

Computer in rete e "grandi dati" Impatto su ricerca e formazione

Mauro Campanella
mauro.campanella@garr.it

Simposio "Scienza e Tecnologia, un dialogo che continua"
LNGS, 22 ottobre 2015

Conclusioni (tesi)

Le tecnologie dell'informazione e comunicazione (ed Internet), nate dalla ricerca circa 60 anni fa, continueranno a svilupparsi ed a essere sempre più diffuse, molto rapidamente.

Stanno introducendo dei cambiamenti strutturali nella vita e nel lavoro.

La scienza sta sfruttando le nuove tecnologie di calcolo e comunicazione per ricerche prima non possibili.

Siamo di fronte ad un diluvio di informazioni (dati) e di tecnologia in tutti i campi che richiede nuovi paradigmi culturali, morali e sociali per il loro uso.

L'istruzione ha il compito più complesso per l'introduzione delle tecnologie informatiche e la modifica della didattica, ma mantenendo il prezioso rapporto umano.

Svolgimento della presentazione

Identificare concetti e tecnologie alla base della costante evoluzione del mondo in mondo digitale (senza scendere in dettagli troppo tecnici) ed il loro sviluppo partendo 'dal basso'.

Mostrare esempi di come ricerca e società stiano già utilizzando le nuove tecnologie (e le sfide che si creano)

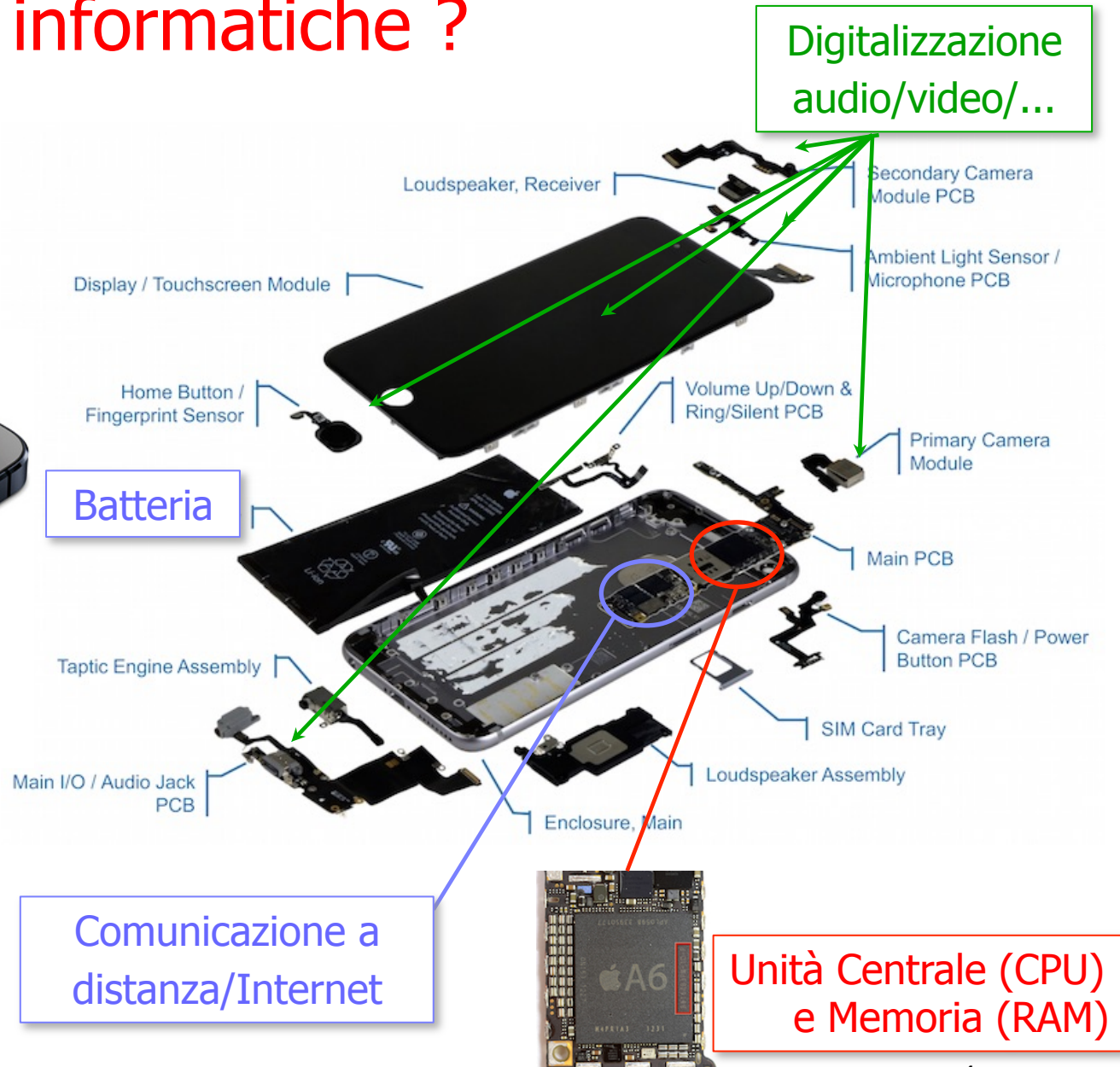


Componenti fondamentali delle tecnologie informatiche ?

Uno strumento che tutti usiamo:



Quali componenti essenziali ha?



Le componenti fondamentali

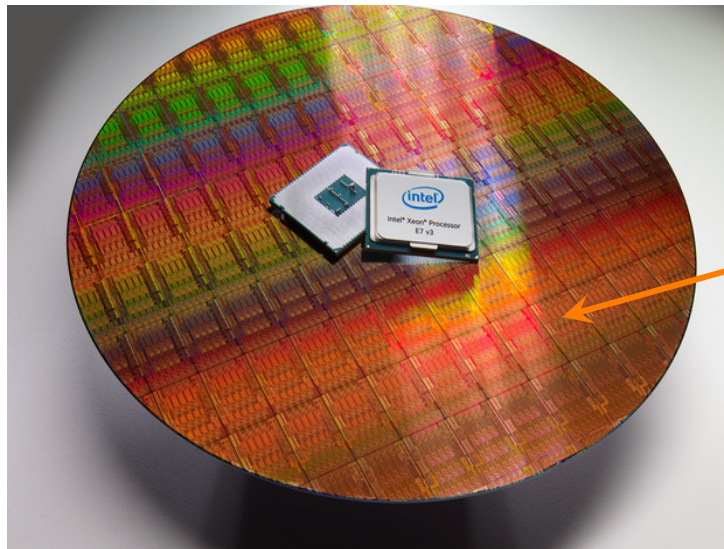
- L'unita' di calcolo centrale (CPU) e la memoria
- La trasmissione dati (ed Internet)
- La digitalizzazione
- Sorgente di energia (*non sarà trattato*)
- **Software** (*non sarà trattato, ma sta diventando "la tecnologia" con cui gestire ogni componente connesso*)

CPU e memoria di massa

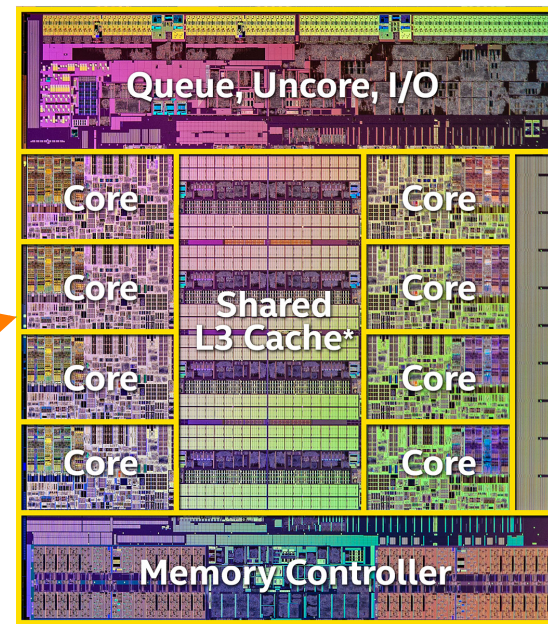
L'unità Centrale di elaborazione (CPU)

Nata dall'idea di usare silicio su cui disegnare elementi base di circuiti elettronici, realizzando circuiti integrati, che sono così 'impacchettati' e 'rimpiccioliti' aumentandone la densità e la velocità (piste sui 15/20 nanometri).

Federico Faggin disegnò il primo Intel 4004 , si occupa di effettuare operazioni sui dati digitali e di controllare e coordinare il funzionamento di un 'oggetto'.



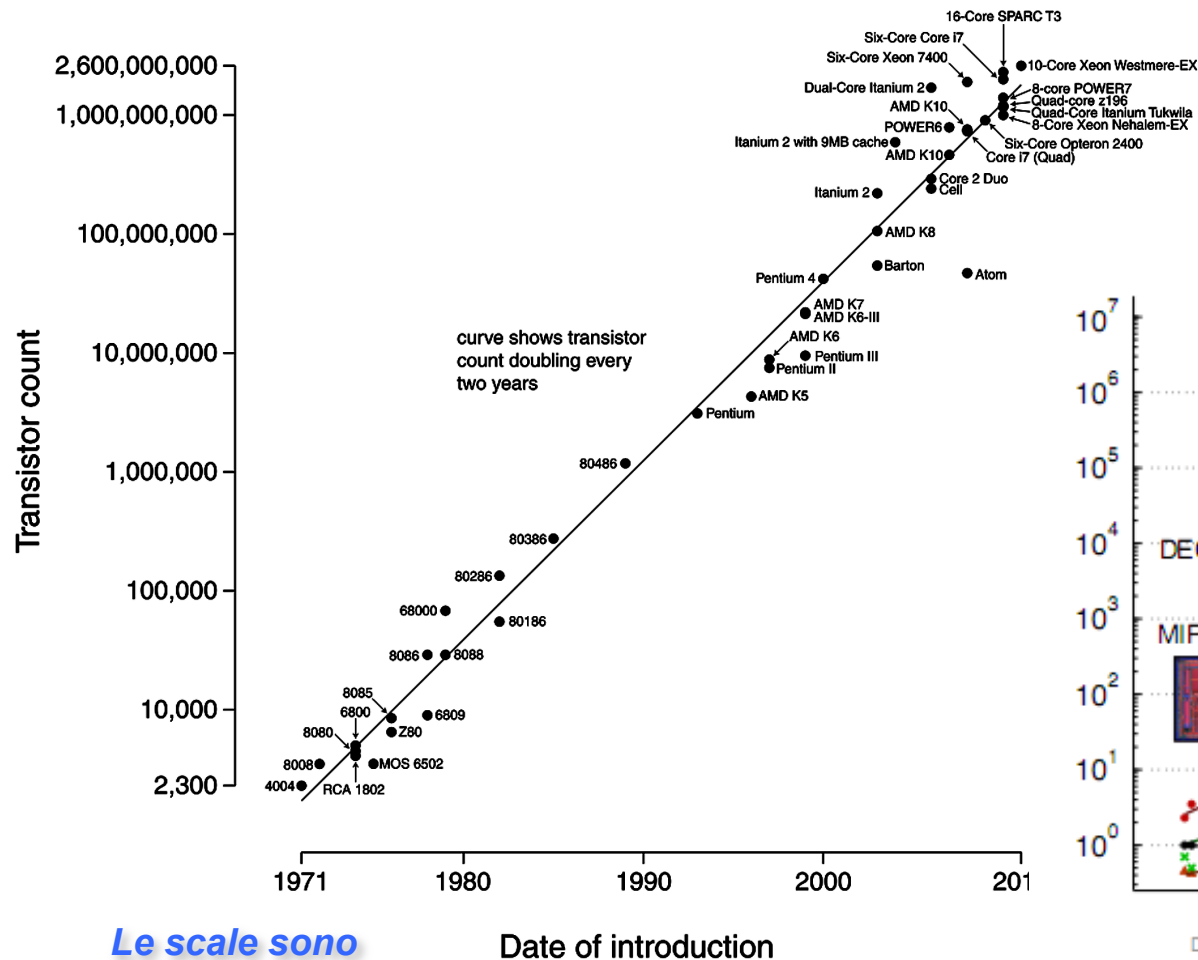
Wafer di Silicio con CPU



Intel Haswell E CPU

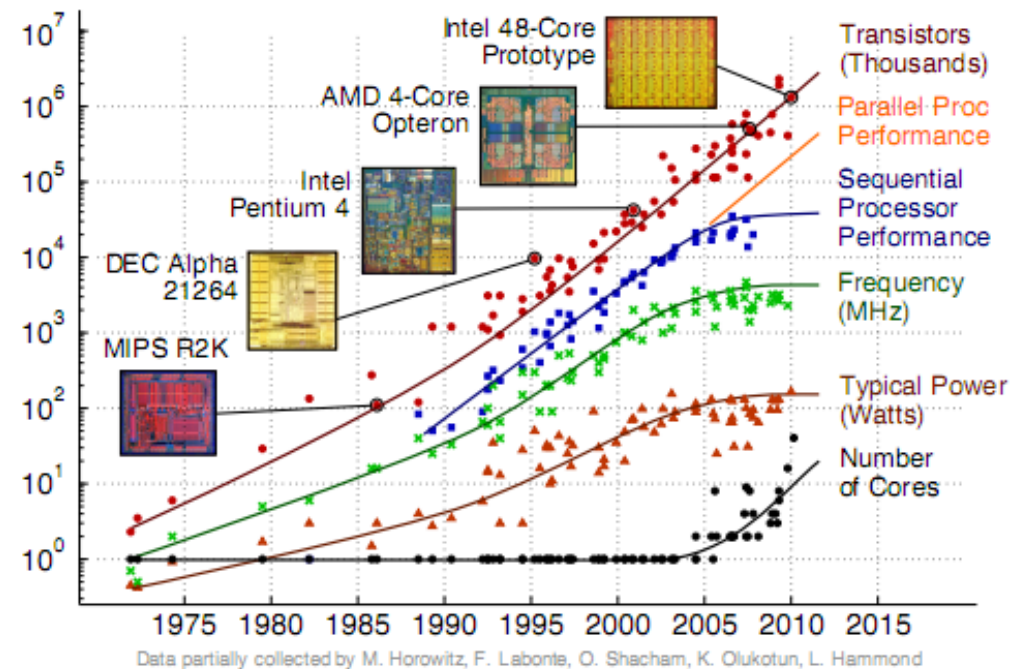
L'evoluzione delle CPU ("Legge di Moore")

