

Seminario “L'interoperabilità nelle Infrastrutture di Dati Territoriali”

Università di Roma “La Sapienza” I°Facoltà di Architettura “Ludovico Quaroni”

Piazza Borghese 9, Aula Magna, 26-17 Febbraio 2004

Interoperabilità e Open Source

G. Bracco

Servizio Centralizzato Informatica e Reti

C.R. ENEA Frascati, Via E. Fermi 45 Frascati (Roma)

bracco@frascati.enea.it, <http://fusfis.frascati.enea.it/~bracco>

Contenuto

- ENEA
- Open Source Software
 - Software il cui codice sorgente è disponibile
- Interoperabilità
 - La capacità di un sistema o di un prodotto di cooperare insieme ad altri sistemi o prodotti senza richiedere sforzi speciali da parte dell'utilizzatore

ENEA - Ente per le Nuove Tecnologie l'Energia e L'Ambiente

11 centri di ricerca + Sede a Roma

Servizio Centralizzato Informatica e Reti (**INFO**)

Servizi Informatici Gestionali
Gestione delle Reti

Calcolo Scientifico INFO

- Centri di calcolo nei centri principali: Casaccia, Frascati, Bologna, Trisaia, Portici
- Risorse multiplatforma per calcolo seriale, parallelo, grafica avanzata

Sistemi di calcolo dipartimentali e individuali



ENEA GRID

Il problema della **interoperabilità** degli strumenti per il calcolo scientifico in ENEA

Risorse di calcolo eterogenee (AIX, Linux, Irix, Solaris) distribuite geograficamente.

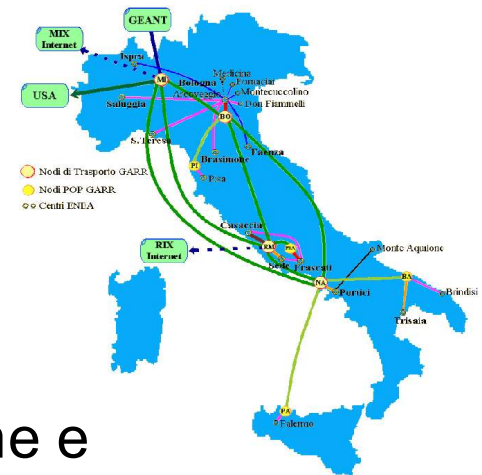
Esigenza di accedere all'insieme dei sistemi in modo integrato: **ENEA-GRID**

- **Architettura di ENEA GRID**

- File system distribuito: AFS
- Gestore di risorse: LSF Multicluster
- Interfaccia di accesso unificata: Citrix Metaframe e EnginFrame

- **Integrazione con i sistemi di calcolo dipartimentali**

- File system distribuito, gestione Licenze



ENEA GRID

- **Software proprietario e open source in ENEA GRID**
 - AFS ==> OpenAFS IBM messo in Open Source dal 2000
 - LSF, Citrix prodotti proprietari
 - Sistemi di calcolo Linux di INFO: 4 cluster di calcolo + vari server
- **Open Source in ENEA**
 - 2003 progetto pilota per l'introduzione prodotti Open Source sul desktop
 - qualche utenza Linux (maggioranza Win e Mac)
 - sistemi di calcolo dipertimentali con Linux

Open Source Software

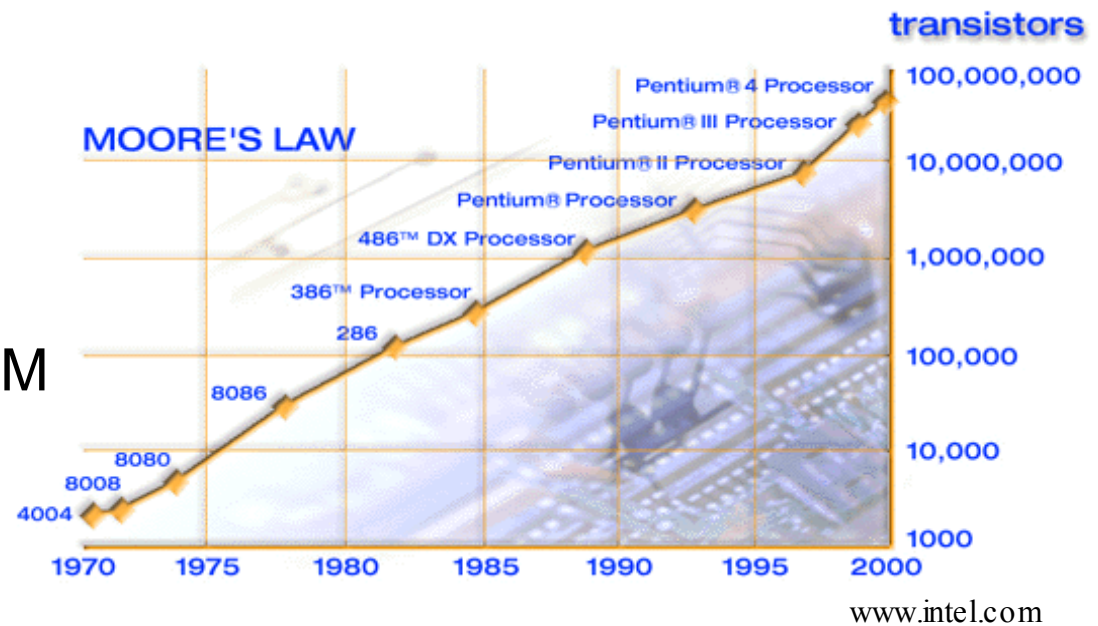
- Un po' di storia
- Che cosa è l'Open Source
- Le licenze
- Gli sviluppatori, le aziende, i progetti
- I governi
- Un esempio: Open Office

