

PER UNA TEORIA DEL VALORE-RETE: BIG DATA E PROCESSI DI SUSSUNZIONE[♥]

di Andrea Fumagalli

Saggio tratto da: D. Gambetta (a cura di), Datacrazia, Politica, cultura algoritmica e conflitti al tempo dei big data, DEditore, Ladispoli (Roma, 2019, pp. 46-69

1. Introduzione

Negli ultimi anni, abbiamo assistito ad una notevole accelerazione tecnologica. Diversi settori sono stati contaminati. E si tratta di settori che hanno sempre più a che fare con la gestione (governance) della vita umana. La lettura del genoma umano, a partire dal 2003, ha aperto spazi enormi nella possibilità di manipolazione della vita individuale e della sua procreazione. Gli effetti sulla medicina, a partire dall'utilizzo delle cellule staminali, sono stati assai importanti. Così come la tavola periodica degli elementi naturali di Mendeleev del 1869 ha aperto la strada alla creazione di materiali artificiali che hanno profondamente inciso sulle innovazioni di processo e di prodotto del periodo fordista, oggi la decrittazione del genoma umano apre alla possibilità di creare artificialmente tessuto umano e combinarlo con elementi macchinici, altrettanto artificiali. Siamo così di fronte alla nascita di una nuova tecnologia biopolitica, ovvero di una "bio-tecnica".

Contemporaneamente, lo sviluppo degli algoritmi di II generazione stanno consentendo un processo di automazione senza precedenti nella storia umana. Applicati, tramite le tecnologie informatiche e le nanotecnologie, alle macchine utensili, essi sono in grado di trasformarle in strumenti sempre più flessibili e duttili, sino ad assimilarsi alle potenzialità umane nella capacità sensoriale. Tali tecnologie sono soprattutto applicate (e spesso nascono o ne vedono la sperimentazione) nei settori militari e nei luoghi bellici (ad esempio, i droni) ma sempre più anche nei settori legati alla logistica internazionale, dalla gestione di un magazzino, al trasporto, alla grande distribuzione. Lo sviluppo dell'Intelligenza artificiale e della robotica ha oramai sperimentato con successo lo sviluppo di magazzini e sistemi di pagamento del tutto automatizzati, con il rischio che nei prossimi anni si debba fronteggiare una crescente disoccupazione tecnologica.

[♥] Ringrazio Giorgio Griziotti e Tiziana Sacconi per una prima lettura e per i miglioramenti suggeriti. Per il supporto psichedelico, mi sono avvalso, come al solito, della musica dei Grateful Dead, Jimi Hendrix e The Phish.

Se il paradigma tecnologico dell'ICT ha duramente colpito i livelli occupazionali dell'industria manifatturiera, la nuova ondata bio-tecnologica rischia di avere effetti ancor più pesanti sui settori del terziario tradizionale e avanzato, che nelle ultime decadi aveva svolto un ruolo di compensazione all'emorragia di posti di lavoro nell'industria tradizionale.

La bio-tecnica attuale, tuttavia, non sarebbe possibile se non si fosse sviluppata una serie di tecniche che hanno accelerato (rispetto al recente passato) il grado di raccolta e manipolazione di dati in grandissima quantità in spazi sempre più ristretti e con velocità sempre più elevata.

Condizione necessaria (anche se non sufficiente) perchè un algoritmo sia in grado di essere sfruttato alla massima potenza è l'esistenza di un processo di standardizzazione catalogazione dei dati necessari al suo funzionamento in relazione allo scopo prefissato. Ciò è reso possibile dalle tecniche di manipolazione dei cd. "big data" e da i processi di "data mining" (estrazione dei dati).

In questo saggio, cercheremo di analizzare tale fenomeno come fonte di valore economico ad uso e consumo del sistema delle grandi corporation con brevi cenni anche alle forme di organizzazione del lavoro che vi sottostanno.

In ultima analisi, cercheremo di proporre una teoria del valore dei big data.

2. Verso una teoria del valore-rete: il quadro di partenza

L'utilizzo e la raccolta dei dati ha sempre fatto parte della storia umana sin dai suoi albori. Ma è solo con l'inizio della rivoluzione industriale che le tecniche di calcolo, affinate dalla rottura "metodologica" iniziata da Cartesio e Galileo, cominciano ad essere applicate non più solo alla necessità di "misurare" nel campo fisico-naturale (esigenza, che, come è noto, è stata alla base dello sviluppo della geometria e della matematica già nell'antichità, dagli egiziani, ai greci, agli arabi) ma direttamente al controllo e alla gestione dell'attività di produzione. Nel momento stesso in cui, con l'avvento del sistema capitalistico di produzione, assistiamo all'irrompere della "macchina" come fattore produttivo immediato, l'atto della produzione (finalizzata all'accumulazione) diventa sempre più discrezionale, sganciato dai capricci della natura, e quindi necessita, appunto, di una o più unità di misura.

Il (plus)valore prodotto dall'accumulazione capitalistica necessita, infatti, di essere noto per poter poi influenzarne la distribuzione.

Finché la produzione capitalistica è stata prevalentemente materiale, sia nel capitalismo artigianale dell'Ottocento che nel periodo taylorista del Novecento, le unità di misure

convenzionalmente fissate per la misura della natura (metro, chilo, litro, volt, watt, cavalli vapore, numerazione decimale, ecc.) erano più che sufficienti.

Quando invece, con la crisi del paradigma fordista, la produzione tende a diventare sempre più immateriale e il capitale sempre più intangibile, il problema della misura acquista una dimensione che va al di là delle tradizionali geografie naturali. Le stesse fonti di valorizzazione si modificano e l'innovazione tecnologica, fondata ieri sull'ICT oggi sulle bio-tecniche, richiede un approccio del tutto nuovo.

