



Università degli Studi di Milano
Jean Monnet Centre of Excellence

“The impact of European Union Research and Innovation
Policy upon Services of General Interest”

With the support of the Erasmus+ Programme of the European Union



Modulo 2

CONOSCENZA, DISEGUAGLIANZE, POLITICHE PUBBLICHE

Lezione 2.4

Digital Europa: come riprenderci i nostri dati

Introduzione



- È successo diverse volte nella storia della tecnologia adozione di una innovazione da parte delle autorità pubbliche abbia creato opportunità per gli investimenti privati:
- i sistemi idrici negli imperi orientali, le strade consolari romane, il servizio postale e telegrafico
- Nel secolo scorso: **telefono, elettricità, Internet, i satelliti per il radio-posizionamento**, e molto altro



- Gli stati nel passato mantenevano un certo controllo sulle innovazioni tecnologiche, se non altro per ragioni **militari**
- Oggi: Da un lato la rinuncia ai **confini virtuali** ha consentito a miliardi di utenti, individui e imprese, di sviluppare le loro attività private
- Ma paradossalmente si sono creati dall'altro lato degli **oligopoli globali di imprese basate sull'economia digitale**

- Nel processo di concentrazione oligopolistica sono in gioco dinamiche strutturali:
- secondo Autor et al. (2020) **l'emergenza di imprese superstar (come le piattaforme digitali) è correlata alla caduta della quota del lavoro sui redditi,**
- il più rilevante aspetto della crescente diseguaglianza nelle nostre società

- **La convergenza fra reti di comunicazione elettronica** di vario genere (dalla TV via cavo alla telefonia mobile, dai social media alla voce via Internet) in quanto associata **all'oligopolio su dati personali e su ogni tipo di Big Data** (compresi quelli amministrativi e quelli relativi alla salute umana) si accompagna ad uno squilibrio di potere e di concentrazione della ricchezza
- **Questo squilibrio potrebbe alla lunga essere incompatibile con i sistemi politici democratici e con società aperte, per l'enorme sproporzione di influenza che deriva dal controllo dell'informazione**

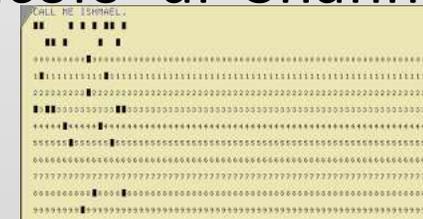
- **Come possiamo contrastare gli effetti indesiderabili del capitalismo digitale, dell'accumulazione di ricchezza e di influenza basata sulla proprietà effettiva dei dati?**
- Le imprese chiave dell'oligopolio digitale sono soprattutto statunitensi e cinesi, con alcuni eccezioni in Europa, Giappone ed in altri paesi asiatici
- L'Unione Europea si sta rendendo conto dello squilibrio

- Nel discorso sullo Stato dell'Unione 2020, von der Leyen ha annunciato che il 20% dei fondi di Next Generation EU (quindi 150 miliardi di Euro) andranno a progetti per una **Europa digitale**
- un **“cloud” europeo** per i dati industriali (in realtà tutti quelli non gestiti direttamente dalle persone, ritenendo ormai persa la competizione per i dati consumer-to-business);
- 8 miliardi di Euro andrebbero ai **supercomputer del futuro**, ad **investimenti nelle tecnologie 5G e 6G**, ad una **infrastruttura di banda larga**, ad una **e-identity** europea, ad **investimenti nella intelligenza artificiale e in spazi comuni digitali**

- Una ***e-identity*** significherebbe potenzialmente connettere i dati delle persone
- in modo tale da semplificare enormemente i sistemi di rilascio di certificati,
- riscossione delle imposte,
- pagamento delle pensioni e altre forme di assistenza,
- gestione di emergenze e tracciamento sanitario,
- con risparmi di tempo che con la pandemia sono emersi in tutta la loro portata

Come è andata

- L'idea di ridurre l'**informazione** a segni semplici è antica
- come l'invenzione della scrittura e dei numeri.
- L'adozione di un **codice binario** (0/1) che mette in corrispondenza due alternative logiche con due stati di un sistema fisico, come la configurazione aperto/chiuso di un circuito elettrico, è più recente
- Ci sono vari antecedenti ottocenteschi all'idea di un segnale binario, come il codice Morse per il **telegrafo** (breve/lungo),
- le schede perforate per i **telai Jacquard** e per le **macchine calcolatrici**, i **nastri perforati per telescriventi**
- Il '**bit**' compare per la prima volta in un celebre articolo di Shannon (1949), "*A Mathematical Theory of Communication*"