Un po' di storia (1)

- La legge di Moore

- 1970 e prima

- Informatica proprietaria con organizzazione “**verticale**” software & hardware, es. IBM
- 1969 in AT&T nasce **UNIX**
- 1971 microprocessore



- 1980

- 1980 **IBM-PC** 16 bit Microsoft, Intel
- Unix si diffonde per i server e nel mondo accademico
- 1984 **Richard Stallman** lancia il Progetto GNU, fonda FSF [Free Software Foundation], definisce la licenza GPL
- 1989 **Internet**: 100.000 sistemi connessi

Un po' di storia (2)

- 1990
 - 1991 Il CERN rilascia WWW: info.cern.ch
 - 1991 **Linus Torvalds**: Linux per PC-Intel, licenza GPL
 - 1995 Le distribuzioni Linux Red Hat, SUSE
 - 1998 **Eric Raymond**: Open Source Initiative, definisce le licenze OSD [Open Source Definition]
- 2000
 - 2000 IBM supporta Linux, anche sui mainframe
 - 2003 Linux Web server OS : 30%
 - 2004 **Apache Web server: 70%!**
 - 2004 supercalcolo: N°3 della classifica www.top500.org, cluster di 1100 nodi Apple, Mac OS X Kernel “Darwin” OSS (Virginia Tech. US)

Richard Stallman, USA 1957

- 1984: **Free Software Foundation**
 - Progetto GNU: (GNU's Not Unix)
 - dimensione sociale e culturale del software libero
- **Software libero => 4 libertà:**
 - di eseguire per qualsiasi scopo
 - di studiare e adattare alle proprie esigenze
 - di redistribuire le copie
 - di migliorare e distribuire la versione migliorata
- **Da copyright a copyleft: licenza GPL**
 - GPL [Generic Public License]
 - modificazione e distribuzione ma conservando la licenza GPL
 - sorgente a disposizione
 - senza garanzie [al più garanzia ma a pagamento]



Roma 2003, C.Pelliccia

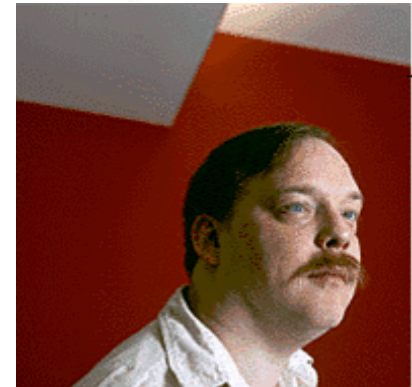
Linus Torvalds, Finlandia 1969

- 1991: **Linux sistema operativo per PC Intel sotto GPL**
 - Kernel + Software GNU : **GNU/Linux**
- **Sistema di sviluppo del kernel:**
 - Linus Torvalds: “Benevolent Dictator”
 - due famiglie di versioni:
 - stabile (pari: 2.2, 2.4, 2.6) sperimentale (dispari)
 - Utilizza BitKeeper (software proprietario!) per la gestione
 - Migliaia di sviluppatori; sempre più piattaforme supportate
- **Le distribuzioni di Linux**
 - 1993: Debian [l'ultima versione (2002), 100 milioni linee di codice!]
 - 1995: prima versione Red Hat, SUSE
- **Diffusione iniziale nel mondo scientifico/accademico e internet**
 - Dal ~1998 anche le grandi aziende: IBM, Sun, Oracle, Apple,.....



Eric Raymond, USA 1954

- 1998: Insieme ad altri lancia OSI: Open Source Initiative
 - Scrive **“The Cathedral and the Bazaar”**
 - Open source come fattore di progresso tecnico
- OSI: Open Source Definition: meno restrittiva di GPL
 - Disponibilità del codice sorgente
 - Redistribuzione del software senza royalties
 - Nessuna discriminazione su uso e utilizzatori del software
 - La condizione copyleft non è obbligatoria (per evitare gli effetti potenzialmente “virali” per il software proprietario)
- OSI esamina le licenze e dichiara il rispetto di OSD
 - Attualmente +40 Licenze registrate, inclusa GPL
 - OSD permette di creare prodotti che mischiano software open source e software proprietari



Che cose e' il software libero/software aperto?

