoscenza distribuita**



Semantic Web

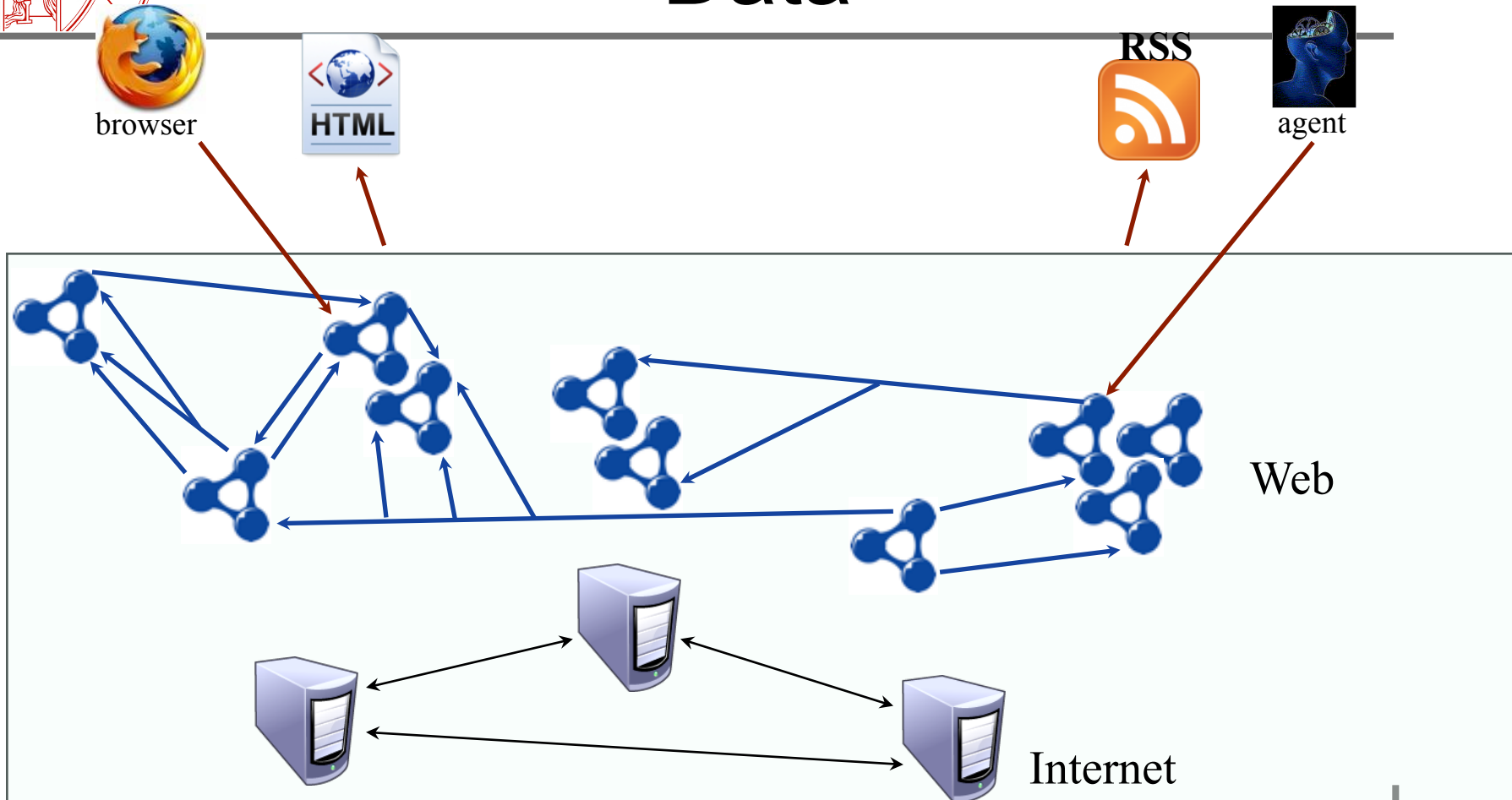
“The Semantic Web is a **web of data.**”