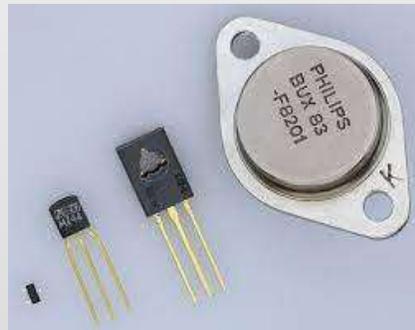




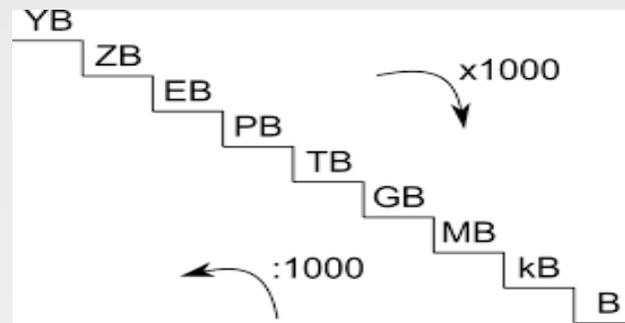
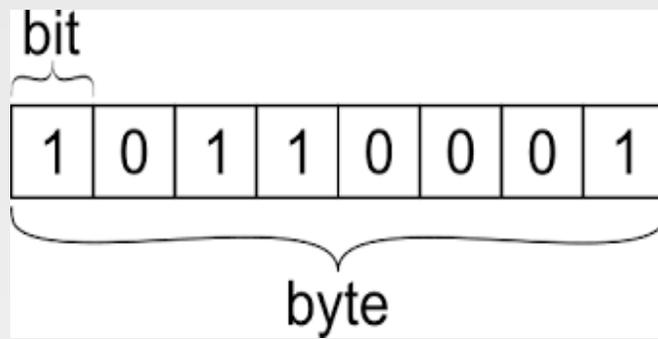
- L'articolo non casualmente era pubblicato dal Bell System Technical Journal, la rivista dei **Bell Telephone Laboratories**
- <https://www.bell-labs.com/about/history/#gref>
- Questi erano stati fondati nel 1925 congiuntamente da **AT&T**, allora dominante la telefonia negli USA, e Western Electric, impresa fornitrice delle apparecchiature, a sua volta controllata da AT&T



- i Bell Laboratories sono stati protagonisti di molti innovazioni cruciali,
- fra cui la versione di transistor a tre terminali su materiale semiconduttore che consente di regolare il trasferimento di una corrente elettrica e tradurre i segni in segnali fisici



- Settanta anni dopo viviamo immersi in un mondo di informazione digitalizzata su base binaria (e quindi di transistor)
- Molti oggetti e servizi di uso quotidiano hanno una dimensione definita in termini di **bytes** (una sequenza tipicamente di otto bit, cioè 2^3) e in multipli crescenti in ordini di grandezza mano a mano che flussi di informazione e memoria crescono
- **Dallo smartphone al supercomputer, dal radiotelescopio al televisore, dal robot al videogioco, dalla lavastoviglie al microscopio elettronico, tutto ha una dimensione di informazione codificata,**



- L' uso del byte come misura nasce in IBM nel 1956 (nel corso della progettazione di uno dei primi computer) ed è diventata lo standard anche nella comunicazione commerciale
- (sia pure con una certa ambiguità perché la base 2 è talvolta sostituita dalla base 10, per cui un **Gigabyte** (GB), viene definito come 10^9 bytes, corrispondente a circa 2^{30} bytes)