- L'approccio “software libero” vuole garantire:
 - il creatore del software, che mantiene il copyright
 - la disponibilità del codice sorgente importante dal punto di vista socio/culturale o/e come vantaggio tecnologico.
- Il mezzo:
 - una licenza accuratamente definita
- Si vuole evitare
 - La possibilità di appropriazione da parte di altri, possibile nel caso di software “di pubblico dominio” cioè privo di copyright.
- Si basa su una comunità di sviluppatori/utenti distribuita in tutte le aree geografiche: i protagonisti del fenomeno.
 - sourceforge.net (+40.000 progetti, +400.000 users), freshmeat,...
 - forum, mailing list, newsgroup, linux groups (www.lugroma.org)

Alcuni “miti”

- **Il software libero è gratuito**
 - Molto del software libero è gratuito ma ci sono programmatori che vivono dalla vendita e manutenzione dei loro programmi “liberi”
- **Il software gratuito è libero**
 - Molti programmi proprietari vengono distribuiti gratuitamente
- **Il software libero è privo di copyright**
 - Gran parte del software libero è distribuito sotto licenza
- **L'introduzione del software libero abbassa i costi**
 - può essere vero ma dipende, è necessaria un'analisi accurata dei costi complessivi.

[<http://softwarelibero.it/documentazione/softwarelibero.html>, F. Potortì]

Le Licenze

- Molte licenze rispettano la Open Source Definition
 - Grandi progetti BSD, Apache, Perl,..... (su sourceforge 70% GPL)
 - Licenze create dalle grandi aziende: IBM, Sun, Apple,...
 - Progetti ibridi: licenza GPL e versione proprietaria (es. MySQL)
- Chi controlla /sancisce il rispetto delle licenze?
 - Il tribunale della “opinione pubblica” (es. Epson Kowa Corp)
- Problema legale complesso; questione brevettabilità del software:
 - **US:** dal 1980 il software può essere coperto da copyright
 - Enorme crescita del numero di brevetti (+100.000) 20.000/anno
 - **Europa:** 2003 direttiva del parlamento europeo: i dispositivi sono brevettabili, il software invece no.
 - +70% degli sviluppatori OSS sono europei (rap. FLOSS,2002)

Gli sviluppatori

- **Motivazioni**

- **Divertimento**, stimolo intellettuale della creazione del codice: ne segue il desiderio di semplicità, eleganza, modularità
- **Spirito artistico** combinato a sfiducia nelle grandi aziende
- **Risposta ad un bisogno** immediato: il software proprietario spesso è poco configurabile/flessibile, sviluppatore/utente.
- **Sviluppo delle capacità** di programmazione
- Riconoscimento di reputazione tecnica:
 - **Segnale di esistenza verso il mercato**
- Controllo di qualità a basso costo

- **Open Source come fenomeno interessante in se**

- <http://opensource.mit.edu>: +100 lavori di analisi
- sourceforge come “parco” per economisti, sociologi, etnografi.....

Le aziende (1)

- **IBM: licenza OSD “IBM Public License”**
 - 1998 applicazioni proprietarie su Linux, uso di OSS:
 - WebSphere, basato su Apache; porting di Lotus, DB2
 - 2000 porting di Linux ai mainframe S390, supporto linux sui server
 - 2000 rilascio come OSS di OpenAFS, Jibes Java Compiler
 - 2003 IBM ha 9 centri IBM-Linux; 17% server venduti con Linux
- **SUN: licenza OSD “Sun Community Source License”**
 - 1998 Acquista StarOffice, 2000 rilascia **OpenOffice**
 - 2000 rilascia come OSS il proprio sistema operativo Solaris
- **Apple: licenza OSD “Apple Public Source Licence”**
 - 1999 rilascia Darwin, software di base di Mac OS X.
- **ORACLE**
 - 2003: ~1000 sviluppatori dedicati all'ambiente Linux

Le aziende (2)

- Le grandi aziende finanziano sviluppatori e istituzioni no-profit che sviluppano OSS:
 - Open Source Development Labs** [<http://www.osdl.org>] (dal 2003, ci lavora Linus Torvalds)



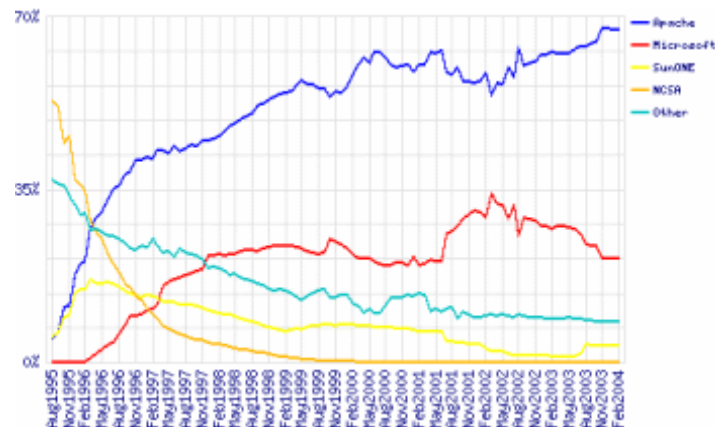
Le aziende (3)