“The Semantic Web is about two things. It is about **common formats** for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about **language** for recording **how the data relates to real world objects.** That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected **not by wires but by being about the same thing.**”

Introduction to the W3C Semantic Web Activity



Dal Web of Content al Web of Data





Dati e affermazioni

- Il Semantic Web usa un modello di rappresentazione dei dati basato su triple, chiamato RDF (a volte chiamato impropriamente “linguaggio”, esistono diverse sintassi per esprimere le stesse informazioni)
- Basato su **triple soggetto-predicato-oggetto** dette **statement**
- *“Umberto Eco è autore de Il Nome della Rosa”* può essere espresso come affermazione RDF assegnando a:
 - “Umberto Eco” il ruolo di soggetto
 - “è autore di” il ruolo di predicato
 - “Il Nome della Rosa” il ruolo di oggetto



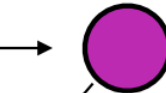
Esempio in RDF

`<https://twitter.com/umbertoeco_>`

IRI che identifica
Umberto Eco su
Twitter



`<http://xmlns.com/foaf/0.1/made>`



IRI che identifica IL
nome della rosa su
Anobii

`<http://xmlns.com/foaf/0.1/name>`

`<http://purl.org/dc/terms/title>`

“Umberto Eco”

“Il nome della rosa”