Unità di misura decimali				
Unità	Simbolo	Valore	Totale byte	Note
kilobyte	kB	1000 byte	$10^3 = 1\ 000$	La "k" del prefisso è minuscolo
megabyte	MB	1000 kB	$10^6 = 1\ 000\ 000$	$10^6 = (10^3)^2 = 1000^2$
gigabyte	GB	1000 MB	$10^9 = 1\ 000\ 000\ 000$	$10^9 = (10^3)^3 = 1000^3$
terabyte	TB	1000 GB	$10^{12} = 1\ 000\ 000\ 000\ 000$	$10^{12} = (10^3)^4 = 1000^4$
petabyte	PB	1000 TB	$10^{15} = 1\ 000\ 000\ 000\ 000\ 000$	$10^{15} = (10^3)^5 = 1000^5$
exabyte	EB	1000 PB	$10^{18} = 1\ 000\ 000\ 000\ 000\ 000\ 000$	$10^{18} = (10^3)^6 = 1000^6$
zettabyte	ZB	1000 EB	$10^{21} = 1\ 000\ 000\ 000\ 000\ 000\ 000\ 000$	$10^{21} = (10^3)^7 = 1000^7$
yottabyte	YB	1000 ZB	$10^{24} = 1\ 000\ 000\ 000\ 000\ 000\ 000\ 000\ 000$	$10^{24} = (10^3)^8 = 1000^8$

- Le infrastrutture di ricerca, dalla fisica delle particelle alla radioastronomia, dai supercomputer alle banche di biodati, generano una **quantità di informazione digitale** che costringe ad usare unità di misura di dimensione sorprendentemente elevata,
- **flussi e stock di dati: petabytes, exabytes, zettabytes, yottabytes, rispettivamente (10^{15} , 10^{18} , 10^{21} , 10^{24} bytes).**

- Si possono fare molti esempi di dimensione in bytes dell'informazione che circola (come flusso per unità di tempo) o che si accumula (come stock in memoria) nel nostro pianeta
- Un **DVD-ROM** tipicamente ha una capacità di 4,7 Gigabytes,
- una puntata di una serie **Netflix** ad alta definizione corrisponde a 7 Gigabytes,
- la memoria di un **MacBook** a 16 Gigabytes



- nel 2016 in **YouTube** corrispondevano a 24 Terabytes, circa due volte e mezza i dati trasmessi dal telescopio spaziale Hubble;
- 10 miliardi di foto su Facebook corrispondono a **1,5 Petabyte**
- Si dice che i dati posseduti da **Google** (qualche anno fa) fossero nell'ordine di 15 exabytes



- Qualche esempio può aiutare visualizzare questi ordini di grandezza
- Il primo sequenziamento di un genoma umano (1990 - 2003) da parte di **Human Genome Project** può essere visto come la produzione di informazione su tre miliardi di basi di DNA, approssimativamente **3 gigabytes**
- La prima volta sono stati necessari circa 3 miliardi di dollari e oltre dodici anni

- Oggi la stessa operazione può essere completata ad un costo ben inferiore ai mille dollari e in meno di un'ora, grazie alla tecnologia sviluppata grazie a HGP **da un piccolo gruppo di imprese specializzate**
- Nello scorso decennio uno dei principali centri di ricerca del mondo in questo campo, **il Broad Institute (Harvard e MIT) produce un sequenziamento ogni dodici minuti, ha accumulato in dieci anni informazioni su centomila individui, e dispone di un archivio di 70 petabytes di dati genetici**

- Lo stesso andamento esponenziale nell'accumulo di informazione genetica si sta registrando presso altri centri con un raggio di studio che comprende migliaia di specie viventi ed estinte
- Una società privata (23andMe) che offre un sequenziamento parziale (per poco più di cento dollari) ha dieci milioni di clienti e quindi i dati sul loro genoma
- **Si prevede che nel 2025 i dati sul genoma corrisponderanno ad una quantità di informazione superiore a quella, che pur sembra sconfinata, dei dati cumulati di Facebook, Youtube e altri social media (per quanto ipotetiche siano queste previsioni)**



- Altri due esempi nel campo dell'astronomia sono il **Large Synoptic Survey Telescope (LSST) in Cile** che ci si attende che intorno al 2025 inizierà la scansione ottica di 37 miliardi di galassie e stelle in dieci anni, corrispondenti alla necessità di gestire 30 terabyte di dati ogni notte
- Il **James Webb Space Telescope** (NASA, ESA, Agenzia Spaziale Canadese), erede del telescopio Hubble, viaggia nello spazio profondo a 1,5 milioni di Km dall'orbita terrestre, al costo di 9 miliardi di dollari, coprendo galassie distanti non visibili da terra, esoplaneti ed altri oggetti celesti