- Come reagisce Microsoft? Considera Linux e OSS una minaccia.
 - 2001 lancia “**shared source program**”, accesso concordato ai sorgenti per ispezione e studio a grandi organizzazioni e governi, senza diritto di modifica; il programma viene esteso al mondo accademico.
 - 2003 **Windows CE viene aperto agli sviluppatori** che possono rilasciare le versioni modificate
- In **ITALIA**: negli ultimi 3 anni OSS sviluppo 120%/anno
 - 2003 Indagine sulle aziende Open Source Software italiane, progetto ELISS (A. Bonaccorsi, C.Rossi,A,Scateni, Scuola Superiore S.Anna, Pisa)
 - 275 aziende, approccio misto (14% solo OSS)

I progetti

- **Apache: Web Server con 70% del mercato**
 - Apache Foundation
 - Comitato direttivo
 - Elezione annuale
 - I membri: contributo significativo
- **GIMP: grafica bitmap**
 - coordinato attraverso GIMP developer mailing list
- **GCC: compilatore C**
 - GCC “Steering Committee” eletto dai partecipanti
- **MySQL: database relazionale**
 - Doppia licenza (il doppio regime richiede il mantenimento copyright: revisione centralizzata)
 - 4.000.000 di utenti, 0.1% paga la licenza proprietaria



www.netcraft.com

I governi (1)

- **Costo del software**
 - US Gov, 2000, 3.7 G\$
 - Italia, Pubblica Amministrazione, 2001, 675 M€
- **Iniziative in molti paesi; alcuni esempi:**
 - **Germania:** Linux sui server della PA, Supporto OSS
 - **Gran Bretagna:** rapporto “Open Source Software” Office of the e-Envoy, UK Government.
 - **Europa:** rapporto “The IDA Open Source Migration Guidelines” [IDA- Interchange of Data between Administrations] Commissione Europea, 2003
- **In Italia:**
 - Legge Reg. Toscana 26/1/2004; utilizzo di standard aperti e scelta preferenziale di soluzioni OSS

I governi (2)

- MIT Ministero per l'Innovazione Tecnologica
 - 2002, Commissione Open Source nella Pubblica Amministrazione
 - 2004, Direttiva MIT: **”Sviluppo dei programmi informatici delle PA”** criteri con cui le PA devono fare comparazioni tecno/economiche per l'acquisizione di software:
 - **“Interoperabilità”** e cooperazione
 - Indipendenza da un unico fornitore/unica tecnologia
 - Disponibilità **codice sorgente**, per ispezione tracciabilità e mantenimento, se il supporto non viene più garantito.
 - Tra i formati di esportazione almeno un **“formato aperto”**
 - **Riuso** delle applicazioni custom (60% del bilancio)
 - Attuazione promossa da CNIPA <http://www.cnipa.gov.it>



Un esempio : OpenOffice (1)



- 2000 SUN rilascia OpenOffice come OSS
- 2003 versione OpenOffice 1.1
 - office suite: oowriter, oocalc, ooimpress, oodraw
- il formato di openoffice: documento.sxi
 - file zip : formato compresso “aperto” (www.info-zip.org)
 - se si espande il file compresso **che contiene?**
 - documenti testo:
 - mimetype
 - content.xml; meta.xml; settings.xml; styles.xml **<= XML**
 - cartelle:
 - meta-inf; Pictures **<= tutte le immagini nel formato originale!**
- **Formato aperto: ==> INTEROPERABILITA' !**

Un esempio : OpenOffice (2)

- Piccolo confronto con MS Office:



- MS Word utilizza per i documenti un formato proprietario “chiuso” ma usa anche **RTF** (Rich Text Format), formato “aperto” ed quindi interoperabile
- Dimensione (byte) di un semplice documento [**Buongiorno!**]
 - MS Word 2003 19968 RTF **3279**
 - OOwrite 5542 RTF **1846**
- Aggiungendo una figura semplice, GIF 1235 byte (800x600 pix)
 - MS Word 2003 24068 RTF **126864**
 - OOwrite 7247 RTF **1988+1235**
- MS Word include nel documento RTF una copia bitmap della figura che ne aumenta la dimensione, senza evidenti vantaggi per l'utente.
- Con Office 2003 MS offre un formato XML per alcuni dei componenti.
 - struttura aperta ma è stato richiesto il brevetto...

Interoperabilità

- Definizione
- Evoluzione della rete:
 - dal Web al Grid Computing e al Semantic Web
- XML, lo strumento base
- Web Services

Interoperabilità

- **COSA E'**
 - la capacità di un sistema o di un prodotto di cooperare insieme ad altri sistemi o prodotti senza richiedere sforzi speciali da parte dell'utilizzatore.
- **PERCHE'**
 - facilitare al meglio lo scambio e il riutilizzo delle informazioni e insieme garantire un appropriato livello di controllo del loro flusso.
- **punti cruciali:**
 - **formato**: modo in i cui dati vengono rappresentati
 - aperto/proprietario
 - **protocollo**: modo in cui i dati vengono trasmessi
 - aperto/proprietario

PRIMA DELLA RETE

Centri di calcolo e reti interne
alle grandi aziende/istituzioni,
isolate tra loro.

