



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 2.7

Digital Europa: come riprenderci i nostri dati

Le opportunità sprecate

- I dati non sono di per sé un **bene pubblico**
- Possono essere **escludibili** ad esempio con *firewall* o criptandoli o con altri mezzi,
- ma in linea di massima il loro consumo non è **rivale**,
- -come in generale la conoscenza- il loro **uso da parte di un soggetto non toglie nulla alla possibilità di un altro soggetto di usarli a sua volta**

- Ma per essere usati i dati grezzi debbo essere **'puliti'**,
- gli algoritmi di intelligenza artificiale debbono essere **allenati**,
- **l'archiviazione sicura**
- ed **estrazione affidabile** non è senza costi
- come potranno essere definiti **i diritti di proprietà** e di accesso ai dati?

- Il movimento a favore degli '*open data*' si contrappone a quello che propone di creare delle **istituzioni di mercato**, che necessariamente implicano diritti di proprietà
- Open data confonde la conoscenza, come possibile bene pubblico, con i dati, di per sé sono escludibili e che – se non gestiti – non hanno alcun valore

- La risposta dei governi alla tensione fra considerare i dati come una merce e considerarli come bene pubblico è la legislazione sulla *privacy*,
- come nelle norme dell'Unione Europea incluse nella **GDPR**, che comunque non risolve **la contraddizione fra valore economico appropriabile, su cui si basa l'oligopolio digitale delle Tech Giants, e valore sociale che si distribuisce a valle**

- Più in generale alcuni ritengono che i grandi centri di archiviazione dei dati, come quelli di Amazon Web Services) siano una risposta inefficiente al diluvio di informazione codificata e che essa andrebbe elaborata il più vicino possibile a dove essa viene prodotta (***edge computing***)
- **Ma dove sono in effetti generati i dati?**

- Oggi si generano soprattutto nelle interazioni sui social media,
- attraverso la posta elettronica,
- attraverso prenotazioni e acquisti online,
- uso di carte di credito e POS,
- attraverso la semplice visita ad un website
- celle della telefonia mobile
- anche nelle interazioni con la pubblica amministrazione e i sistemi sanitari....

- Domani potrebbero essere generati soprattutto da macchine interconnesse fra di loro, ***internet-of-things***, che in sostanza consiste nel fatto che ogni singolo oggetto dotato di una scheda logica potrebbe generare segnali trasmessi, accumulati, analizzati e processati a distanza per ogni genere di attività
- **Si prevede che 25 miliardi di macchine saranno così interfacciate nel 2025, essenzialmente *wireless***
- **La tecnologia 5G potrebbe supportare, a quanto si dice, un milione di connessioni per km quadrato**

Prendiamo ad esempio il potenziale offerto dai **dati biomedici personali**. Se di un individuo si conoscesse il genoma completo,

- le patologie che gli sono state diagnosticate,
- i farmaci che assume,
- i parametri vitali in continuo
- il tempo dedicato all'attività fisica (ad esempio con un *device* da polso come Fitbit),
- ma *anche* il suo carrello della spesa al supermercato e acquisti in genere,
- i suoi spostamenti fisici e in rete,
- il suo *network* sociale,
- reddito dichiarato al fisco,

ne potrebbe derivare una profilazione molto più completa dei rischi di quella desumibile da banche dati non interfacciate

- Se questo scenario appare **inquietante in relazione al singolo individuo**,
- basta spostare l'unità di analisi ad **aggregati statistici** di riferimento, ammesso che ovviamente gli utenti che conferiscono i dati si fidino del gestore delle banche dati integrate,
- cosa che oggi avviene accettando con un *click* le condizioni a piccoli caratteri proposte dai gestori dei vari servizi

- Se le banche dati sono integrate, così possono esserlo i *software* di pulizia del 'rumore' di archiviazione, di interrogazione e analisi
- **In questo senso l'intelligenza artificiale tende al monopolio naturale, nel senso che il costo marginale di gestione del bit è sempre decrescente**
- Questa integrazione non è peraltro affatto banale, come sa chiunque abbia lavorato su grandi banche dati, come ad esempio ORBIS, dove sono accumulati dati su oltre 380 milioni di società con informazioni finanziarie dettagliate su 40 milioni di imprese, compresi oltre ai dati di bilancio anche informazioni sulla proprietà azionaria