Microprocessor Transistor Counts 1971-2011 & Moore's Law

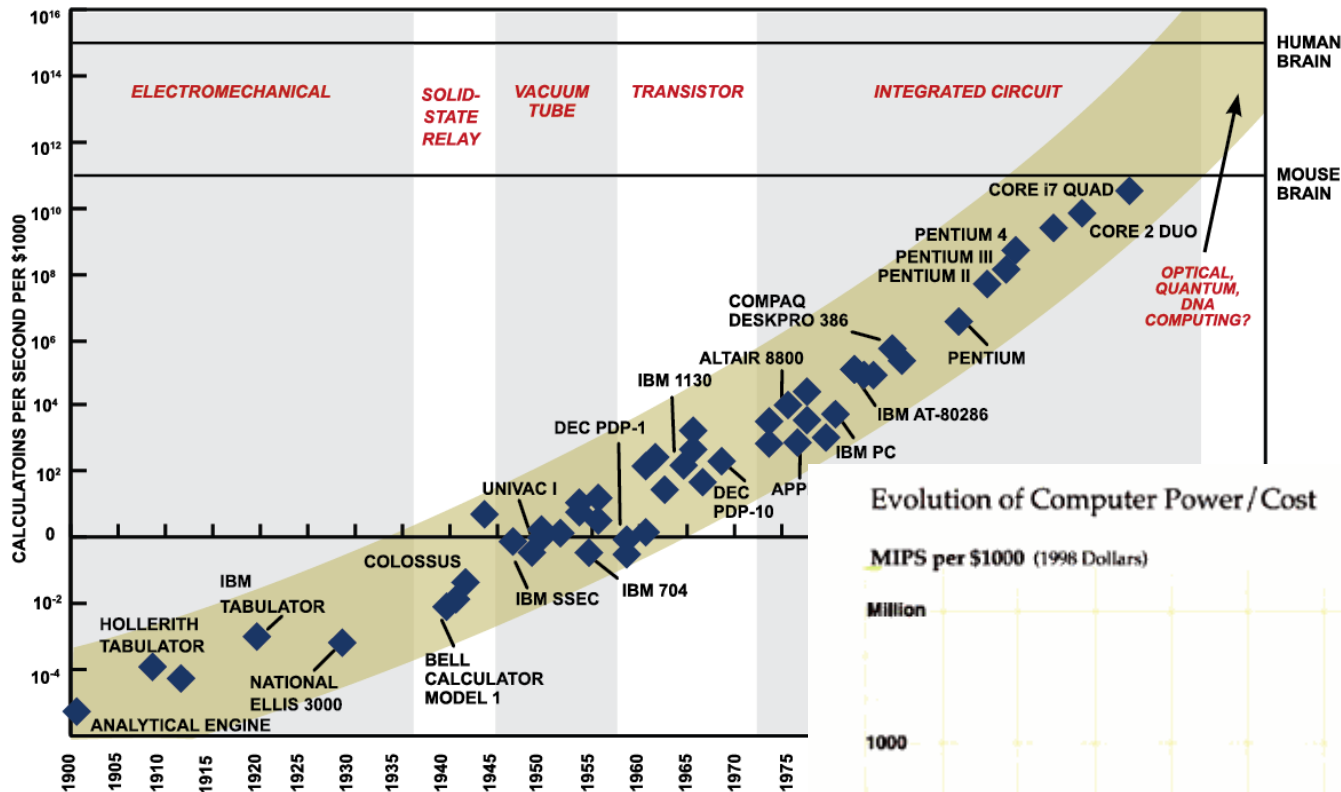


*Le scale sono
logaritmiche ...*

Uno dei fondatori di Intel (Moore) affermò nel 1971 che nei successivi 10 anni la densità dei circuiti sarebbe raddoppiata ogni 18 mesi. Questa previsione è stata vera fino ad oggi ...



Prepared by C. Batten - School of Electrical and Computer Engineering - Cornell University - 2005 - retrieved Dec 12 2012 - <http://www.csl.cornell.edu/courses/ece5950/handouts/ece5950-overview.pdf>



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, THE 2012 REPRESENT BCA ESTIMATES.

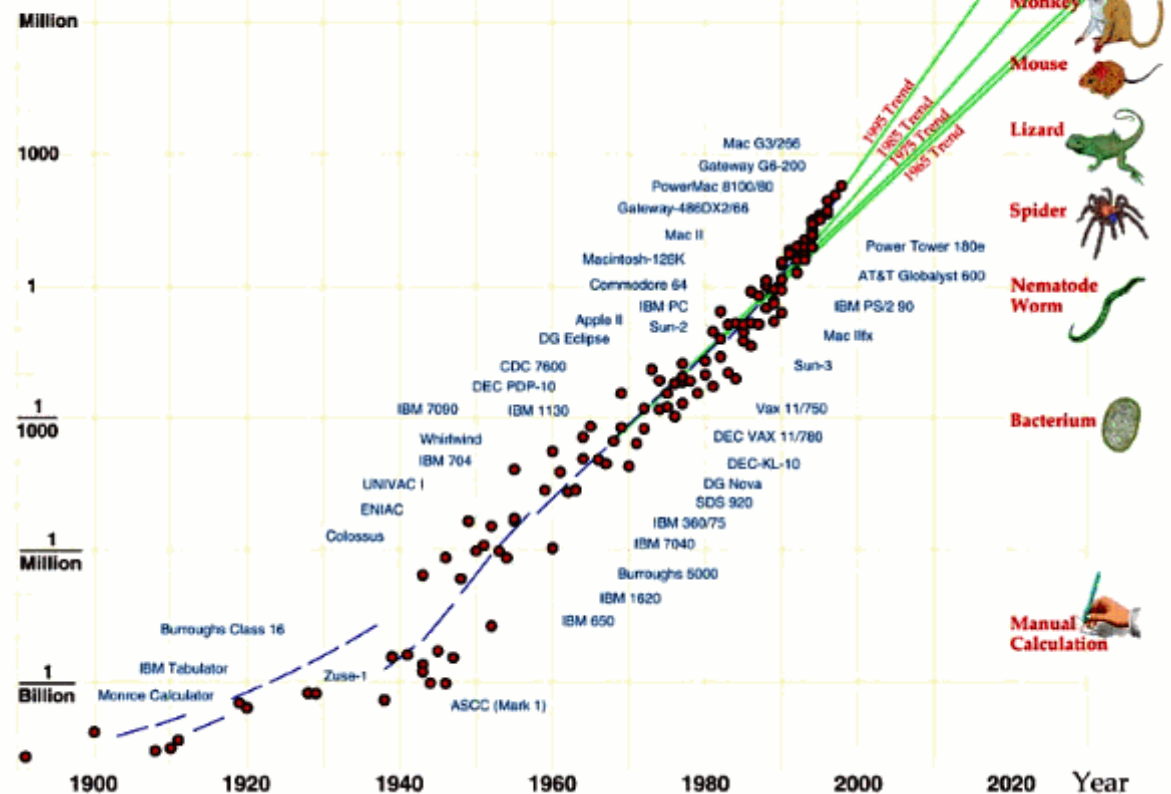
MA

il cervello è un oggetto sia digitale che analogico e non sono confrontabili in base al numero di calcoli al secondo che possono eseguire

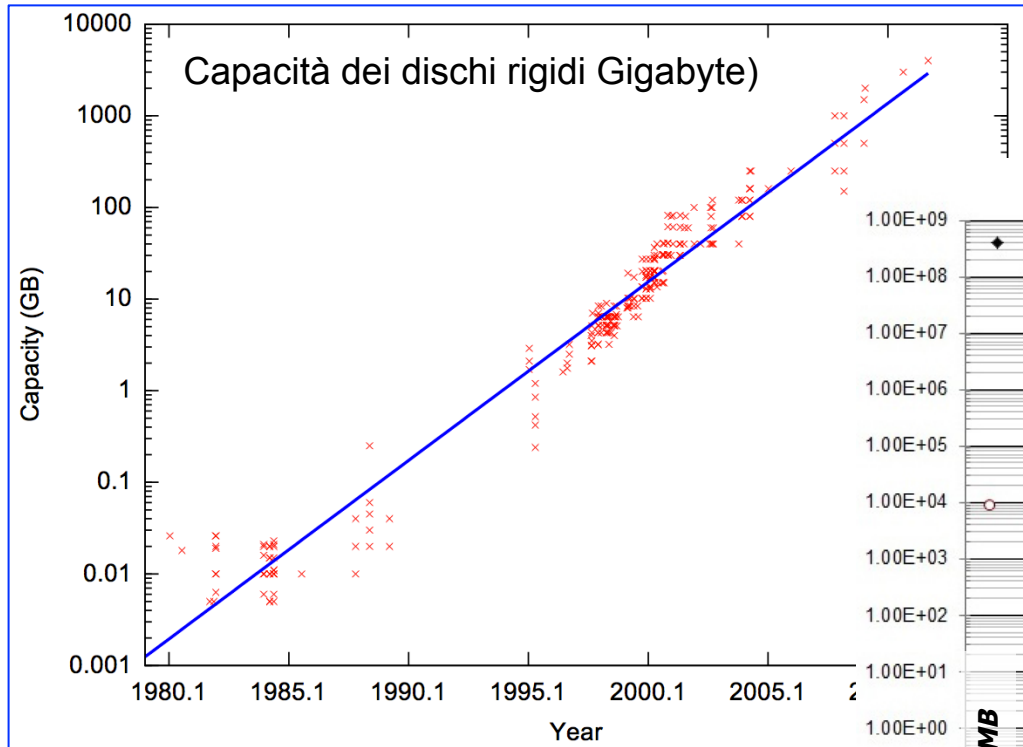
E-voluzione ?

Evolution of Computer Power / Cost

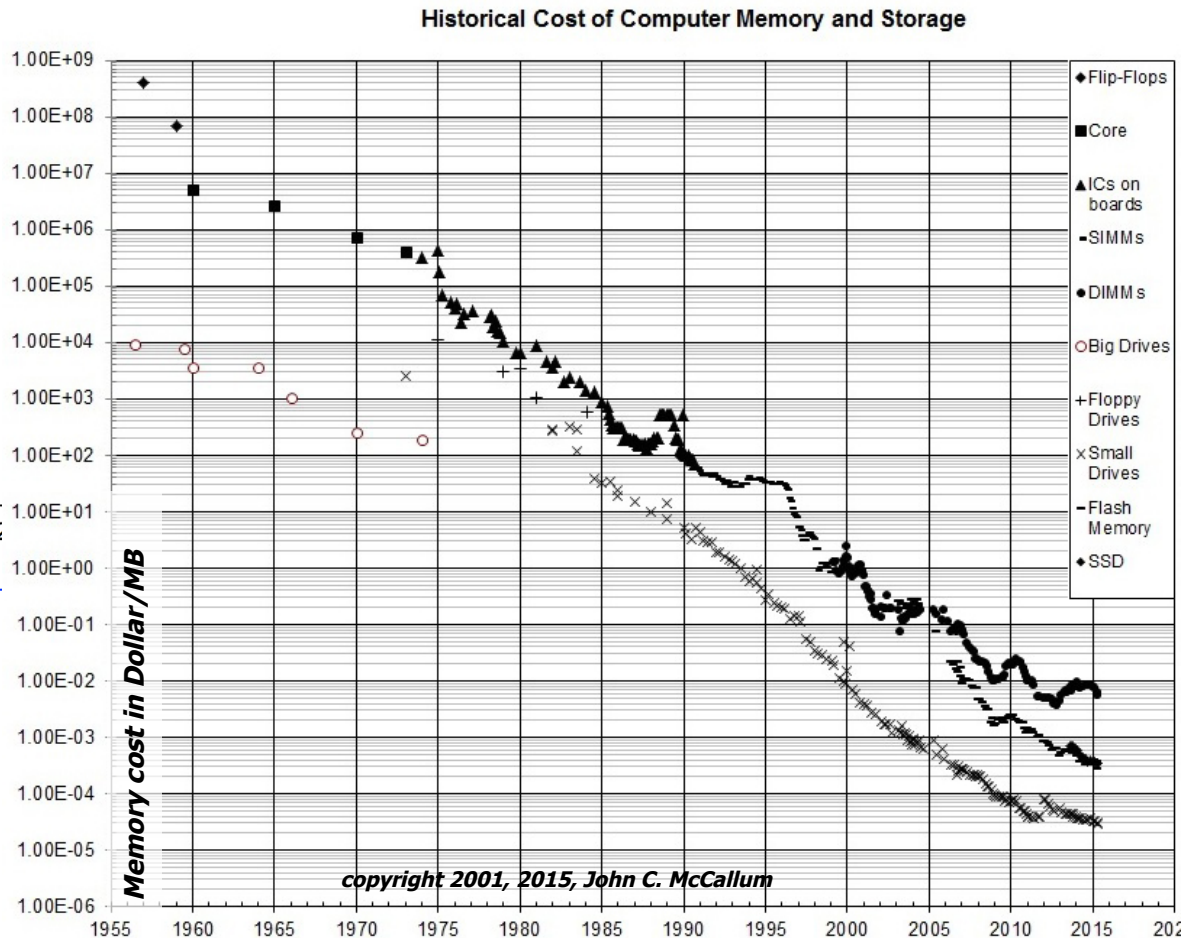
MIPS per \$1000 (1998 Dollars)



Dischi di memoria di massa



Il costo per byte della memoria e' destinato a scendere ancora e i dischi a diventare sempre più veloci passando a stato solido



Il costo di un bit in memoria di massa ?

Un disco esterno base da 2000 GigaByte (2000 miliardi di byte – 1 byte=8 bit e contiene un carattere) costa 60 Euro (2015), quindi un **1 byte costa ~3 miliardesimi (10^{-9}) di centesimo di euro**. Se il disco è usato per N anni, il costo va diviso per N.

(fate tante copie dei vostri dati personali)

Wikipedia (10 volte in dimensione l'enciclopedia britannica) è composta da circa 10.000 milioni di caratteri (tralasciando le immagini) , cioè **10 GB**. Che equivalgono 30 centesimi (solo testo).

Un libro di testo di 500 pagine e varie immagini è pari a qualche centinaio di MB e quindi ha un **costo di memorizzazione** (in byte su disco) **inferiore a 3 centesimi**. Il **solo testo**, senza figure, ha un costo di memorizzazione su disco di **0,007 centesimi**.

Internet

Il valore dell'informazione digitale condivisa: Internet

"Internet" è estesa a livello mondiale ed é cresciuta varie centinaia di milioni di volte dalla creazione e sta ancora crescendo e modificandosi.

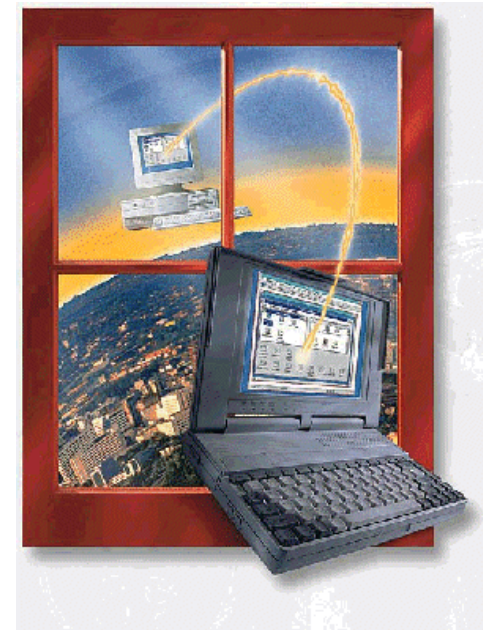
E' La rete principale per il trasporto dell' informazione.

E' pubblica, non proprietaria, neutrale ed aperta a tutti⁽¹⁾ .

Tutti possono partecipare alla sua crescita ed innovazione.

E' basata su pochi principi semplici e chiari.

Ha creato un modello collaborativo di creazione ed uso dell'informazione.