- Si prevede che si dovranno risolvere complessi problemi di **classificazione** delle immagini con l'ausilio di computer,
- creazione di un **catalogo** che incrocia diverse lunghezze d'onda dei segnali,
- **evoluzione temporale** delle osservazioni,
- calibrazione di **modelli** interpretativi,
- oltre ovviamente ai problemi di **comunicazione** affidabile dei dati dal telescopio spaziale alle stazioni di terra

- Può essere utile confrontare l'archivio di dati di **Facebook** con quello del **CERN**
- Nel primo caso una stima di qualche anno fa considerava che circa 300 petabytes di contenuti fossero archiviati nei vari server (HIVE) di Facebook con un flusso quotidiano di 3-4 petabytes con forse oltre 2 miliardi di utenti attivi
- **Il Data Center del CERN dispone di 10000 server con 174000 processori fisici e gestisce 1 petabyte al giorno di dati, l'equivalente di 210000 DVD**



- Dopo le operazioni di filtro, 500 petabytes di dati sono attualmente archiviati, mentre 10 gigabytes al secondo vengono inviati alla **Worldwide LHC Computer Grid**, l'erede del **WWW**, che connette 170 nodi in 42 paesi a loro volta connessi a migliaia di altri computer
- **In altre parole il CERN, con la sua relativamente piccola comunità di utenti scientifici (forse diecimila persone) genera e gestisce più dati di Facebook con i suoi oltre due miliardi di utenti, e con risorse economiche in termini di entrate annue (2020) che sono circa 70 volte quelle del CERN**

- Infrastrutture di ricerca, *internet companies*, così come pubbliche amministrazioni e società di comunicazione digitale (dalle Telecom ai gestori di contenuti online) gestiscono dunque flussi di dati sulla scala dei petabytes, e in prospettiva di ordini di grandezza superiori, come exabytes e zettabytes all'anno con una crescita apparentemente senza limiti
- **Alcune stime della produzione di dati nel mondo nel 2019 sono nell'ordine di 33 zettabytes nel 2018, cioè settemila miliardi di DVD, 40 zettabyte nel 2019, con la previsione di 50 zettabyte nel 2020 (senza contare gli effetti della pandemia che aumentano il flusso**

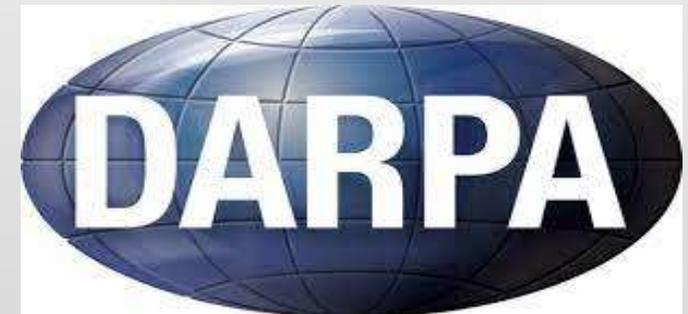
- Tutta questa quantità di informazione non sarebbe gestibile se non si fossero verificate in sequenza o talvolta in parallelo varie rivoluzioni tecnologiche, apparentemente una indipendentemente dall'altra, quasi sempre con il ruolo importante di agenzie pubbliche per innescare il processo
- Diversi autori hanno menzionato il ruolo di **ARPA** nella creazione di **internet** e del CERN come luogo in cui il World Wide Web è stato inventato



- Nel 1957 l'URSS lancia lo **Sputnik** e gli Stati Uniti si sentono minacciati
- L'ex-generale, presidente USA pro-tempore, Eisenhower lancia **l'Advanced Research Projects Agency** (poi ribattezzata DARPA aggiungendo un Defense all'inizio in modo che non ci fossero dubbi sulla missione), con un budget iniziale che espresso in Euro attuali si aggirerebbe sui 3,74 miliardi, da spendere nel finanziamento di progetti di ricerca potenzialmente utili alla difesa

- **Nonostante ARPA non avesse laboratori e neppure uffici (ma un generoso budget e un management capace) come scrive The Economist “l’agenzia ha dato forma al mondo contemporaneo, contribuendo a creare la difesa missilistica, la tecnologia di invisibilità al radar (*stealth*), internet e il personal computer, il laser e il GPS”**