- Solo un disegno ex-ante del data management (compreso spesso l'istituzione di un '**chief data officer**') che sin dall'origine stabilisca dei '**data warehouses**' o '**data lakes**' con le relative procedure integrate può alleviare i problemi di integrazione che spesso ex-post possono risultare intrattabili o costosi
- **Non si tratta di stabilire 'dove' i dati sono archiviati, se in grandi centri accessibili attraverso servizi di *cloud* o in archivi decentrati ma interconnessi, ma se fonti diverse di dati sono in grado di essere analizzate in modo unico con determinate procedure**

- I dati in molti casi peraltro non sono oggetti statici, ma **flussi** quasi continui in tempo reale
- le loro combinazioni non possono essere gestite con **procedure 'atemporal'**,
- un tema che riguarda le più varie applicazioni, dalla *supply chain* globale delle grandi imprese manifatturiere come l'automobile, al monitoraggio dei furti di energia per operatori elettrici come Enel, alla gestione dei contratti assicurativi

- Questo apre un mercato a imprese come Oracle o IBM che si propongono di vendere programmi di gestione di basi di dati costruiti con un approccio di ***datamining*** e **intelligenza artificiale** a chi non ha le risorse per costruire da sé questi programmi (cioè praticamente tutti, date le competenze professionali richieste sono scarse sul mercato del lavoro e contese a caro prezzo)

- Nel mercato entrano continuamente **start-up** che propongono nuovi programmi di AI,
- ma quali e quante raggiungeranno la massa critica per competere con gli oligopolisti nei vari segmenti non può essere previsto,
- saranno presumibilmente poche per l'operare in questo come altri campi di una certa **selezione darwiniana**,
- se una start up si presenta con idee commercialmente interessanti è probabile venga **acquisita da una delle imprese maggiori** con offerte troppo allettanti per essere rifiutate

- In questa prospettiva il dibattito fra *big data centres* e *edge computing* non appare definire una alternativa secca,
- **si possono immaginare sistemi integrati di archivio e trattamento dei dati senza che questi siano fisicamente concentrati:**
- se vi è una organizzazione unitaria che gestisce il processo di estrazione, pulizia, elaborazione, accesso, il **monopolio naturale** si applica all'organizzazione, non ad un suo luogo specifico

- Ad esempio non sembra affatto necessario che i dati generati dalla flotta di **auto *self-driving*** debba affluire in *server* remoti,
- ma ovviamente se i *server* locali sono connessi e interfacciati in modo integrato non fa molta differenza sotto il profilo del meccanismo economico che genera i costi marginali decrescenti

- Sotto il profilo distributivo: la formazione di un **proletariato della lavorazione dei dati**
- **Decine di milioni di proletari del data management e delle attività dell'IT in generale, persone con la vita inchiodata ad un video (come gli operatori dei *call center*), sono un problema sociale (The Economist)**

- **Ma miliardi di esseri umani espropriati dei loro dati per generare la ricchezza di pochi, in cambio di beni spesso di dubbio valore, sono una questione di altro ordine di grandezza**
- È il ritorno su scala planetaria ad un inaccettabile **capitalismo senza regole pubbliche**, che priva tutti della dignità personale, riducendoci a fornitori gratuiti di conoscenza sulle nostre vite in cambio di briciole

- Nella elencazione dei benefici dell'economia digitale, per quanto essi possano apparire grandi al singolo individuo ed alla singola impresa, manca l'identificazione dei **costi-opportunità**:
- che cosa si potrebbe fare, ed invece non si fa, con la scienza e tecnologia e dell'informazione per affrontare **le sfide irrisolte del nostro tempo?**
- **La declinazione concreta di obiettivi e strumenti potrebbe venire dalla costruzione di una infrastruttura pubblica che faccia propria la missione di un'agenda di ricerca, sviluppo e fornitura di servizi nell'interesse collettivo**