(1) vi è un forte controllo in Cina ed altri paesi che filtrano l'informazione ottenibile

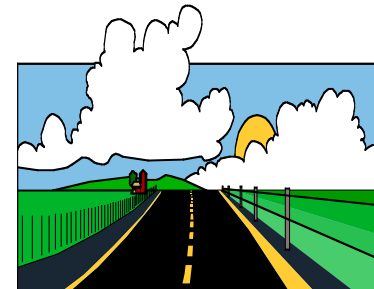
Cosa compone “Internet”

Modello a strati:

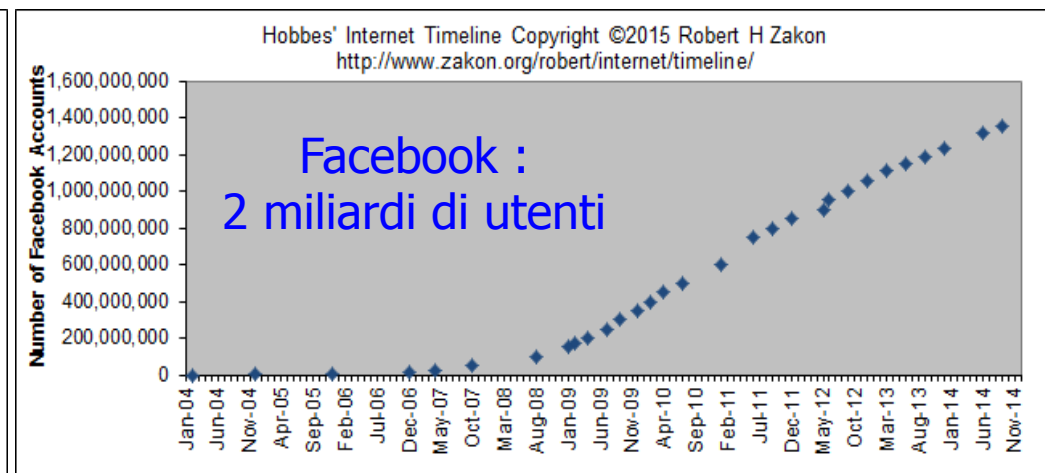
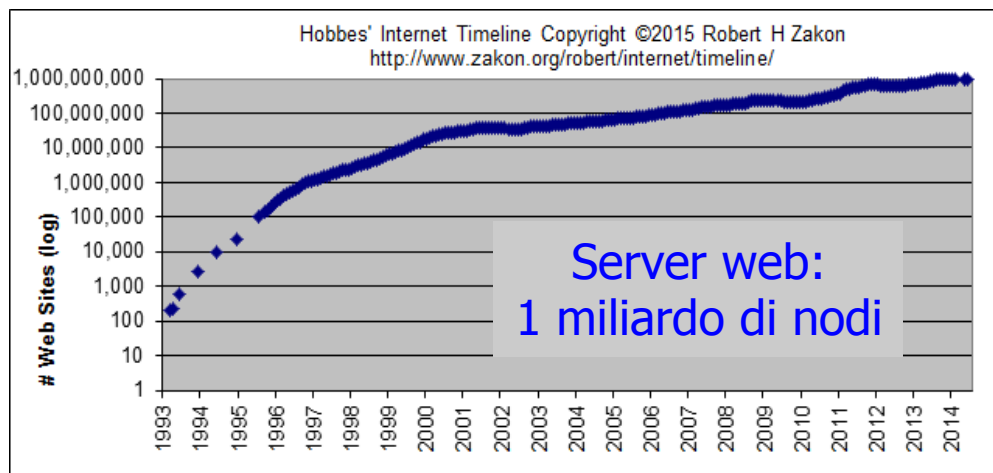
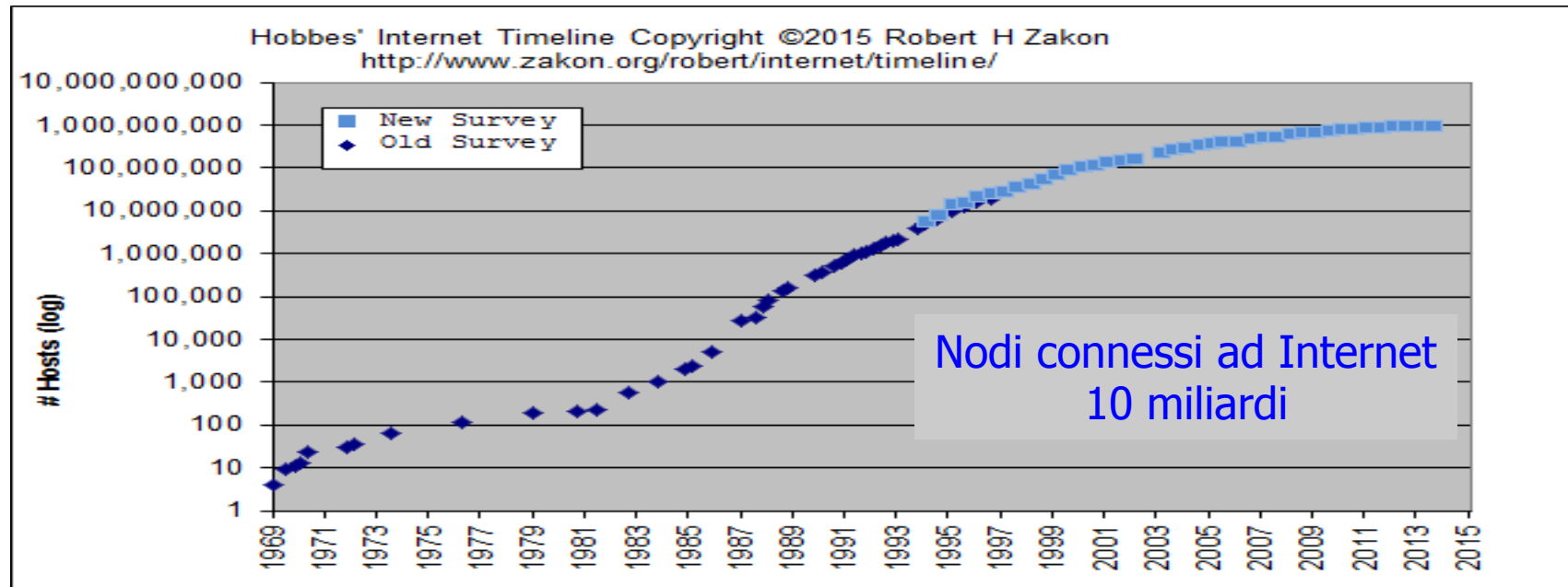
- 4 – applicazioni ed i programmi (posta, browser web, Youtube, facebook, whatsapp,...)
- 3 – I calcolatori
- 2 – La lingua comune per trasmettere dati (protocolli di cui i fondamentali sono TCP ed IP)
- 1 – Tante reti fisiche di trasmissione dati (domini), come una rete autostradale fra paesi diversi gestita in modo cooperativo, generalmente in fibra ottica

Il successo di Internet è dovuto a vari fattori, tra cui:

- Informazione spedita a pacchetti
- Modello decentralizzato con intelligenza ai bordi delle rete.
- Gestione trasmissione fatta dai nodi utente che accedono alla rete e non dalla rete stessa (come nella telefonia)



La dimensione di Internet come sistemi



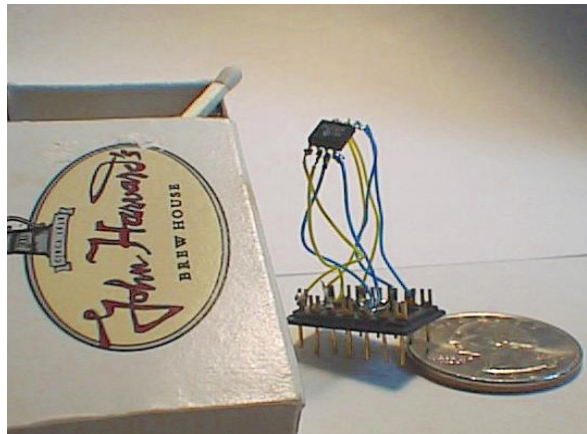
“Grandi” Internet appliances



Tostapane abilitato al
Web + barometro

Il Più piccolo server
WWW del mondo

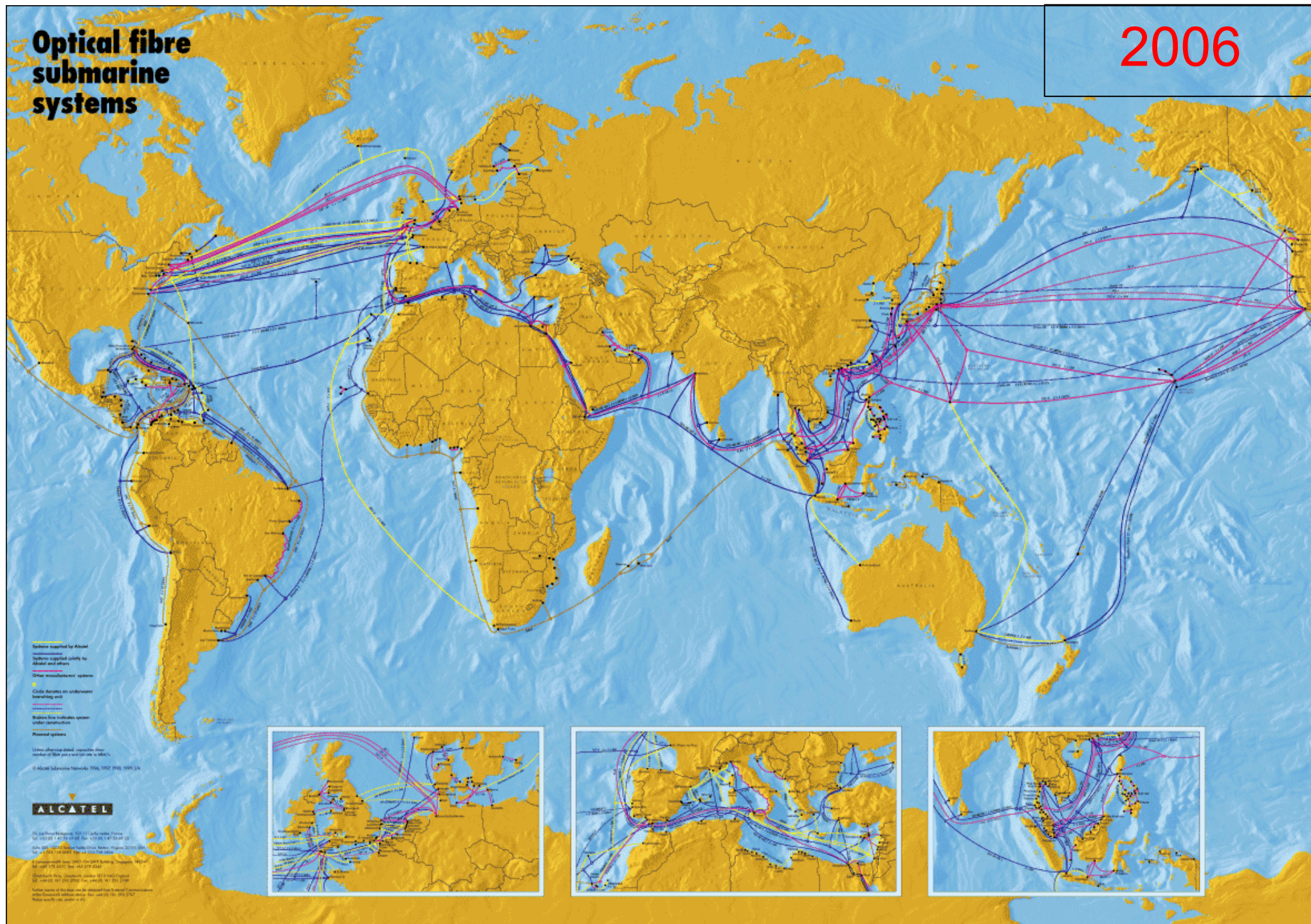
<http://www-ccs.cs.umass.edu/~shri/iPic.html>