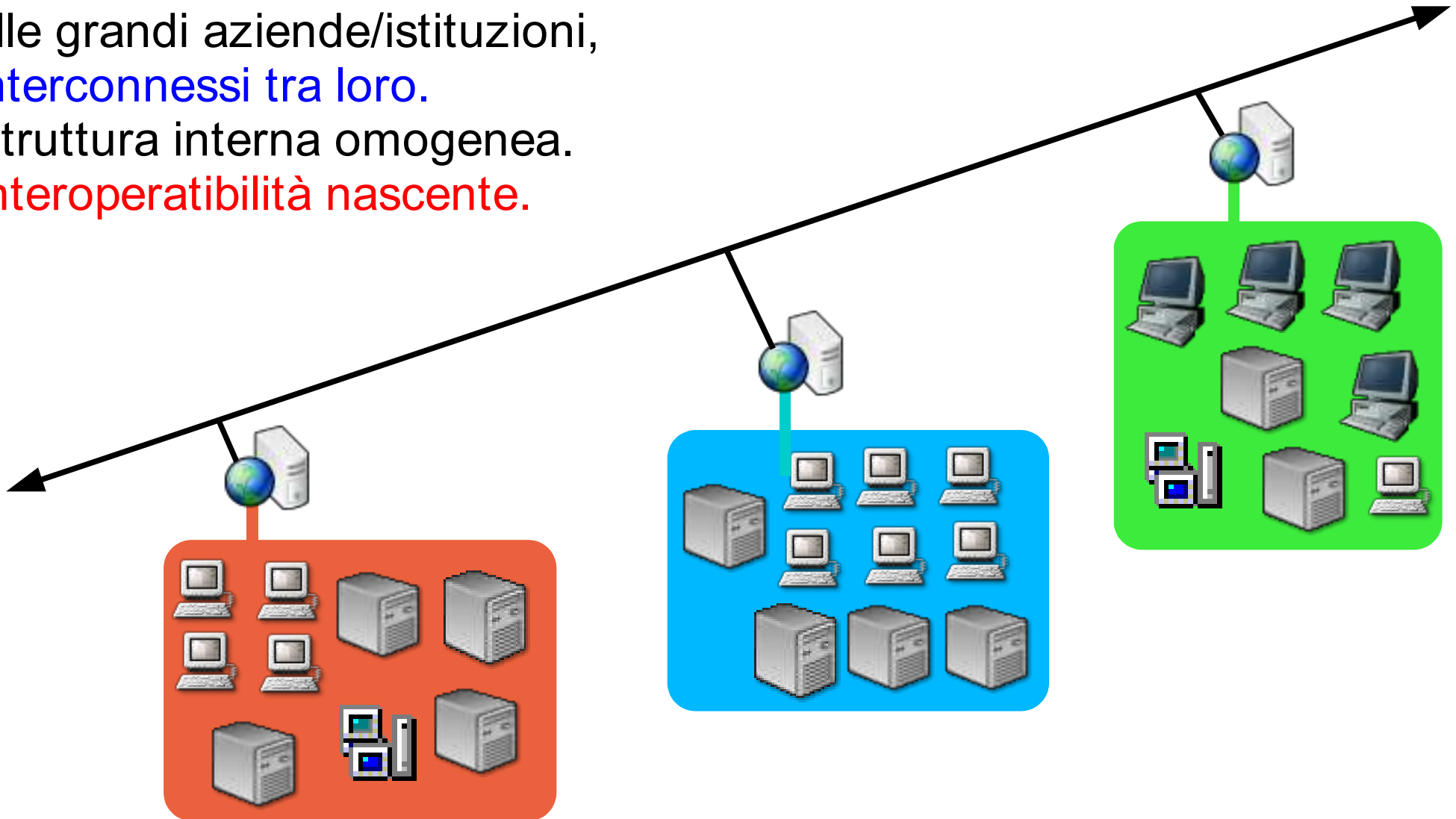
Struttura interna omogenea.

Interoperabilità?