- A quanto sappiamo il management aveva larghissima autonomia e secondo la storia dell'agenzia (Weinberger 2017) la decisione di creare **ARPANET**, il progetto precursore di internet fu presa in una riunione di mezz'ora
- **DARPA a fine 2019 aveva un budget di 3,4 miliardi di dollari, non molto minore di quello iniziale, con uno staff di 220 persone, cioè è rimasta essenzialmente una struttura agile con una grande budget, con cui finanziare progetti di frontiera**



- La storia del World Wide Web al CERN è nota al pubblico non specialistico ed è stata raccontata dallo stesso inventore **Tim Berners Lee**
- Berners Lee era un giovane dipendente del CERN ed aveva osservato lo spreco di tempo per i fisici di particelle del mondo che usavano computer e software fra loro non compatibili, per cui non riuscivano ad interfacciare i loro sistemi e a condividere i dati, se non con procedure ad hoc



- L'idea iniziale del WWW (una serie di codici e protocolli fra cui **http** e **HTML**) non attirò molto interesse e iniziò modestamente, mettendo l'elenco telefonico del CERN nel neonato WWW
- Un po' dopo, fuori dal CERN, i protocolli del WWW vennero adottati per l'archivio elettronico dei documenti di un altro acceleratore di particelle: SLAC a Stanford (USA)

- Dato lo scarso interesse iniziale del CERN stesso per lo sviluppo del progetto Berners Lee passò al MIT nel 1991, mentre il CERN nel 1993 – pur detenendo i diritti del WWW – decise di cederli gratuitamente, così come fece lo stesso Berners Lee
- Decisioni ‘fatali’ che hanno creato **una incalcolabile esternalità positiva** per miliardi di persone e imprese

- In realtà il CERN comprese presto che le potenzialità erano enormi e **la prima conferenza mondiale WWW si tenne al CERN nel 1993**, con il fondamentale apporto di Robert Caillau (che molti considerano co-inventore del web)
- DARPA rientra anche in questa storia perché sostenne il World Wide Web Consortium creato al MIT, con la co-sponsorizzazione della Commissione Europea
- **Il consorzio tuttora persegue una politica di accesso aperto e gratuito alle tecnologie e agli standard che sviluppa**
- **<https://www.w3.org>**



- **Nonostante inizialmente usi commerciali dei codici sembra fossero vietati, già nel 1995 nasce l'e-commerce con Amazon ed eBay**
- Entro pochi anni vi sarebbero state centinaia di start-up e collocamenti in borsa di *internet company*, fra cui **Google** fondata nel 1998
- Le valutazioni speculative portarono alla nota bolla delle *dot.com* (2001), ma le imprese sopravvissute sono oggi al vertice del potere economico mondiale

- **Secondo alcune stime gli utenti del WWW sarebbero 4,66 miliardi di persone, pari al 59 per cento della popolazione mondiale (95 % in Europa), fra cui oltre 4,1 miliardi di utenti dei social media, di cui Facebook, Instagram, Messenger, Whatsapp rappresentano 2,3 miliardi nel 2019**
- **In dieci anni Facebook è passata da 2000 dipendenti a 45mila, Google da 24mila a 119mila**

- Per guardare più ad Oriente, la cinese **Tencent** oggi ha un valore di mercato di 580 miliardi di dollari, è leader nei videogiochi online (che per lo più regala ma poi vende oggettistica connessa), controlla una dozzina di società del settore, anche in Usa, Finlandia ed altrove
- La sua app (**WeChat**) ha 1,2 miliardi di utenti e si stima che due terzi dei cinesi usino videogiochi online dal loro smartphone (negli USA un terzo della popolazione)





- La app **Tik Tok** è stata scaricata 2 miliardi di volte
- ByteDance, la società che la ha sviluppata, potrebbe valere (non è ancora quotata in borsa) fino a 100 miliardi di dollari,
- ha 60mila dipendenti e oltre 100 milioni di utenti sia in Cina che negli USA