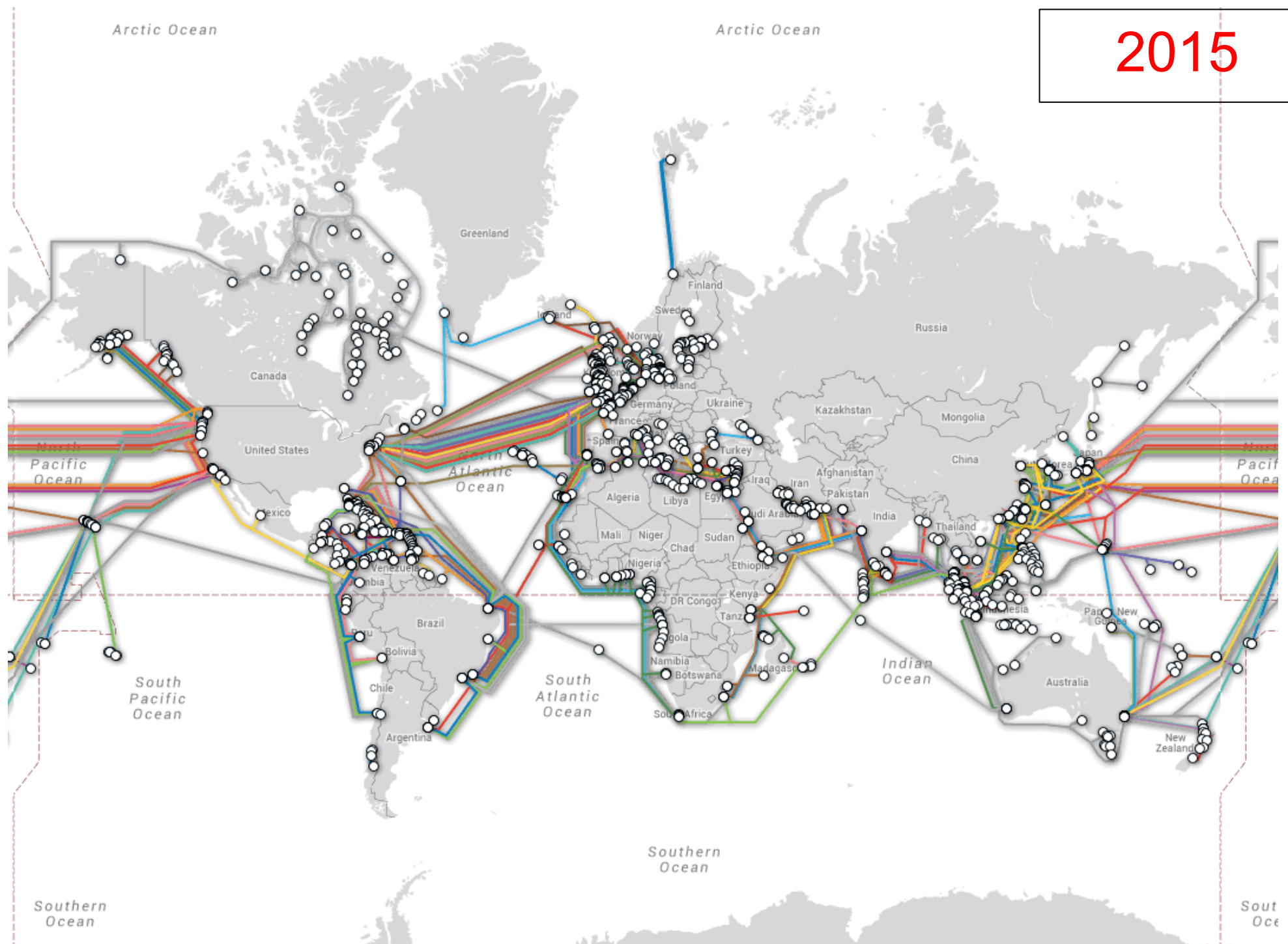
Apple Watch

Optical fibre submarine systems

2006



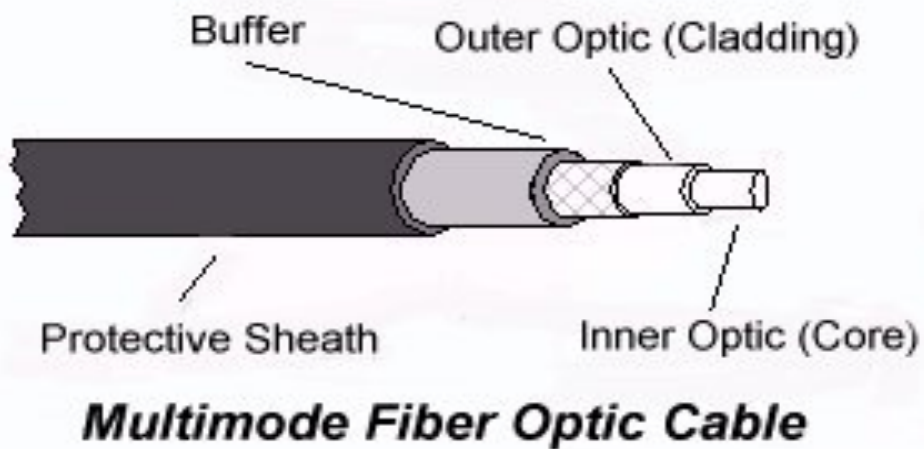
2015



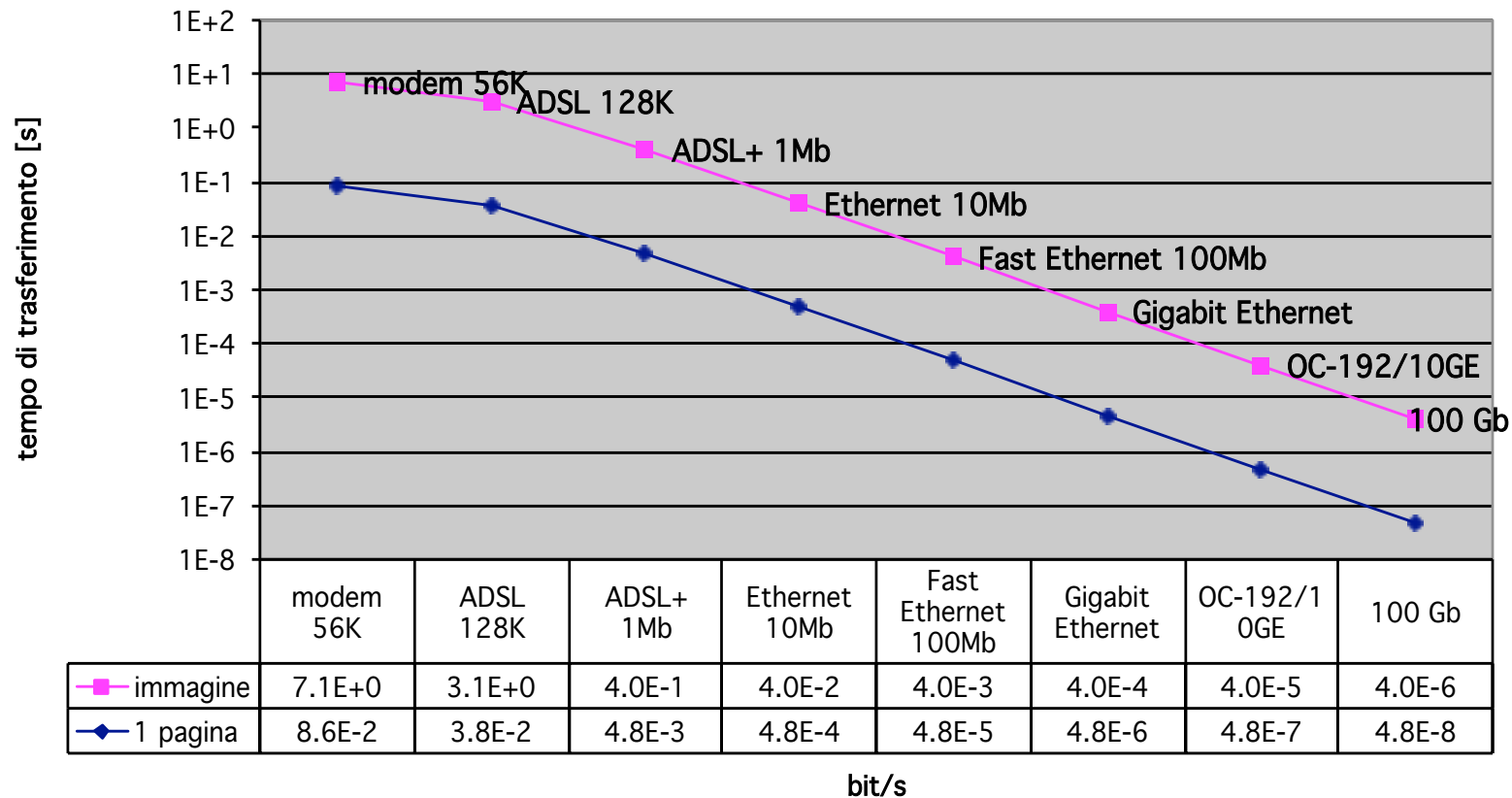
La fibra ottica

Costi molto bassi (composta da silicio)

Alta capacità di trasportare informazione fino a migliaia di miliardi di bit al secondo per ogni fibra.



La velocità trasmissiva



Solo una rete simmetrica in velocità fra trasmissione e ricezione permette la condivisione di informazione creata dall'utente.

Ogni tecnologia asimmetrica (ADSL per esempio) penalizza l'utilizzatore in rispetto alla rete.

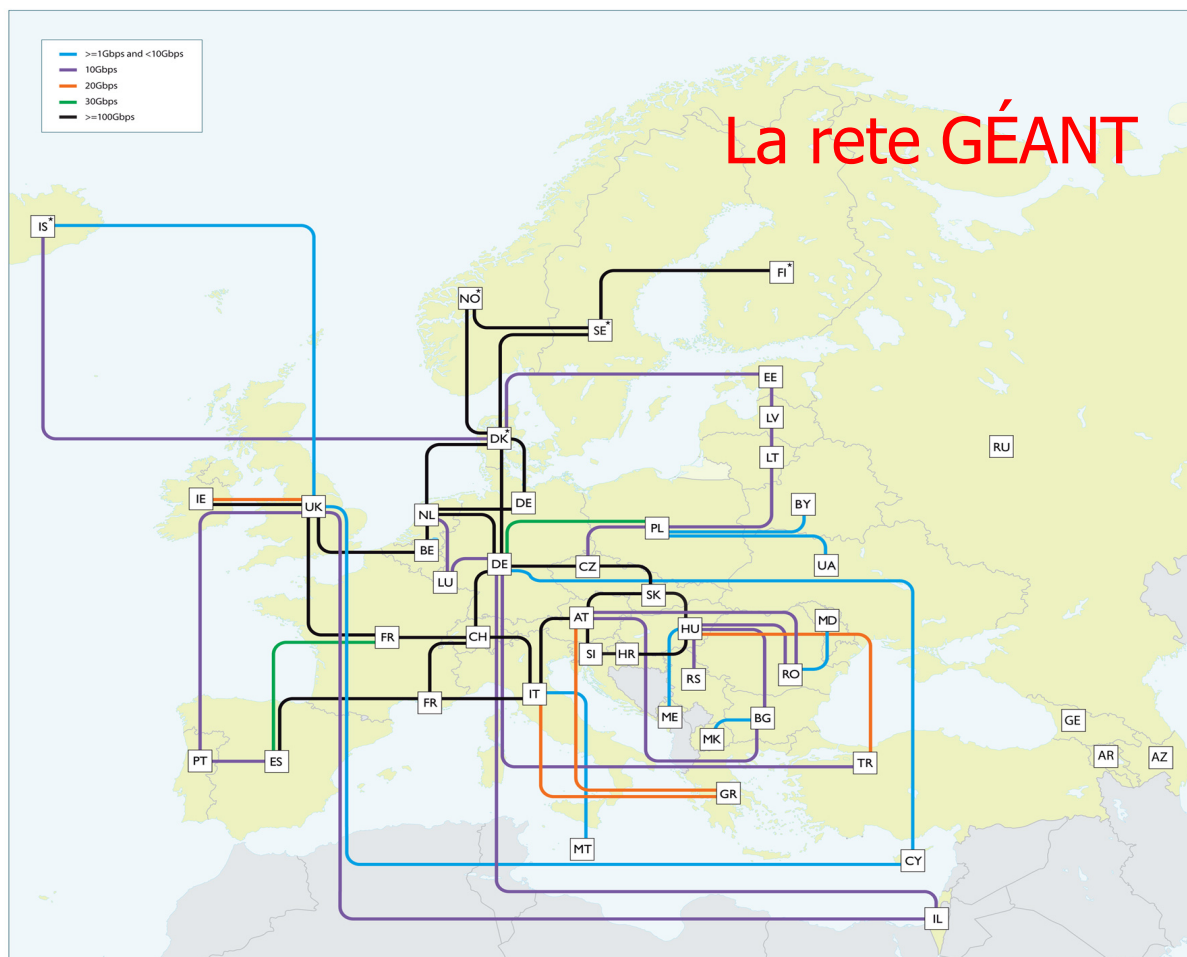
Il costo di trasmissione (per un privato)

Un abbonamento ADSL costa circa 20 Euro al mese per una velocità di 8 Mb/s in ingresso e 512 Kb/s in uscita.

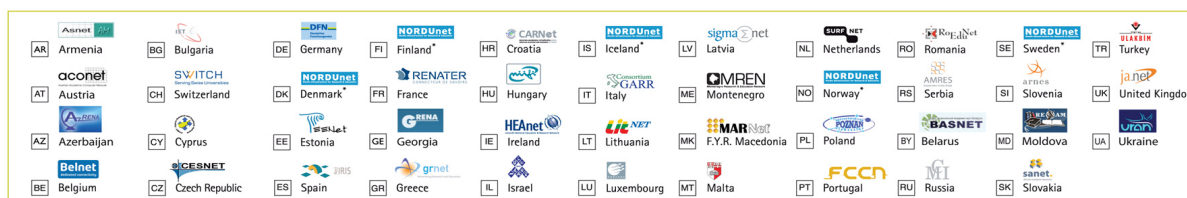
Per trasferire 10 GB (Wikipedia) ci vogliono circa 66 minuti in ingresso, pari a 7 centesimi (43 ore in uscita pari a 1,2 euro).

Un libro di testo di 500 pagine e varie immagini è pari a qualche centinaio di MB e quindi ha un costo di trasferimento in ingresso di 0,03 centesimi.

L'ecosistema delle reti della ricerca



GÉANT connectivity as at January 2014.



CAMPUS Laboratori



Reti della ricerca
(GARR in Italia)



dorsale
GÉANT

Circa 50 e più milioni di utenti in Europa solo di ricerca ed istruzione

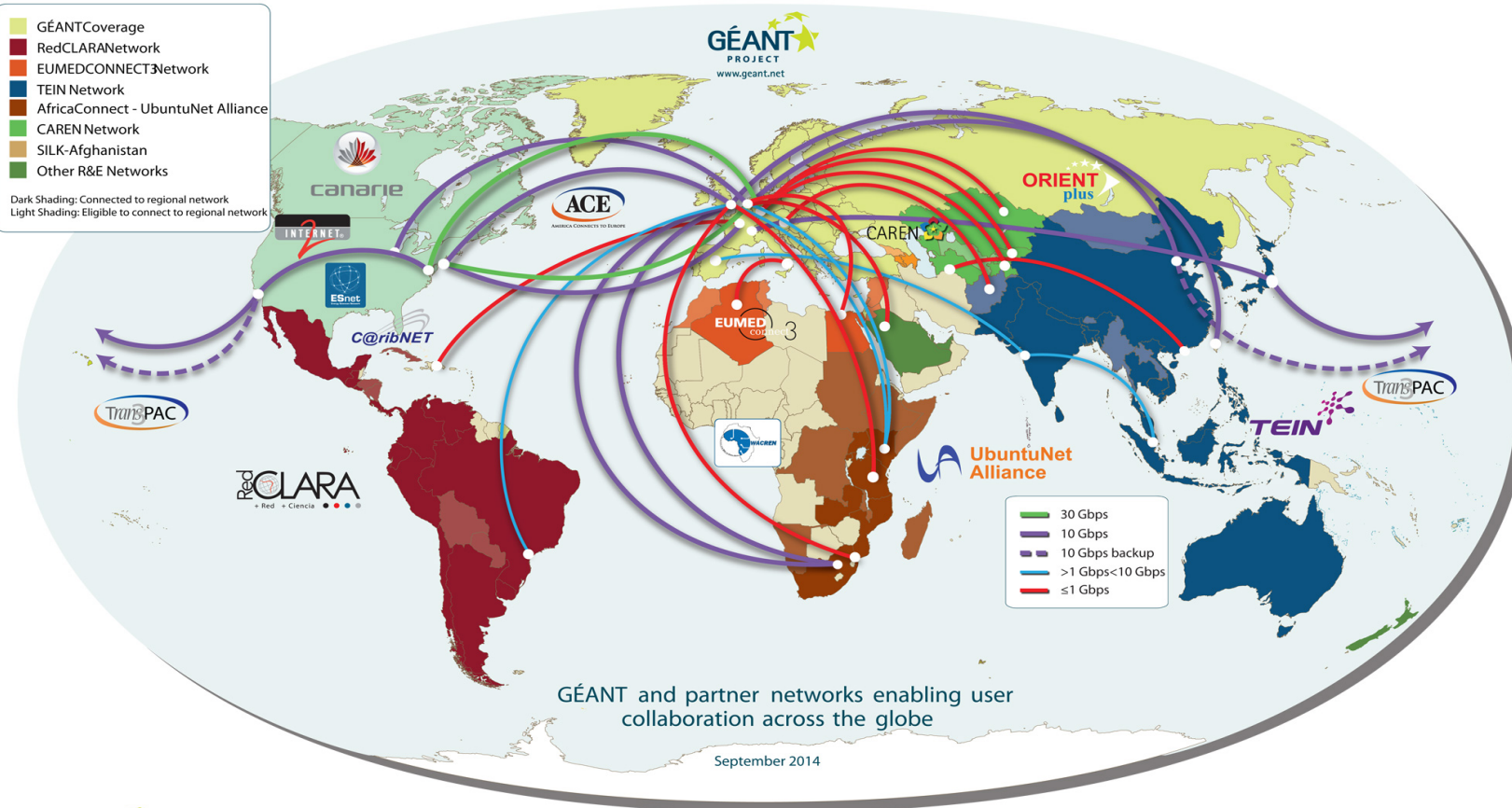
Specializzata in alta velocità di accesso a livello mondiale

Basata su fibre ottiche

GÉANT
PROJECT

- GÉANTCoverage
- RedCLARANetwork
- EUMEDCONNECTNetwork
- TEIN Network
- AfricaConnect - UbuntuNet Alliance
- CARENNetwork
- SILK-Afghanistan
- Other R&E Networks

Dark Shading: Connected to regional network
Light Shading: Eligible to connect to regional network



GÉANT
PROJECT
www.geant.net

connect • communicate • collaborate
GÉANT is co-funded by the European Union within its 7th R&D Framework Programme.



La rete GARR-X

Più di 13.000 km di fibre as uso esclusivo GARR

- ~9.000 Km di dorsale
- ~4.000 Km of circuiti di accesso

Circa 800 siti utente interconnessi

500 Gbps capacità aggregata di accesso

100 Gbps verso GÉANT

Fibre transfrontaliere con Slovenia, Svizzera e Malta (Francia pianificata).

Gestione, pianificazione e progettazione fatte da GARR a Roma (circa 15 persone)



Soci fondatori del Consortium GARR

L'armonizzazione delle reti della ricerca italiane iniziò 1986 con un finanziamento iniziale del MIUR.

Il consorzio è una associazione senza fini di lucro fondata da CNR, ENEA, INFN e Fondazione CRUI (in rappresentanza delle università statali) sotto l'egida del MIUR nel 2002.



Il funzionamento è direttamente finanziato da membri e dalle entità collegate (nessun finanziamento pubblico).