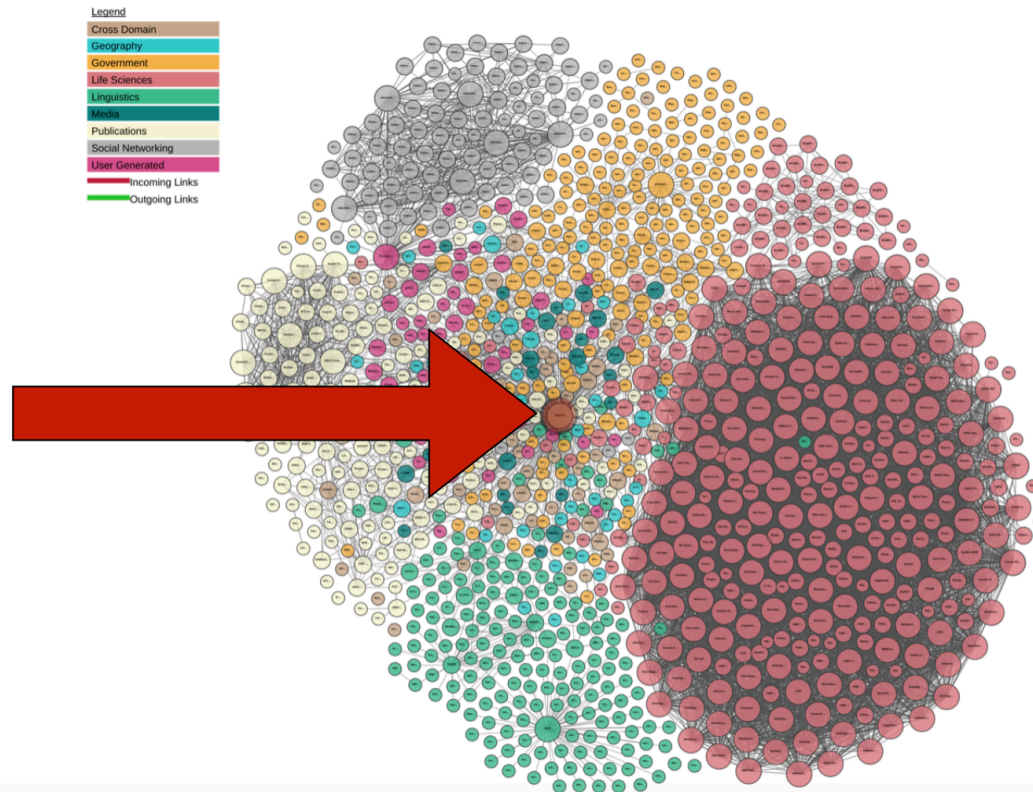
Linked (Open) Data

- Linked Data riguarda l'uso del Web per creare link **tipati** tra **risorse** appartenenti a domini differenti, in modo da esprimere relazioni e proprietà
- Obiettivo: pubblicare dati sul Web in modo che
 - siano machine-readable
 - con un significato esplicitamente definito
 - e abbiano collegamenti verso altri insiemi di dati (dataset) esterni
- Linked Open Data (LOD) è un progetto del W3C che si occupa di estendere il Web tradizionale pubblicando dataset liberi e aperti e mettendo in relazione tra loro dati provenienti da sorgenti diverse



LOD Cloud nel 2017

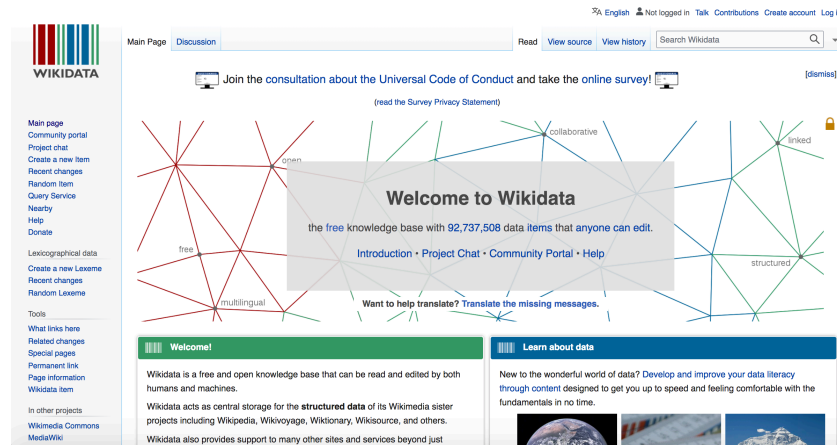
- <https://lod-cloud.net/>





WikiData

- WikiData (<https://www.wikidata.org/>) è una knowledge-base collaborativa, aperta e multilingua
- Costituita da **item (entità)** con un codice identificativo univoco e sui quali sono espresse **proprietà e relazioni** tramite **affermazioni**
- I dati possono essere modificati direttamente e/o importati da altre fonti, inclusa WikiPedia (così come fatto da DBPedia)





Google Knowledge Graph e Schema.org

- Anche Google mette a disposizione un'API per accedere al proprio *Knowledge Graph*, una base di conoscenza usata dal motore di ricerca e che contiene informazioni estratte da Wikipedia e da altre sorgenti (Freebase, CIA World Factbook)
- <https://developers.google.com/knowledge-graph>
- Usa classi e proprietà definite su <https://schema.org/> ed espresse in formato RDF e JSON-LD quindi compatibile con tecnologie Linked Data
- *Schema.org* è un progetto inizialmente fondato da Google, Microsoft e Yahoo e ora supportare da una comunità open-source per definire vocabolari utili a descrivere *Persone*, *Luoghi*, *Eventi*, *etc.*
 - Classi e proprietà specializzate e condivise da applicazioni diverse



Conclusioni

- Il World Wide Web ha una storia relativamente recente ma ha visto diverse **evoluzioni** ed **involuzioni**
- Da un sistema di contenuti ipertestuali si è arrivati ad un sistema di dati e applicazioni, con modalità di accesso sempre più variegate
- Le tecnologie di base tuttavia sono sostanzialmente le stesse ma è cambiato il modo in cui sono sfruttate, il supporto hardware e di comunicazione, l'uso da parte degli utenti
- Conoscere questi cambiamenti e il modo in cui le tecnologie sono state sviluppate e influenzate dal contesto aiuta a comprenderle meglio e metterle in relazione