- Incidentalmente il mercato dei **videogiochi** in genere (Nintendo ed altri) grazie anche alla pandemia è stimato 175 miliardi USD nel 2020 (dati Newzoo) e richiede tecnologie piuttosto sofisticate
- **Se ciascuno degli utenti dei servizi di queste imprese basate su internet pagasse agli inventori una royalty di due centesimi di Euro al mese, il CERN avrebbe il proprio bilancio interamente finanziato dal WWW, senza bisogno di ricorrere agli stati membri**
- Per quanto le storie di internet e WWW siano suggestive del ruolo delle agenzie pubbliche nel creare il mondo digitale che conosciamo, non esauriscono la catena di eventi e forze in gioco

- In primo luogo, come ho ricordato sopra, l'idea stessa di adottare segnali binari per trasmettere informazione ha avuto applicazione su larga scala già con i sistemi telegrafici nel XIX secolo
- Lo stesso **Morse** ebbe l'appoggio del **governo USA** (uno stanziamento del Congresso per la prima linea operativa)
- **L'importanza del telegrafo elettromagnetico per finalità militari venne colta molto presto ovunque, così come era accaduto precedentemente per i sistemi postali**



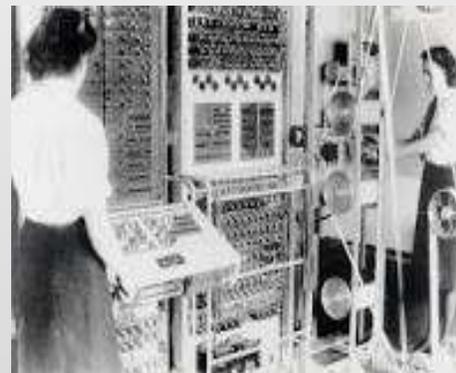
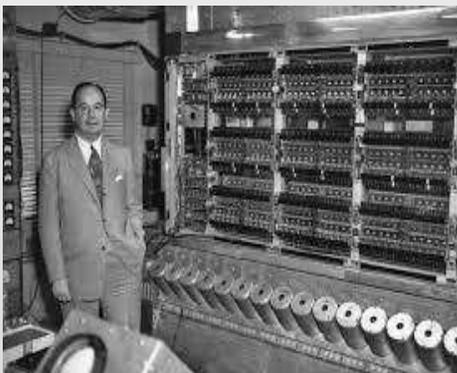
- È da qui che nasce il servizio pubblico di **Poste e Telegrafi che in Italia risale al 1862 (Regie Poste)** <https://www.posteitaliane.it>
- Se la successiva evoluzione della telefonia analogica per la voce si basava su di un principio diverso, con il passaggio dall'analogico al digitale si è tornati all'uso dei codici binari
- Qualcosa di simile è accaduto (o sta accadendo) per la televisione e la radio con segnali digitali



- **Nella grande maggioranza dei paesi, poste, telegrafo, telefono sono stati visti dai governi come parte di una infrastruttura pubblica di comunicazione, come tale, quando non gestita direttamente da imprese pubbliche, affidata ai privati con meccanismi di concessione**
- **É difficile sottovalutare il ruolo innovativo nel secolo scorso delle imprese di telecomunicazione, in larghissima misura pubbliche in tutta Europa**

- In secondo luogo, la nascita del computer, nonostante si trovino macchine calcolatrici meccaniche di vario tipo anche nei secoli precedenti,
- è all'inizio essenzialmente legata ad **impieghi militari di crittazione e decifrazione** di messaggi durante la Seconda guerra mondiale,
- cioè in definitiva all'esigenza di '*hackerare*' (diremmo oggi) le comunicazioni radio in codice

- Il caso più noto ed emblematico è la macchina **Colossus**, basata sulle idee di **Turing**
- costruita per la Royal Navy per decifrare il codice **Enigma** tedesco verso la fine della Seconda guerra mondiale (1944),
- replicata segretamente in numerosi esemplari, con l'impegno di migliaia di tecnici, e capace di decifrare quotidianamente migliaia di messaggi del nemico



- Senza una concatenazione di antecedenti forse non sarebbe esistito il primo calcolatore 'per il mercato ' (**UNIVAC 1951**), sviluppato da parte dagli stessi progettisti che avevano lavorato nei progetti militari
- **Se IBM prima e Apple dopo ci sembrano l'archetipo dell'impresa privata negli USA, oggi sappiamo quanto nella loro storia si deve alla ricerca condotta, essenzialmente con finalità militari, in laboratori e università finanziate dai governi**