LA RETE : Internet

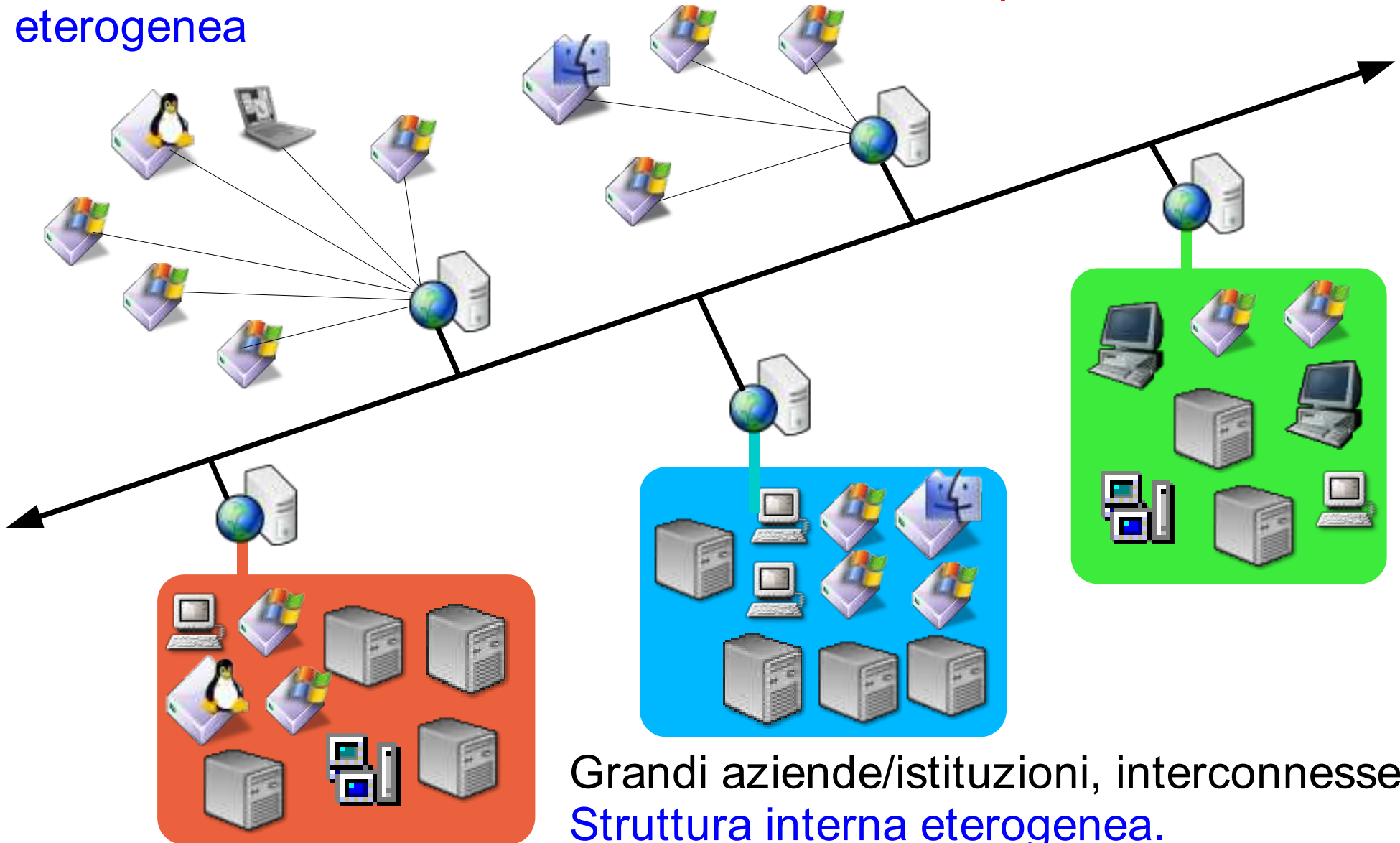
Centri di calcolo e reti interne
alle grandi aziende/istituzioni,
interconnessi tra loro.
Struttura interna omogenea.
Interoperabilità nascente.



LA RETE : Web

Informatica diffusa
ed eterogenea

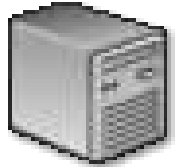
Interoperabilità necessaria



Grandi aziende/istituzioni, interconnesse.
Struttura interna eterogenea.

Internet: i servizi e i problemi della interoperabilità

- Un server offre alla rete i suoi servizi:



- **indirizzo ip**: es. www.enea.it
- servizi: **protocollo e port**
 - WWW: port **80** (es. [apache web server](#))
 - E-mail: port **25** spedizione; port **110** ricezione;
 - FTP: port **21**
 -
- La sicurezza e la “convivenza civile” in una rete **interoperabile**
 - I problemi: Crackers, Virus, Worms, SPAM,.....
 - I rimedi:
 - Il firewall concede l'accesso solo a **port e protocolli prescelti**
 - Metodi di autenticazione, certificati, cifratura,.....