Gli utenti GARR

Circa 800 siti collegati



100 Università



350 Istituti e laboratori di ricerca



60 Istituti di ricerca Biomedica



65 Biblioteche, Musei ed istituzioni culturali



170 scuole

Accordi con i Ministeri:

- MIUR
- Ministero della salute
- Ministero dei beni culturali



*Ministero dell'Istruzione,
dell'Università e della Ricerca*



Ministero della Salute



Ministero
dei beni e delle
attività culturali
e del turismo

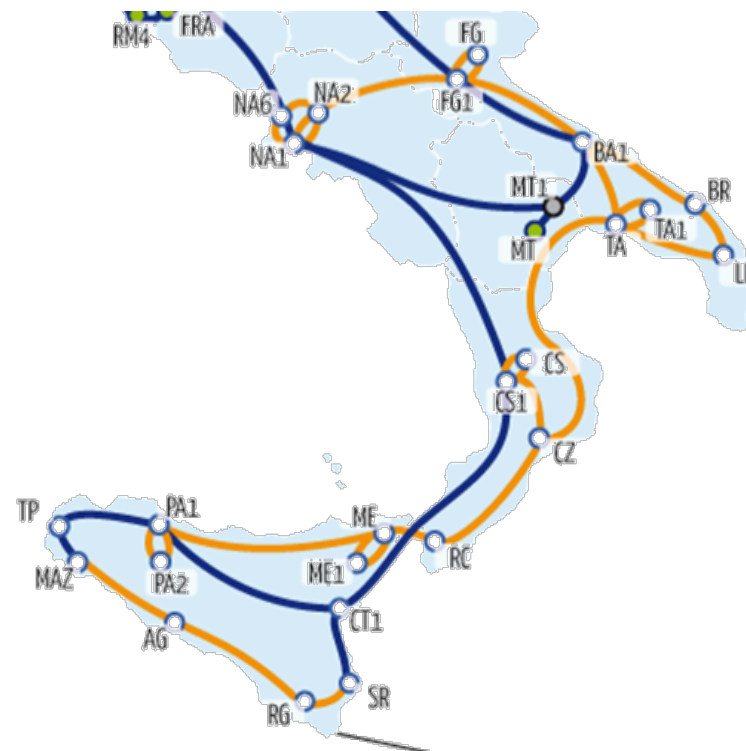
Progetto GARR-X Progress

Iniziato il 1 luglio 2013 (Calabria, Campania, Puglia e Sicilia),

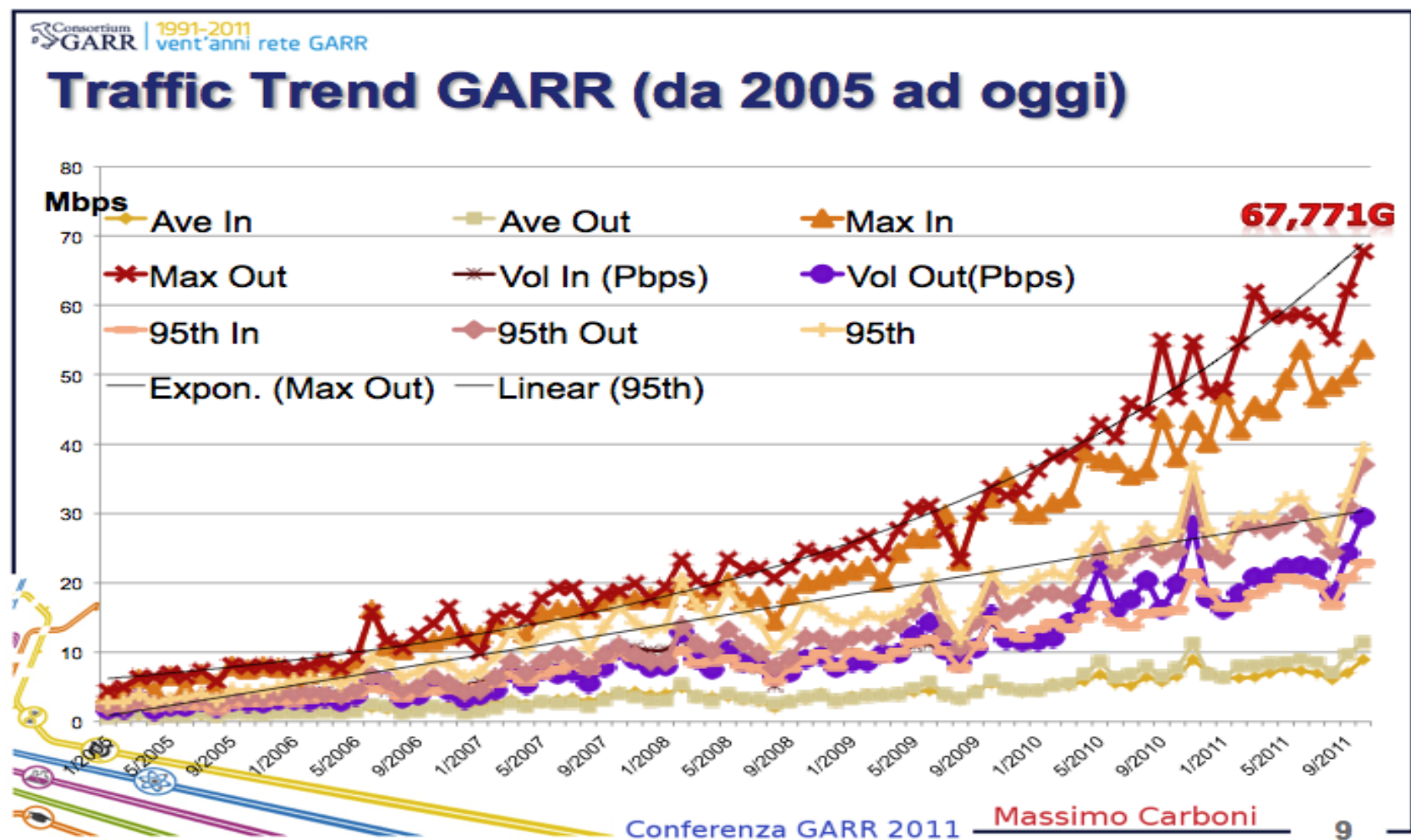
- 46.5 M€ di fondi infrastrutturali con bando MIUR
- Circa 3.900 km di nuove fibre di dorsale
- Circa 2.200 km di nuove fibre di accesso
- 24 Punti di Presenza (PoP)
- Circuiti fino a 100 Gbps

Nuova infrastruttura di calcolo distribuita :

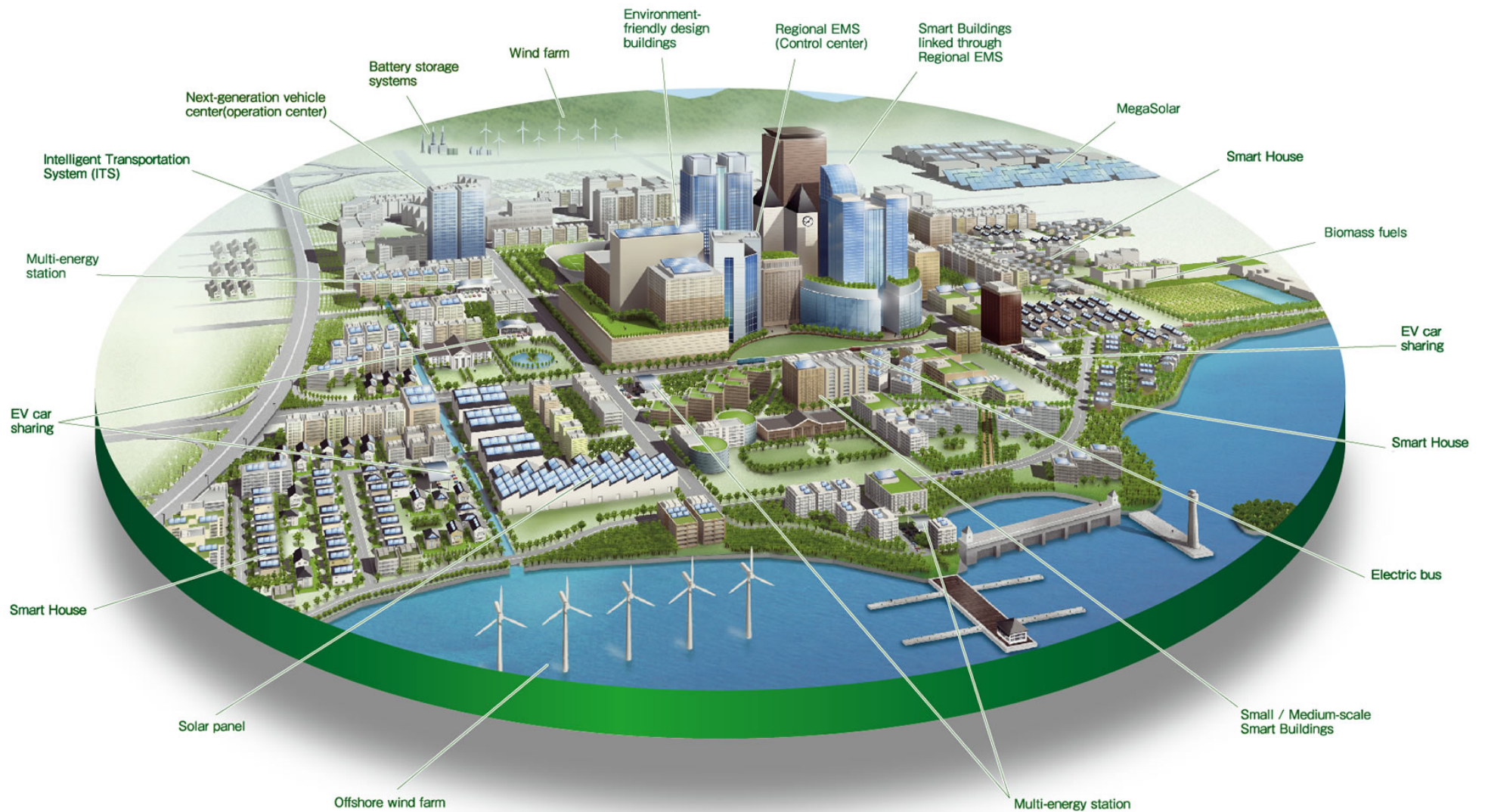
- Più di 10.000 GB di memoria di massa
- Più di 8.000 CPUs (virtuali)
- Termine installazione il 30 settembre 2015



Lo sviluppo di Internet nella ricerca in Italia



Smart Cities



"Digitalizzazione"



La nascita di Venere di Botticelli è disponibile con risoluzione altissima (occhio sinistro della Venere in alto) in

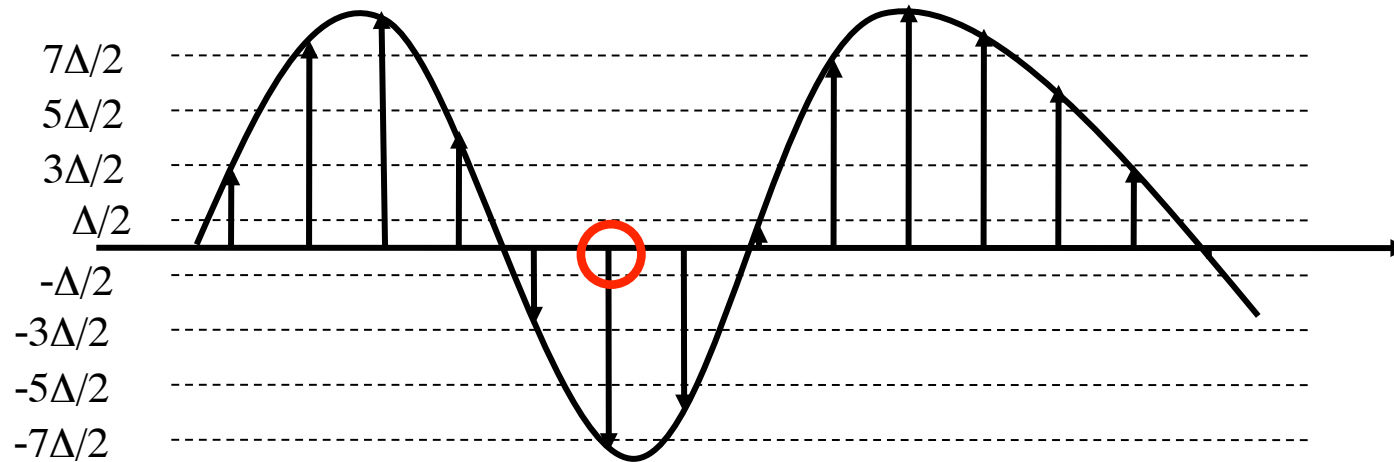
<https://www.google.com/culturalinstitute/home>

Sezione Art Project → Collections → Uffizi. La visione è interattiva.

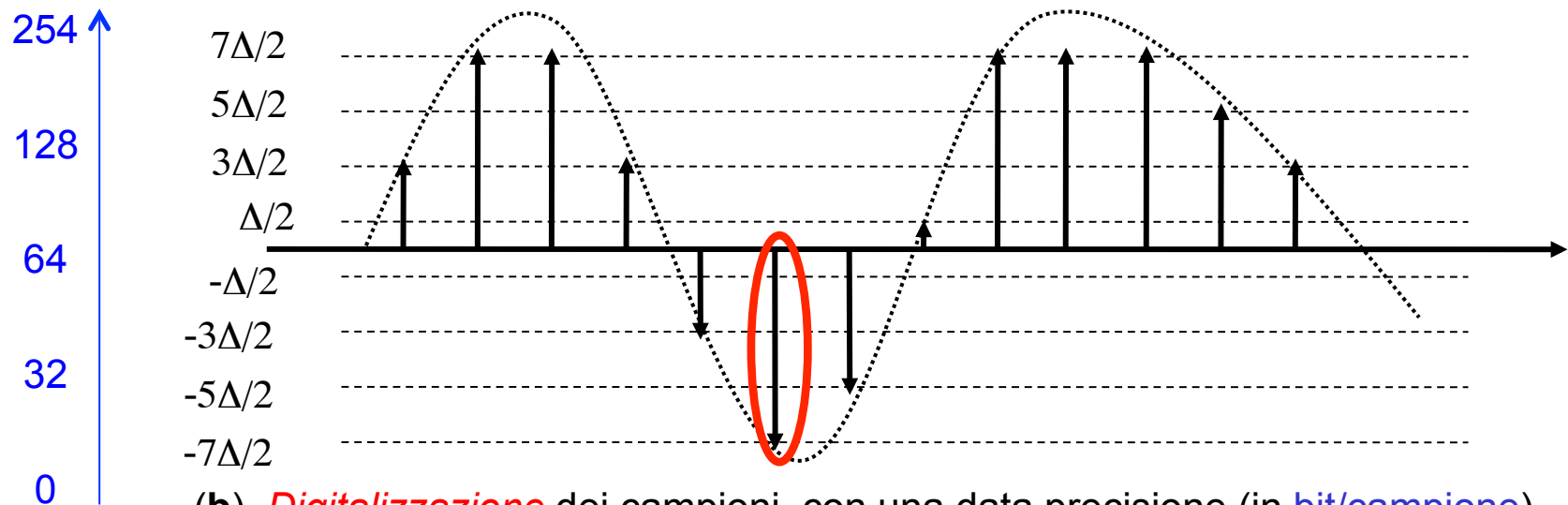
Scienza e Tecnologia, un dialogo che continua . LNGS 22 ottobre 2015



Campionamento e “Digitalizzazione”



(a) **Campionamento** della forma analogica del segnale con una certa **frequenza**



(b) **Digitalizzazione** dei campioni, con una data precisione (in **bit/campione**)