- **Un' agenda digitale declinata in obiettivi connessi a questi temi potrebbe innanzitutto considerare che oggi, per la prima volta nella storia umana, abbiamo la possibilità di investigare in modo esaustivo, sistematico e dinamico, la correlazione fra patrimonio genetico, ambiente naturale e condizioni socio-economiche di ciascuno di noi**

- I **problemi etici**, di *cybersecurity*, di crittazione di queste informazioni in un sistema di archivi digitali integrati sarebbero importanti, ma non insuperabili e non maggiori di quelli che già si pongono attualmente in modo scarsamente omogeneo fra giurisdizioni
- La partita dei Big Data e dell'intelligenza artificiale è troppo importante per non giocarla nel campo della salute con una chiara e preminente **missione pubblica**

- Esiste una lunga tradizione di raccolta di dati sullo **stato dell'atmosfera** su pochi parametri per le previsioni metereologiche, ed una più recente raccolta sistematica di dati sull'inquinamento dell'aria e dei corpi idrici, e su altro ancora
- Ma come per la salute umana vi è un grande **spreco di informazione**

- Singoli programmi di ricerca con durata tipicamente di diversi anni, ma non permanenti, raccolgono ed elaborano certi dati in alcuni contesti,
- manca una **infrastruttura internazionale coordinata di creazione del database sulla salute del pianeta e sulle sue determinanti**
- Nonostante una costosa flotta di **satelliti di osservazione terrestre** ruoti sulle nostre teste generando un flusso di dati preziosi, la maggior parte di questi dati non viene utilizzata perché non sembra avere valore commerciale immediato

- Lo stesso vale per le stazioni di rilevamento a terra dell'**inquinamento**,
- per le osservazioni sullo **stato dei mari e dei fiumi**,
- per il monitoraggio del **traffico veicolare** nelle strade
- e della produzione e consumo di **elettricità** in tempo reale,
- per l'impatto ambientale della **zootecnia**
- e di varie altre attività umane

- I dati vengono raccolti, ma poco e male utilizzati in archivi non integrati.
- **Il paradosso è che la comunità scientifica degli astronomi si pone il problema di come integrare i dati sull'osservazione dell'universo più di quanto i governi si pongano il problema di integrare i dati sul nostro pianeta**

- Come per la salute umana le tecnologie dell'informazione che lo consentirebbero sono a portata di mano, ma gli investitori finanziari non sono interessati. L'orizzonte è troppo lungo per loro.
- Un terzo elemento su cui stiamo sprecando vistosamente le opportunità sono proprio l'accesso universale all'informazione e alla conoscenza e l'interconnessione fra persone
- **Il sistema attuale delle telecomunicazioni e della *web economy* opera di fatto in condizioni di costi marginali decrescenti e spesso nulli**

- **Questo fatto, combinato con la configurazione di oligopoli orientati al profitto, comporta la necessità da parte delle imprese di estrarre le rendite sulla trasmissione ed accesso ai dati in modi inefficienti:**
- con la pubblicità, con canoni di abbonamento e discriminazione di prezzo e in molti altri modi, quando **la soluzione più efficiente sarebbe quella di offrire gratuitamente la comunicazione digitale come servizio pubblico di base**, così come – almeno in Europa- si offrono altri servizi a carico della fiscalità generale

- La persistenza persino in Europa, certamente in Italia come si è visto con la didattica a distanza, di una '**povertà digitale**' è inaccettabile dato il costo non insormontabile, e l'elevato beneficio, che avrebbe colmare i divari, come a suo tempo hanno fatto le imprese pubbliche del servizio elettrico

Iniziative della Commissione Europea e la proposta

- Se si scarta l'ipotesi di un sovranismo digitale, le forme di **regolazione sovranazionale**, come per la UE il GDPR o il Digital Services Act appaiono non risolutive (vedi slides della lezione precedente).
- Le azioni *anti-trust* sono lente ed incerte. Un esempio sono le indagini della CE e della Corte Suprema US sul potere di mercato di Apple con il sistema di pagamento AppStore, installato in esclusiva su ogni iPhone, da cui nel solo 2019 sono transitati 519 miliardi di dollari, con un consistente flusso di commissioni per la stessa Apple