WWW: i punti di forza

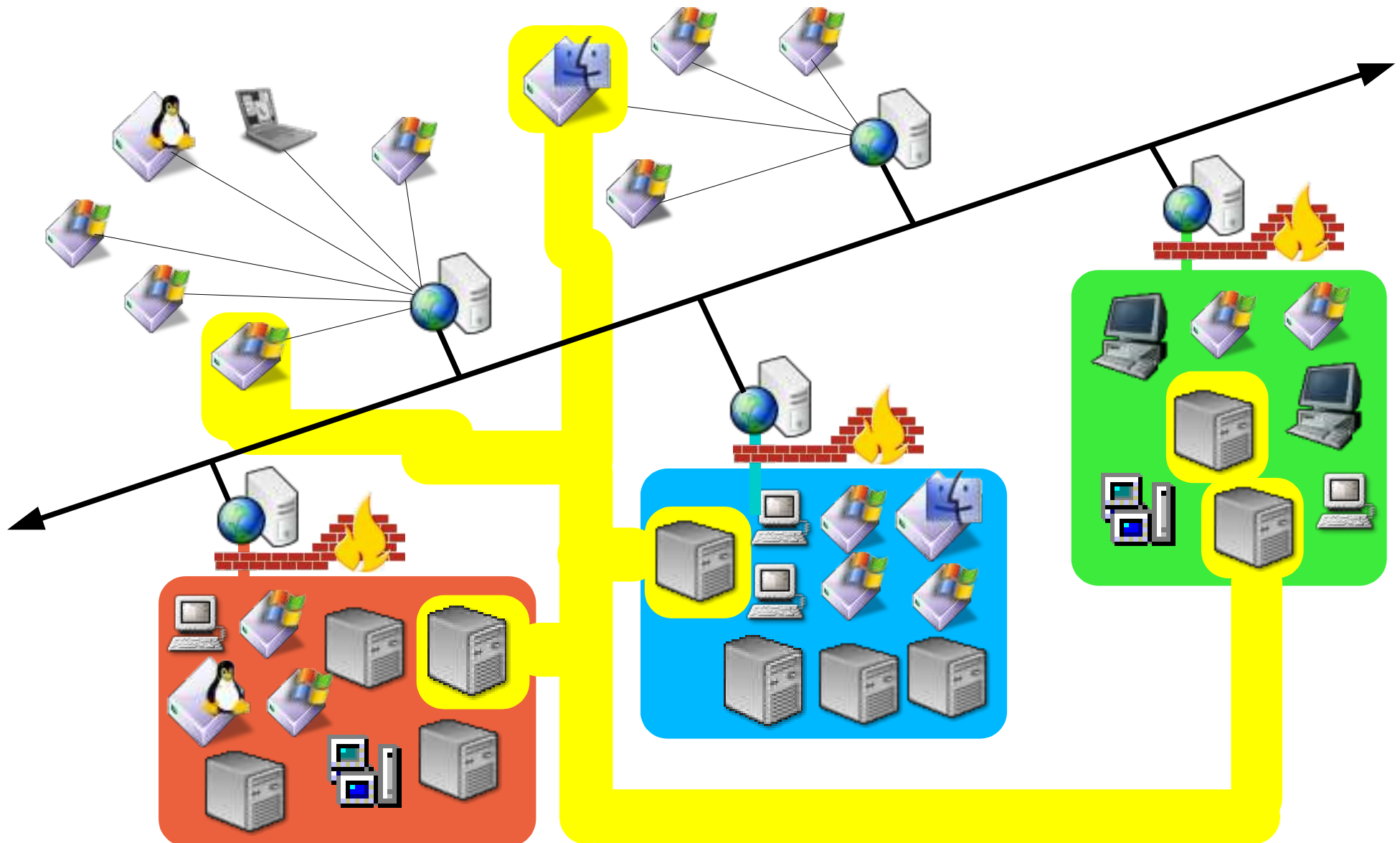
- **Grande successo nell'accesso alle informazioni sulla rete**
- **Semplice gestione**
 - utente: browser gratuito, magari opensource [Mozilla]
 - gestore: installa web server, magari OSS Apache
 - rete: traffico sulla porta 80
- **Elementi base semplici e “aperti”**
 - documenti in **formato testo**
 - linguaggio **HTML**
 - protocollo: **http**, localizzazione: **URL**
 - metodi di accesso alle risorse del server **GET, POST**
- **Multipiattaforma, formati e protocolli aperti**
- **Disponibilità di motori di ricerca: Google** [15.000 server linux!]

HTML; i limiti di WWW e le estensioni

- HTML un linguaggio semplice
 - `<html>`
 - `<head><title>Prova</title></head>`
 - `<body>Buongiorno!</body>`
 - `</html>`
- Limite di WWW: nato per trattare dati testuali
- estensioni per aggiungere funzionalità
 - **server**: php, asp, servlet java,.....
 - **client**: applet java, activeX, plug-ins,.....
- **non tutte le estensioni hanno la stessa interoperabilità**
 - es. java: SI! activeX: NO!
- rischio di smarrire la visione unitaria dello strumento

LA RETE : Grid Computing/ Semantic Web

Unità globale della rete, richiede **interoperabilità "profonda"**



Grid computing/Semantic Web

- una visione unitaria dell'insieme delle risorse e dei contenuti del Web (**massima interoperabilità**)
 - Peer-to-Peer: la condivisione distribuita di qualsiasi informazione sulla rete ma anche la rete come giungla.
 - **Grid computing**: la rete come riserva di potenza di calcolo per collaborazioni scientifico/tecniche su vasta area
 - Ian Foster, C.Kesselman et al. 1995 “Anatomy of the grid”
 - **Semantic Web**: aggiungere ai contenuti del web anche un modo formale per descriverli e di descrivere anche i metodi per poter accedere ed individuare a tali contenuti
 - Tim Berners-Lee, 2001 “Semantic Web”
 - **Una integrazione dei due: Open Grid Service Architecture**
 - Ian Foster, C.Kesselman et al. 2002 “Physiology of the grid”

Semantic Web (1)