Conversione Analogica-digitale

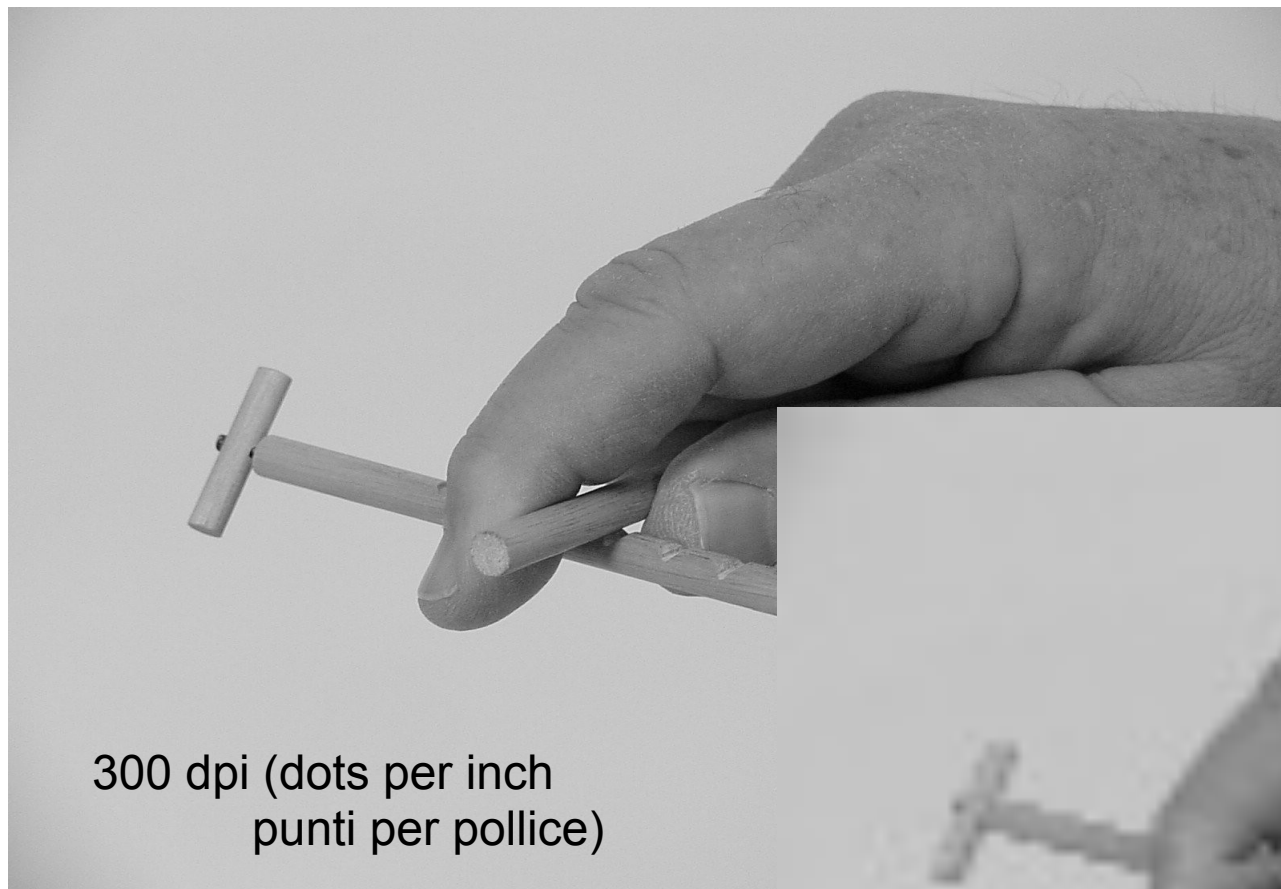
Digitalizzazione divide la quantità analogica in elementi base (campionamento) e quindi digitalizza ogni campione con una precisione definita a priori.

Per esempio il numero di pixel ed il massimo di colori per pixel

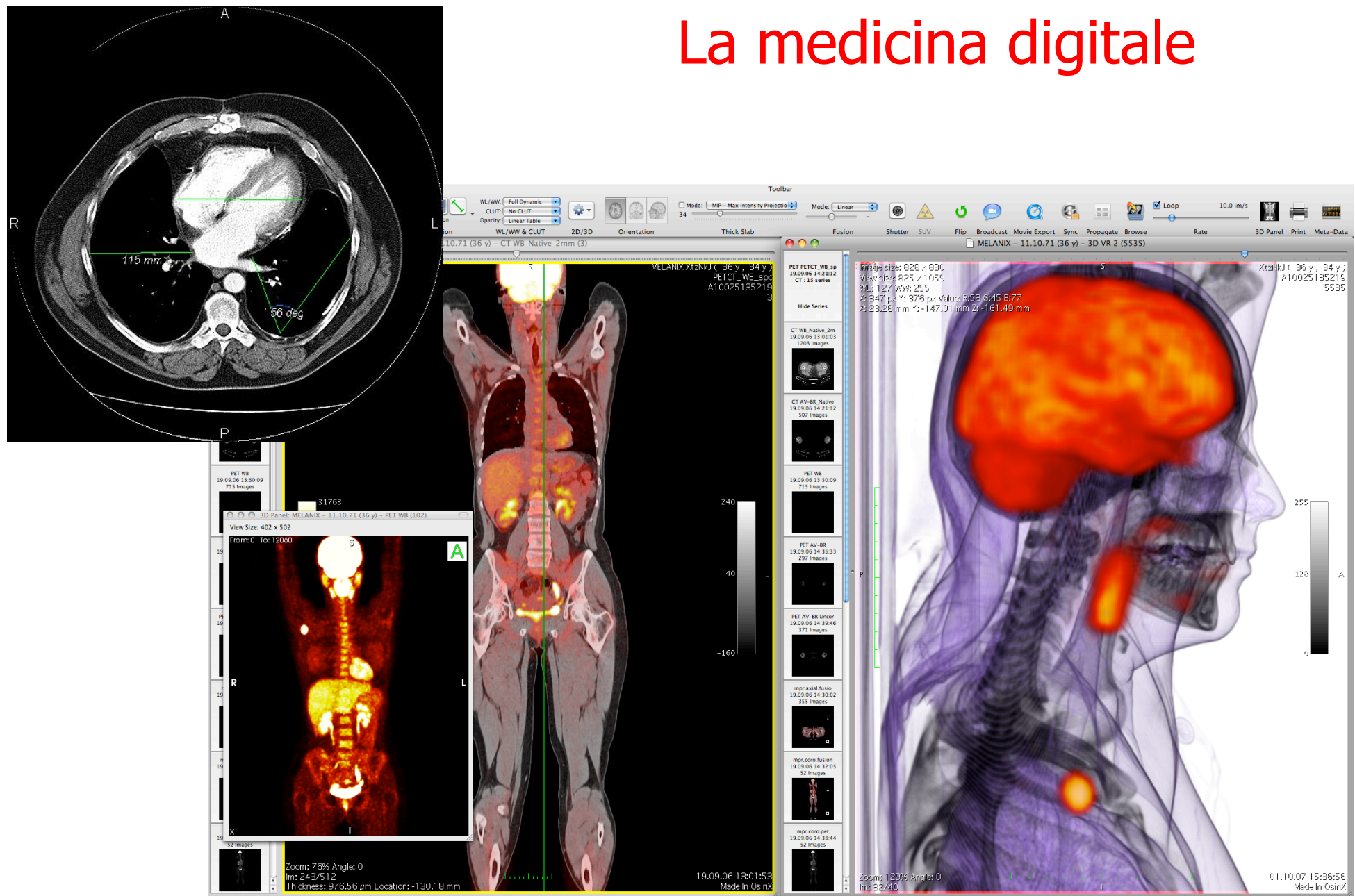
Il campionamento NON decide il **formato** (esempio bitmap, RAW, TIFF, jpeg) in cui si organizzano i dati, nè se il dato verrà compresso con o senza perdita di dati.

Una volta acquisiti i dati digitali possono essere compressi **con perdita** di contenuto (jpeg, mp3, mpeg, avi) **o senza perdita** (wav, TIFF, AAC, DICOM)

Immagine: Puntì/Pollice



La medicina digitale



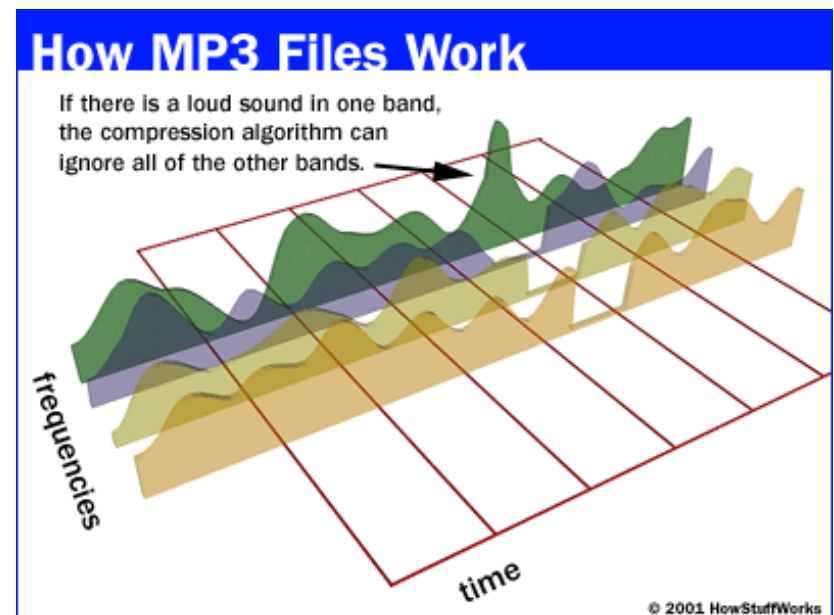
MP3 (MPEG 1/2 Audio Layer 3)

Musica : campionata a **44,100 Hz a 16 bits/campione** (se di ottima qualità, a volte a 12 bit). Due flussi separati per destra e sinistra, ciascuno da **705 Kbps** (176K byte per secondo) equivalenti a circa 32 MB per 3 minuti.

Algoritmo di compressione con perdita creato da “Moving Picture Experts Group” (MPEG), fondato nel 1988 da Hiroshi Yasuda (Nippon Telegraph and Telephone) e Leonardo Chiariglione (CSELT Torino).

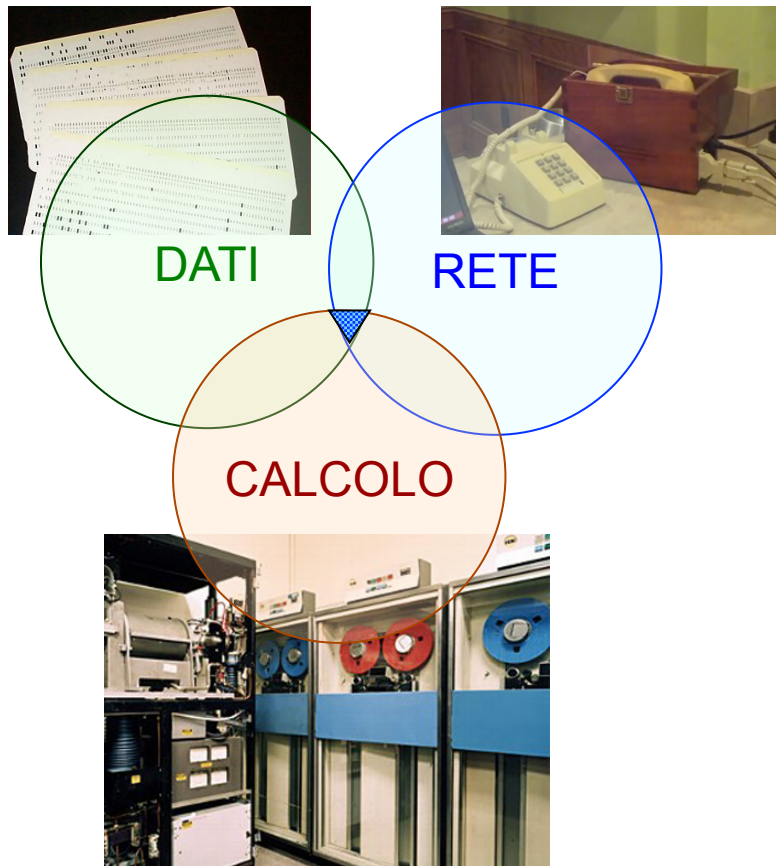
L'algoritmo prevede:

- Una analisi **spettrale**
- **Rimuovere** i suoni **al di fuori della sensibilità** “media” dell’orecchio
- Rimuovere i **suoni coperti** da altri suoni
- Scegliere un **bit rate adatto** al tipo di suono (tra 96 e 320 Kbps)

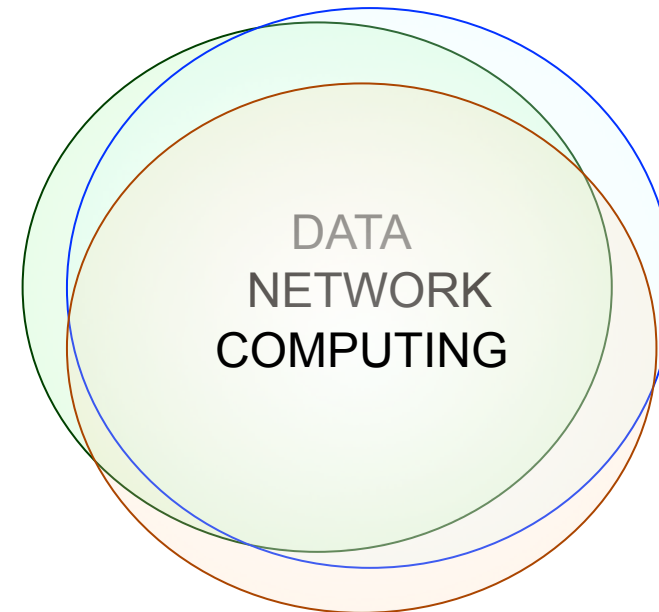


Riassumendo: mattoni dell'informatica

Ieri (1980 analogico e digitale)



Oggi (digitale)



Una sovrapposizione quasi completa della tre componenti che le rende mutuamente legate

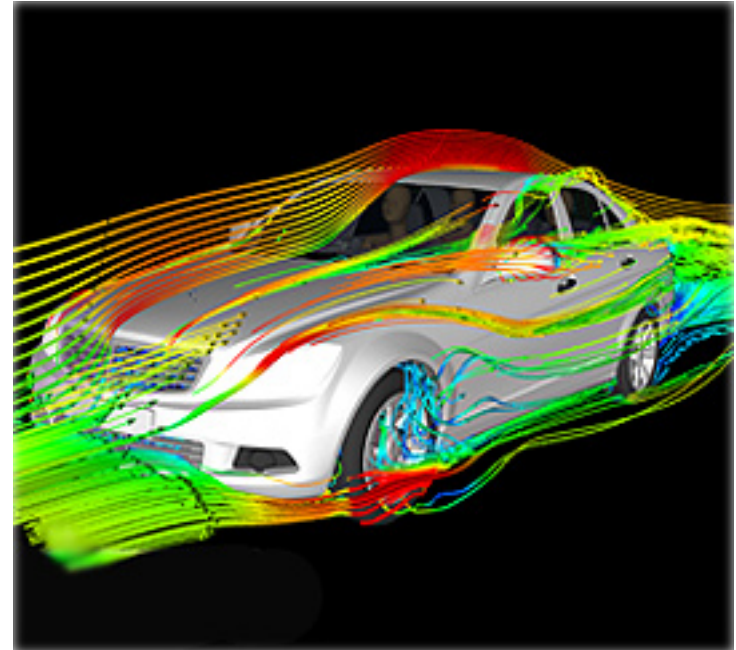
Indispensabile **usare** tutte e tre le aree per il quotidiano e la ricerca (la colla è il software)

Software

Simulazioni

La qualità delle risposte di un calcolatore dipende dalla qualità del programma e che viene eseguito e del modello che ne è alla base.