- Forse alcuni Tech Giants potrebbero optare, come Facebook, per sistemi di auto-regolazione, sia con comitati e procedure di appello sui contenuti, sia con algoritmi che danno agli utenti un certo controllo sui dati
- La via maestra tuttavia sarebbe la creazione – quanto meno a livello europeo di una **infrastruttura pubblica in grado di competere con le Tech Giants e di offrire servizi che queste non sono al momento interessate ad offrire**

- Un sintomo embrionale di questa ambizione è il progetto **GAIA-X**, inizialmente proposto dalla Germania nel Summit Digitale di Dortmund (2019), ma cui successivamente ha aderito la Francia e cui si sono detti interessati altri paesi ed imprese, istituti di ricerca ed organizzazioni
- L'idea inizialmente pare fosse ambiziosamente quella di creare un “**Airbus Cloud**” sul modello del consorzio che ha effettivamente rotto la dominanza di mercato della Boeing nell'industria aeronautica

- Il programma **Digital Europe** è considerato complementare ad altri programmi della UE, in particolare a Horizon Europe per la ricerca e a Connecting Europe Facility per progetti di fibra ottica in aree periferiche ed altri investimenti
- **Nonostante questi segnali incoraggianti, difficilmente senza una vera infrastruttura pubblica sul modello delineato nei capitoli precedenti l'Europa recupererà il divario che si è creato con USA e Cina nella tecnologia dell'informazione e settori connessi**

- Occorrerebbe immaginare un soggetto sovranazionale europeo che non abbia solo funzioni di **coordinamento**,
- ma abbia **l'autonomia manageriale**,
- di **bilancio**,
- di **capitale tangibile e intangibile**,
- di **personale dedicato**
- Come per Green Europe, anche in questo caso si tratterebbe di **una combinazione di infrastruttura di ricerca e di impresa pubblica, o in breve di una infrastruttura pubblica**

- **L'agenzia sovranazionale** che gestisse questa infrastruttura dovrebbe certamente ricorrere a collaborazioni con terze parti, come nello schema di Gaia X, ma in una chiara distinzione di ruoli
- **Il segreto del successo del “modello di Ginevra”**, del CERN (ma anche di altre IR) consiste nel fatto che le collaborazioni scientifiche sperimentali con la loro struttura ampia e ‘federata’ hanno un punto di riferimento permanente nella infrastruttura scientifica, tecnologica, manageriale del CERN

- Lo stesso può dirsi per **EMBL-European Bioinformatics Institute**, che mentre assicura un servizio a centinaia di migliaia di utenti nel modo più flessibile che si possa concepire,
- è anche solidamente presidiato da una **propria infrastruttura fisica di server**,
- da proprio **personale e risorse** derivanti dall'essere parte di EMBL, un organismo sovranazionale,
- non è una **federazione** di istituti partecipanti

- Come per Green Europa, non si tratta di immaginare un sistema chiuso e autoreferenziale, ma al contrario aperto a ***partnership*** con organizzazioni pubbliche e private già esistenti, prioritariamente negli Stati membri, ma costruendole a partire da risorse adeguate comparabili con i budget annui di ricerca e sviluppo dei grandi attori in campo
- La scala di un progetto Digital Europa dovrebbe essere quella di alcuni miliardi di Euro all'anno in capo alla nuova agenzia (molto maggiore di quanto potrebbe essere stanziato per la *digital economy* in Next Generation EU, peraltro in modalità piuttosto dispersive)

La scala che servirebbe?

The World's Biggest R&D Spenders

Companies with the highest R&D investment expenditure in 2020 (in billion U.S. dollars)



* formerly Facebook (until Oct 2021)

Sources: Nasdaq.com, corporate reports



- **Serve reclutare le persone giuste**
- **finanziare i progetti,**
- **investire nelle attrezzature sia in un campus che funga *hub* che in sedi decentrate**
- **i finanziamenti pubblici iniziali e periodici potrebbero rientrare in cassa in larga misura direttamente attraverso una politica di equa tariffazione di alcuni tipi di servizi**
- **e di gestione nell'interesse pubblico di eventuali brevetti**