- lo strumento per il massimo di Interoperabilità
- il metalinguaggio XML: eXtensible mark-up language
- HTML
 - `<html>`
 - `<head><title>Prova</title></head>`
 - `<body>Buongiorno!</body>`
 - `</html>`
- XML
 - `<ricetta> <nome>Pane</nome>`
 - `<ingrediente>farina</ingrediente> <ingrediente>acqua</ingrediente>`
 - `<ingrediente>sale</ingrediente><ingrediente>lievito</ingrediente>`
 - `<temperatura>220</temperatura><tempo>10</tempo>`
 - `...</ricetta>`

Semantic Web (2)

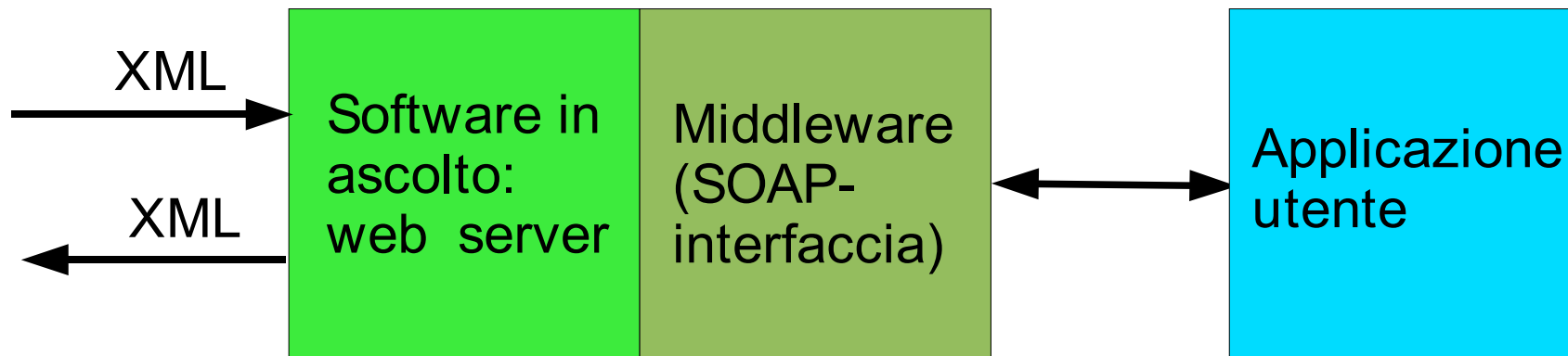
- Utilizzando XML come linguaggio di base si costruiscono gli altri componenti
 - **RDF**: Resource Description Framework, **descrive formalmente i dati con un soggetto, verbo ,complemento**
 - (il progetto)(possiede)(una home page)
 - **URI**: Uniform Resource Identifier, una estensione del concetto di **localizzazione** di URL che comprende anche i “verbi”
 -
- Se le risorse sulla rete sono descritte in modo interoperabile diventa possibile e fruttuoso il ruolo “**agenti**” che **esplorano la rete in modo automatico alla ricerca delle risorse (dati e servizi) disponibili**

Web Service (1)

- Cosa e'
 - “Un nuovo tipo di applicazione Web, che fornisce non solo un servizio ma anche la sua descrizione; in tal modo si mette a disposizione anche degli altri web services sulla rete”
- Dove risiede?
 - server web, ad es. Apache
- Come viene interrogato?
 - attraverso comandi forniti in XML , SOAP , Simple Object Access Protocol.
- Come ne vengono definite le proprietà?
 - in XML con WSDL , Web Service Definition Language
- Come puo' venire scoperto da altri web services?
 - utilizzando UDDI, Universal Description, Discovery and Integration Service), anche questo basato su XML

Web Service (2)

- Altri componenti basati su XML per gestire:
 - transazioni, autenticazione PKI, cifratura, file systems,....
- Schema di funzionamento:



Web Service (3)

- Implementazioni
 - Microsoft: .Net
 - Sun: Sun One Platform
 - IBM: WebSphere
 - Globus Tool Kit 3
- Limiti attuali
 - prestazioni: file di testo, concentrazione sul Web server
 - compatibilità tra piattaforma
- Garanzia di standard comuni
 - W3C: supporto da ~400 organizzazioni, HTML,
 - OASIS: “Organization for the advancement of Structured Information Standards” ~600 organizzazioni, e-business standards

Conclusioni

- **Open Source: fenomeno tecnologico, socio-economico, culturale**
 - la conoscenza del dettaglio tecnologico può essere una **garanzia per il controllo dell'uso delle tecnologie**.
 - può favorire lo sviluppo di **soluzioni tecnicamente efficienti** e stimolare la **concorrenza** su tale piano.
 - può favorire l'affermarsi di **formati aperti** ==> interoperabilità
- **Interoperabilità ==> la rete globale**
 - impossibile valutarne ad oggi le ricadute culturali oltre che tecnologiche e socio-economiche.
 - può portare ad un **nuovo livello di integrazione e di diffusione** lo sfruttamento delle risorse disponibili sulla rete
 - deve contenere caratteristiche di **flessibilità** e di **concorrenzialità** interna che evitino i rischi di **appiattimento** ed **omologazione** senza finire preda di fenomeni di degenerazione.