Le simulazioni sono quindi uno strumento, per capire e confrontarsi con la realtà



In questo senso, il risultato di una simulazione o di un sistema esperto ha un valore limitato.

La simulazione di realtà virtuale (per quanto sofisticata ed utile) non equivale ad un diretto confronto con il mondo fisico (si vedano le previsioni del tempo).

La tecnologia informatica in azione



L'idea è di mettere a disposizione di un utente con accesso alla rete tutta una serie di strumenti che normalmente sono applicazioni su calcolatore

- Posta (Gmail.google.com)
- Video (video.google.com)
- News (news.google.com)
- Mappe (maps.google.com)
- Voce (Google Talk)
- Lingue (translate.google.com)
- Arte <http://www.googleartproject.com/> (musei)

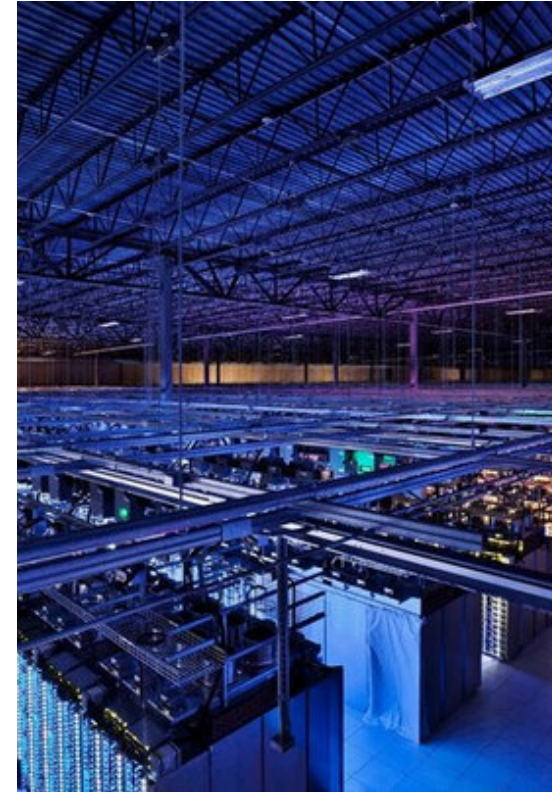
E naturalmente ricerche:

- Libri (books.google.com)
- Codici (programmi)
- Images.google.com
- Blog
- News

Ma qual'è il grado di protezione del dato personale dell'utente ?

L'analisi dei nostri dati è particolarmente pericolosa per noi.

GOOGLE – una "cloud"



La "cloud" è calcolo e servizi concentrati in **pochi** siti giganteschi raggiungibili via Internet ad altissima velocità

Stampa 3D – La rosa di Nylander

E' stampata
usando una
singola formula,
ispirata ad un
esperimento di
fisica sulla
Risonanza
Nucleare
Magnetica



A silhouette of an oil pumpjack is shown against a purple and blue sky. The word "DATA" is overlaid in large, white, sans-serif capital letters on the left side of the image.

DATA

sono il nuovo petrolio

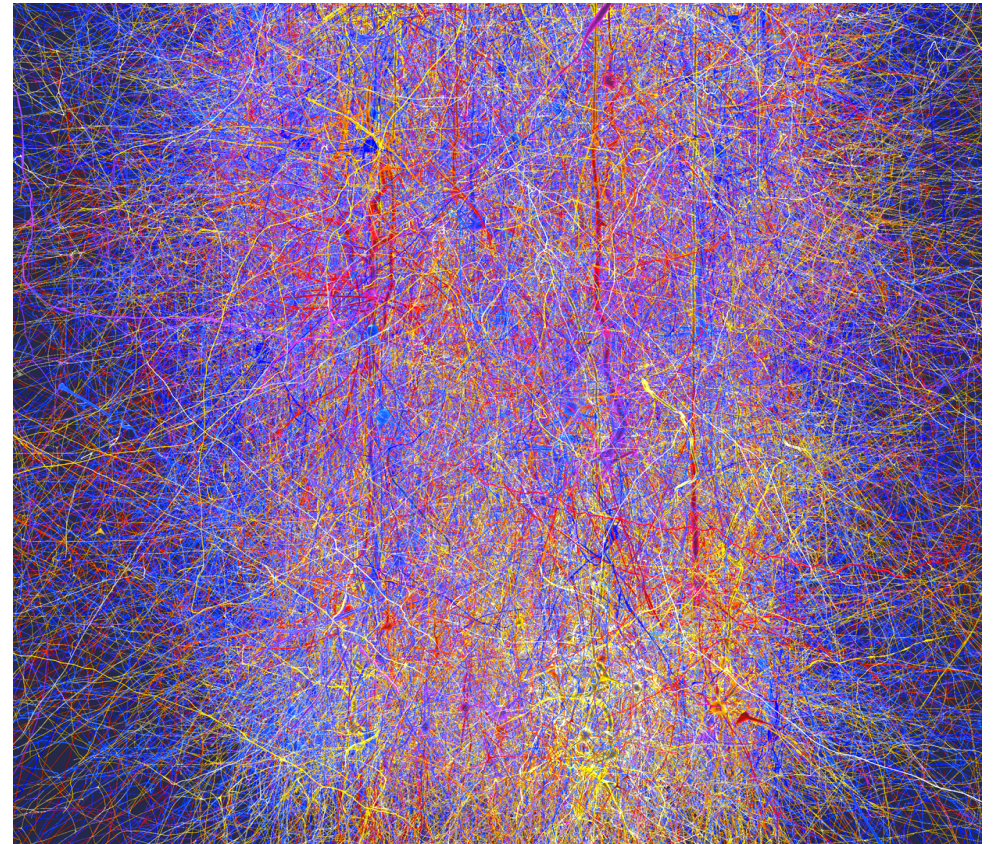
Human Brain



Progetto 'flagship' finanziato dalla Commissione Europea con 1 miliardo in 10 anni.

80 enti partecipanti con coordinamento dalla Ecole Polytechnique Fédérale di Lausanne (EPFL) in svizzera.

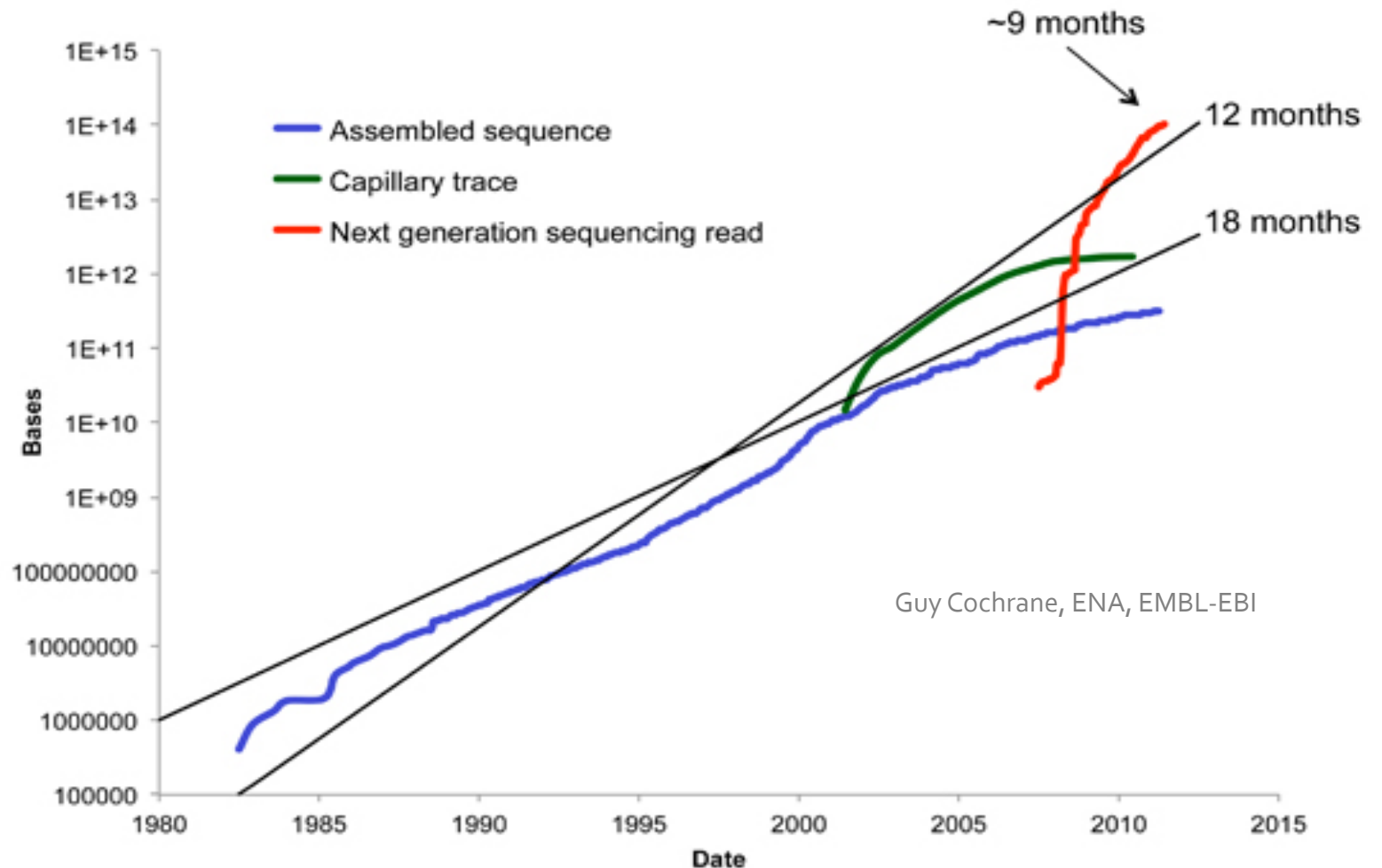
Utilizzerà l'informatica per simulare il cervello per curare le malattie degenerative



Tessuto cerebrale simulato

Big Data (DNA)

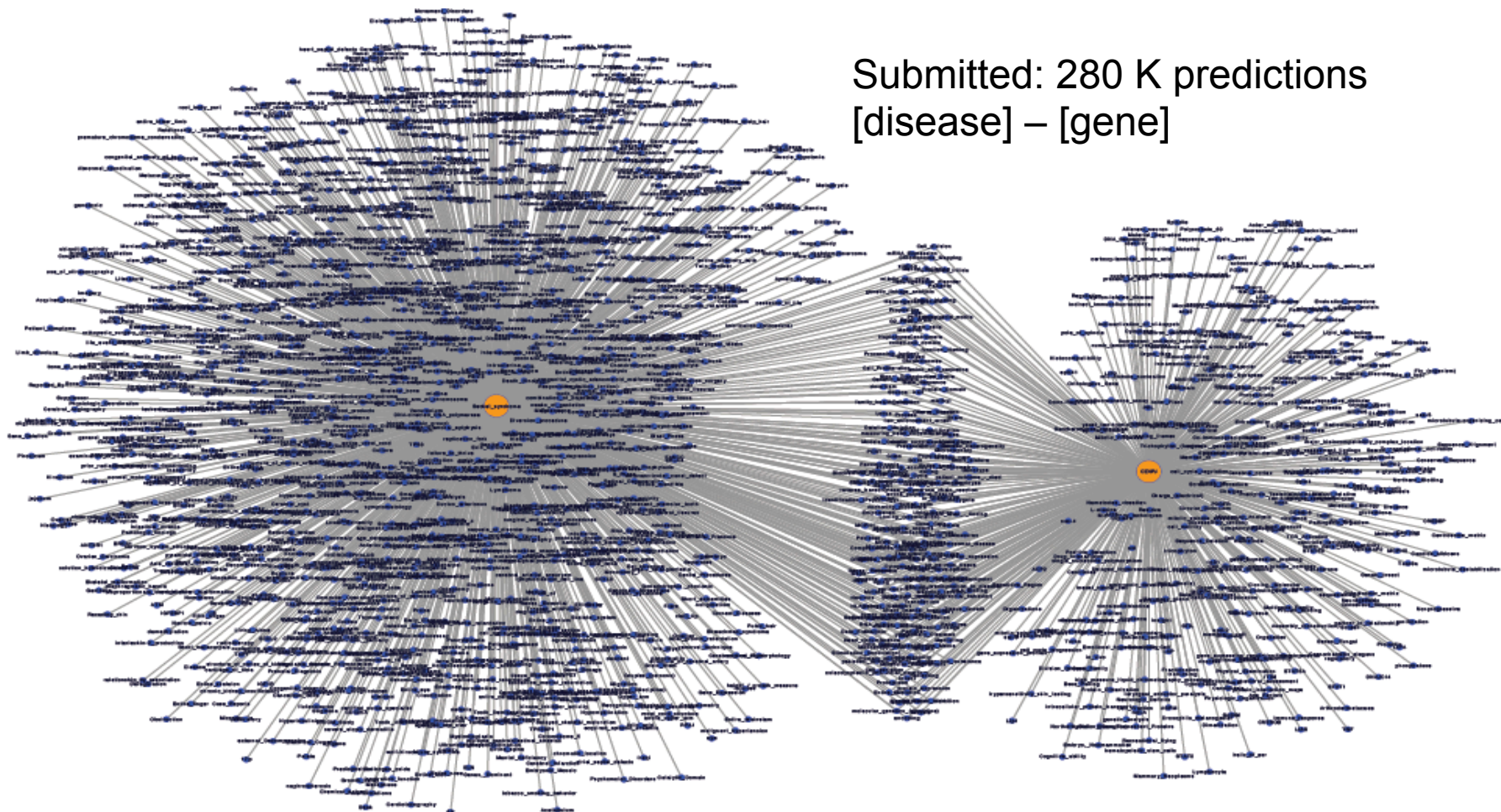
I dati sulle
sequenze di
DNA sono
raddoppiati
ogni 6-8 mesi
negli ultimi tre
anni e si pensa
che continuerà
per i prossimi
dieci almeno.



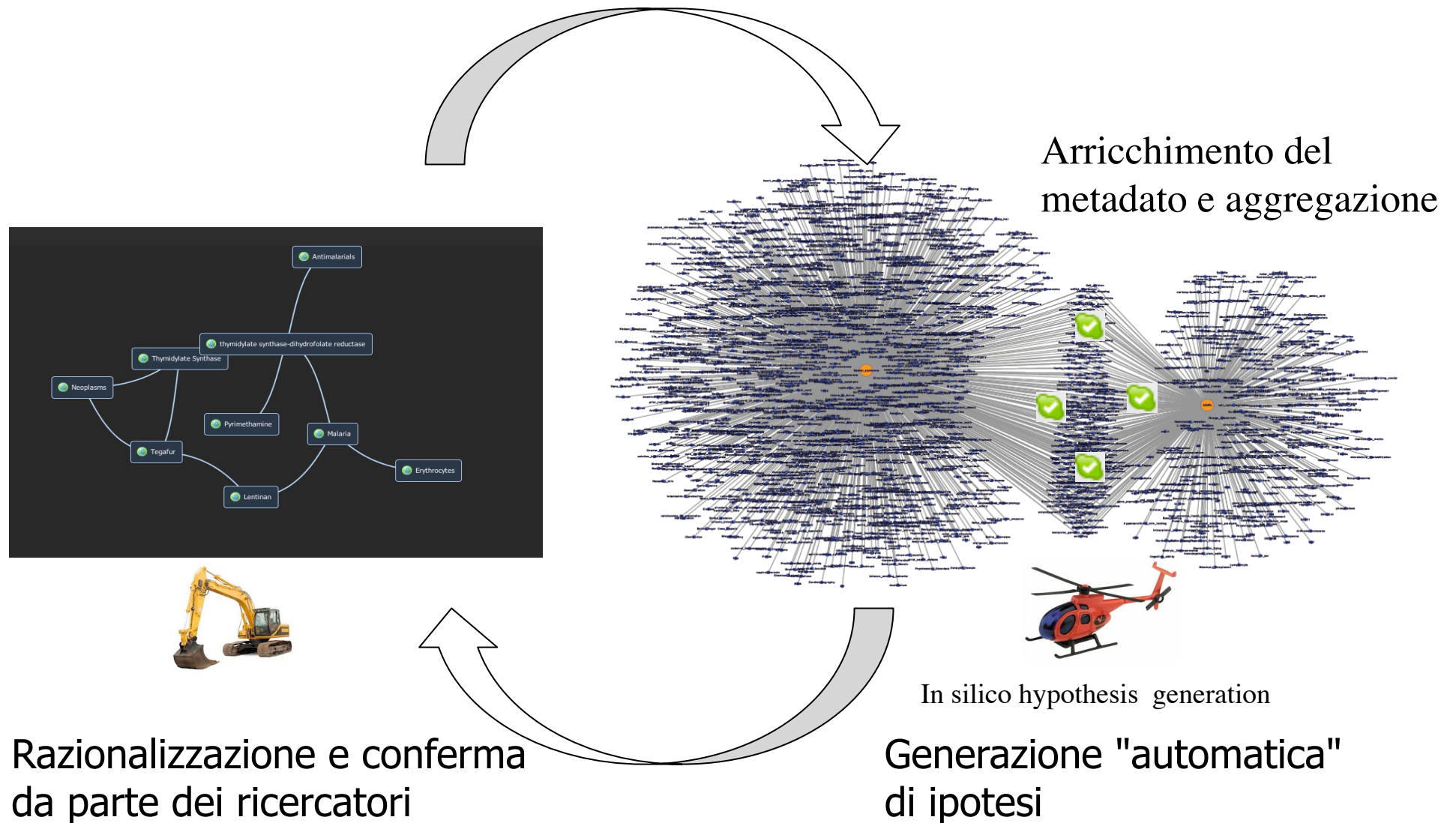
Da Barend Mons, TNC14 <https://tnc2014.terena.org/core/presentation/81>

Incrocio via computer dati pubblicati

Submitted: 280 K predictions
[disease] – [gene]

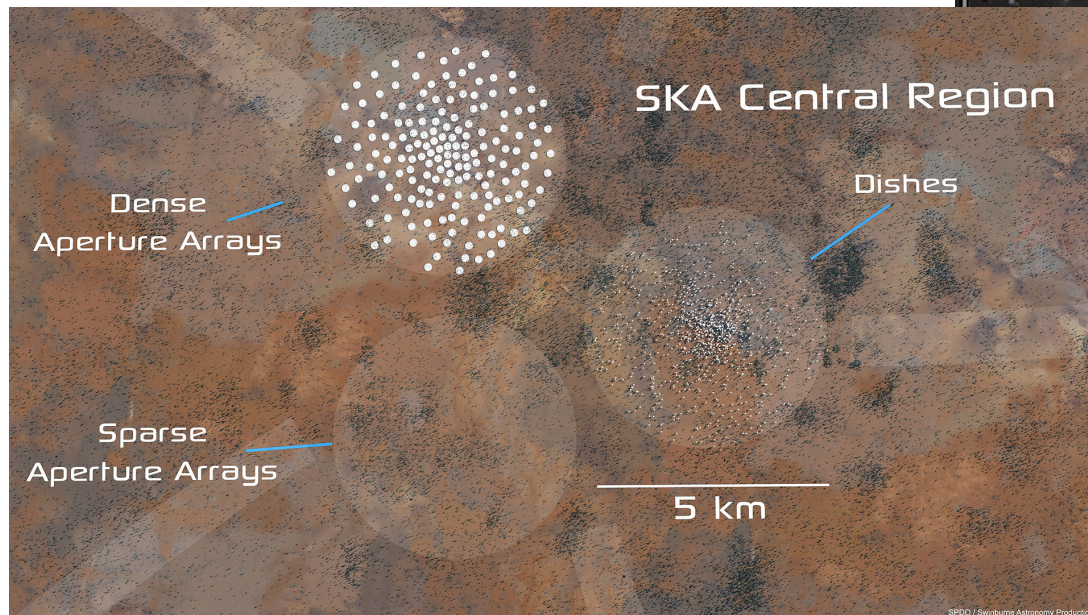
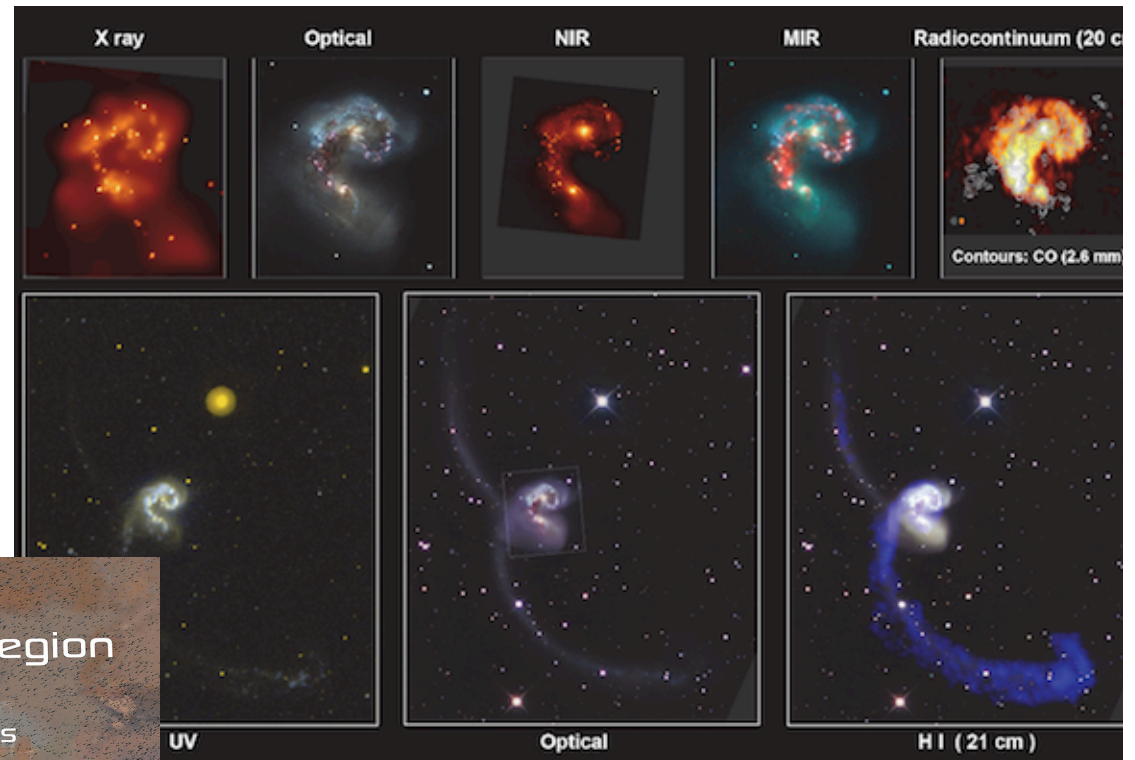


Nuova conoscenza "dal silicio"...



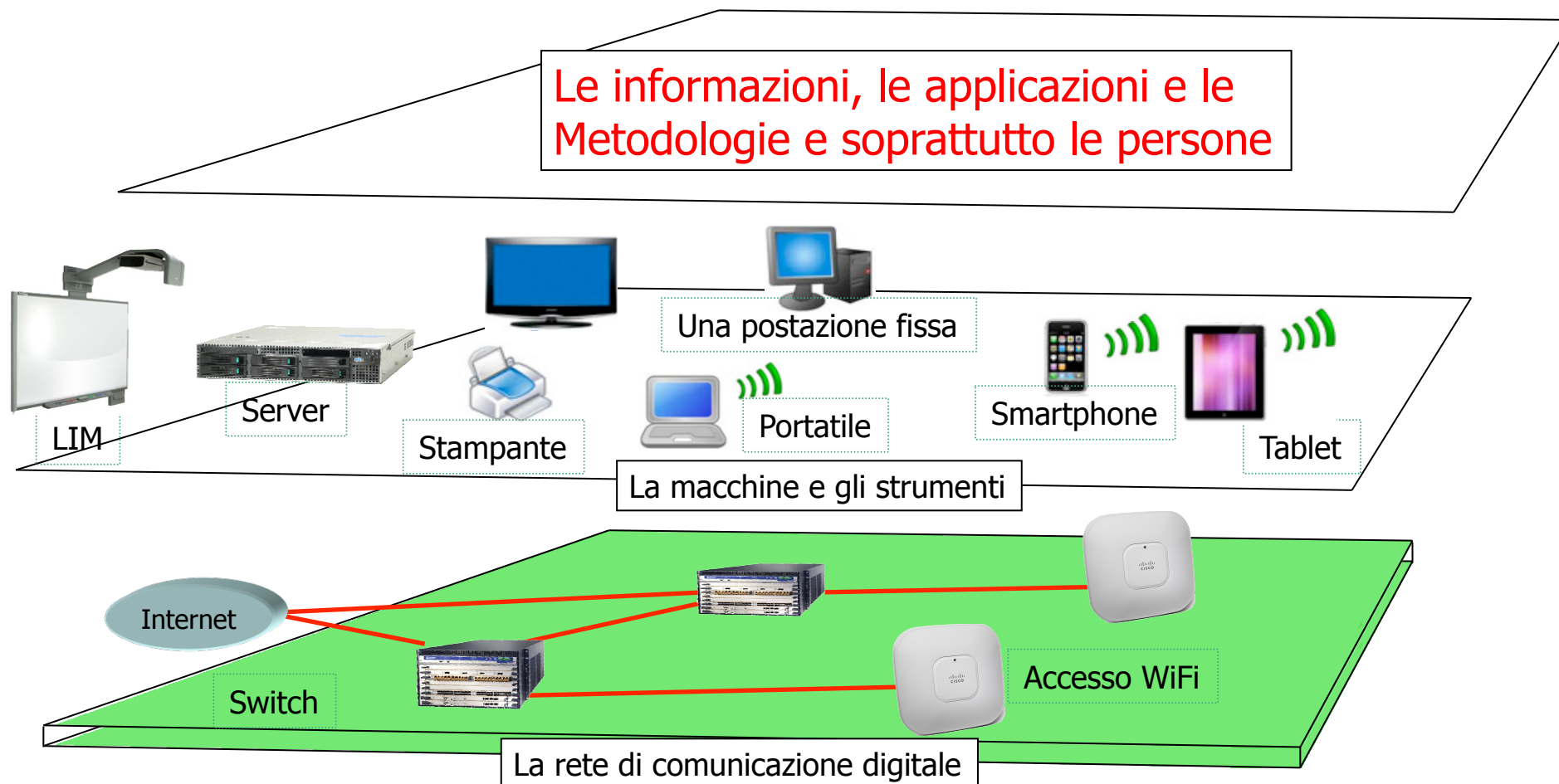
Square Kilometer Array (SKA)

Progetto di radioastronomia da circa 2 miliardi di euro che vuole analizzare in profondita' lo spazio con centinaia di antenne in Australia ed in Sudafrica



La quantità di dati creata e analizzata sarà comparabile al giorno con l'intero traffico Internet, così come le strutture tecnologiche da creare

Tecnologia Informatica in Istituto scolastico



Rete fissa o senza fili ?



Vanno usate entrambe in progetto coerente.
Le alte velocità necessarie alla ricerca ed
all'educazione sono ottenibili solo attraverso la fibra
che deve essere stesa dentro un grande Istituto.

Conclusioni (tesi)

Le tecnologie dell'informazione e comunicazione (ed Internet), nate dalla ricerca circa 60 anni fa, continueranno a svilupparsi ed a essere sempre più diffuse, molto rapidamente. *Sono basate su Silicio e Fotoni*
Stanno introducendo dei cambiamenti strutturali nella vita e nel lavoro.

La scienza sta sfruttando le nuove tecnologie di calcolo e comunicazione per ricerche prima non possibili.

Siamo di fronte ad un diluvio di informazioni (dati) e di tecnologia in tutti i campi che richiede nuovi paradigmi culturali, morali e sociali per il loro uso. *Il software ci può aiutare nell'evoluzione, ma solo se accoppiato con chiare scelte morali, legali e sociali*

L'istruzione ha il compito più complesso per l'introduzione delle tecnologie informatiche e la modifica della didattica, ma mantenendo il prezioso rapporto umano. Il compito è culturale e non può che comprendere un uso (ragionato) delle tecnologie informatiche.

Un ultima domanda ...

Stiamo creando delle forme di vita basate su silicio ?



Con cui discuteremo,
ma non umanizziamo
la tecnologia ex ante



Referenze

Le referenze sono da intendere a scopo divulgativo ed esplicativo (se non specificato diversamente)

Big Data

- https://en.wikipedia.org/wiki/Big_data
- Analisi di grandi quantità di dati (manuale universitario): "Mining of Massive Datasets" Jure Leskovec, Jeffrey D. Ullman Stanford Univ., Anand Rajaraman Millilway Labs <http://infolab.stanford.edu/~ullman/mmds/book.pdf>
- <http://www.cs.washington.edu/research/bigdata>
- <http://www.prnewswire.com/news-releases/for-big-data-researchers-network-and-compute-capabilities-are-lynchpin-to-success-300064774.html>

Referenze

Internet

- <https://www.bbvaopenmind.com/en/article/the-impact-of-the-internet-on-society-a-global-perspective/?fullscreen=true>
- <http://www.vox.com/a/internet-maps>
- <http://www.submarinecablemap.com>
- <https://www.telegeography.com/telecom-maps/index.html>

Studenti ed ICT

- <http://www.oecd.org/edu/students-computers-and-learning-9789264239555-en.htm>

Referenze

Arte e digitalizzazione

- http://www.moma.org/explore/inside_out/2013/01/03/artists-choice-trisha-donnelly-in-the-making/
- <https://www.google.com/culturalinstitute/asset-viewer/the-birth-of-venus/MQEeq50LABEBVg?projectId=art-project> ALTA risoluzione
- <https://www.youtube.com/watch?v=UoBPNTAFZMo> differenze MP3 e audio non compresso

Referenze

CPU

- <http://preshing.com/20120208/a-look-back-at-single-threaded-cpu-performance/>
- <http://www.forbes.com/sites/tomcoughlin/2014/12/22/hdd-area-density-and-tb-trends/>
- <http://www.ilpost.it/2015/04/15/legge-moore-spiegazione-10-esempi/>
- <http://www.kurzweilai.net>

Intelligenza artificiale

- <https://www.quora.com/Is-the-human-brain-analog-or-digital>
- http://futureoflife.org/AI/open_letter_autonomous_weapons

Informatica e didattica.

Esempi di risorse

- Analisi di qualsiasi parola greca (inglese):
 - <http://www.perseus.tufts.edu/hopper/morph>
- Analisi di qualsiasi parola latina (inglese):
 - <http://www.perseus.tufts.edu/hopper/morph?l=chartis&la=la>
- Corsi di ogni livello, grado ed in lingue diverse:
 - <http://www.apple.com/education/itunes-u/>
- Simulatori di Fisica: <http://phet.colorado.edu/>

Corsi Massivi Aperti On-Line

- Coursera <https://www.coursera.org/>
- edX <https://www.edx.org/>
- UDACITY <https://www.udacity.com/>

Reti della ricerca

Per informazioni sulle reti della ricerche europee e non solo:
<http://www.terena.org/compendium/>

Rete italiana della ricerca ed istruzione : <http://www.garr.it>

Rete dorsale europea della ricerca ed istruzione :
<http://www.geant.net>