A partire dalla diffusione dell'informatica, la velocità di calcolo è esponenzialmente aumentata. Il volume dei dati creati ha richiesto, non a caso, nuove forme di misurazione, continuamente in fase di ridefinizione, perché velocemente diventano obsolete.

Se inizialmente le tecniche di *data-mining* erano la sofisticata evoluzione di tecniche di calcolo statistico (e ancora oggi vengono studiate in questa prospettiva impolitica e neutrale¹) oggi sono sempre più fortemente interrelate alle caratteristiche personali, in grado di definire raccolte differenziate (individualizzate) di dati da commerciare poi liberamente.

Un noto esempio, su cui si è soffermato Matteo Pasquinelli, riguarda l'algoritmo *Pagerank* di Google². Ecco come tale algoritmo viene descritto da Nicholas Carr³:

“Al cuore di [Google] si trova l'algoritmo PageRank che Brin e Page scrissero mentre erano studenti a Stanford negli anni 90. Notarono che ogni volta qualcuno con un sito Web pone un link ad un altro sito, costui esprime un giudizio, dichiara che considera quel sito importante. In seguito realizzarono che mentre ogni link contiene un po' di intelligenza umana, tutti i link insieme contengono una grande quantità di intelligenza – molto di più, in effetti, di quanto qualsiasi singola mente possa possedere. Il motore di ricerca di Google scava in questa intelligenza, link dopo link, e la usa per determinare l'importanza di tutte le pagine del Web. Più grande è il numero di link che porta ad una pagina, più grande è il suo valore. Come dice John Markoff, il software di Google “sfrutta sistematicamente la conoscenza e le scelte umane su ciò che è significativo”. Ogni volta che scriviamo un link, o anche solo lo clicchiamo, alimentiamo il sistema di Google con la nostra intelligenza. Rendiamo la macchina un po' più intelligente – e Brin, Page e tutti gli azionisti di Google un po' più ricchi”⁴.

¹ Ad esempio, Susi Dulli, Sara Furini, Edmondo Peron, *Data Mining*, Springer Verlag, Roma-Berlino, 2009 e Paolo Giudici, *Data Mining*, Mc Graw – Hill, Milano, 2005. Per data-mining si intende “l'insieme di tecniche e metodologie che hanno per oggetto l'estrazione di un sapere o di una conoscenza a partire da grandi quantità di dati (attraverso metodi automatici o semi-automatici) e l'utilizzo scientifico, industriale o operativo di questo sapere”: https://it.wikipedia.org/wiki/Data_mining

² Matteo Pasquinelli, “L'algoritmo PageRank di Google: diagramma del capitalismo cognitivo e rentier dell'intelletto comune” in F. Chicchi and G. Roggero (a cura di), *Sociologia del lavoro*, Milano: Franco Angeli, 2009.

³ Nicholas Carr, *The Big Switch: Rewiring the World, from Edison to Google*, New York: W.W. Norton, 2008 [trad. it. di Matteo Vegetti: *Il lato oscuro della rete. Libertà, sicurezza, privacy*, Milano: Etas, 2008].

⁴ Brano tradotto da Matteo Pasquinelli.

L'algoritmo, oggi, si sta affermando come l'espressione del *general intellect*, è la sua espressione fenomenologica. Una espressione varia e flessibile a seconda dell'ambito di riferimento.

Non riguarda direttamente il *bios* ma il *cognitivo*.⁵ E' oggi lo strumento per misurare il valore dell'intensità cognitiva. E' allo stesso tempo sussunzione reale e formale. Ma è anche qualcosa di più. E' misura matematica del *valore di rete*, in grado di condensare il *wetware* e il *netware* sulla base di un *software*. E' quindi base di accumulazione e valorizzazione.

“Ciò che PageRank identifica e misura è un valore di rete in una forma molto numerica. Se una merce è descritta tradizionalmente da un valor d'uso e da un valore di scambio, il valore di rete può essere considerato un ulteriore livello annesso ai precedenti che ne descrive la rete delle relazioni sociali. Questo termine può risultare alquanto ambiguo in quanto può essere frainteso come 'valore delle reti' secondo la tanto celebrata 'ricchezza delle reti' descritta da Yochai Benkler⁶. Al contrario, una nozione di plusvalore di rete dovrebbe essere qui introdotta per fare maggiore chiarezza⁷. In effetti, PageRank produce quello che Deleuze e Guattari descrissero come plusvalore macchinico riferendosi al plusvalore accumulato attraverso il dominio cibernetico, ovvero la trasformazione di un plusvalore di codice in plusvalore di flusso⁸. Attraverso PageRank, Google non ha semplicemente conquistato una posizione dominante nel controllo e possesso di estesi indici del web, ma soprattutto un monopolio nella produzione di tale *valore di rete*”.⁹

L'esempio citato è paradigmatico di un'evoluzione dei processi di valorizzazione contemporanea che, partendo dal cognitivo, hanno sempre più pervaso il *bios*, al punto che l'evoluzione tra uomo e macchina tende a diversificarsi sempre più lungo due direttrici parallele e sinergiche: il rapporto tra *soggettività e macchina* e quello tra *corpo fisico e macchina*.

Del primo si è scritto molto, a partire dai primi anni Settanta quando si è indagato il rapporto tra mente e macchina. E su tale ibridazione che Franco Berardi conia nei primi anni 2000 il termine *cognitariato*¹⁰. La definizione del termine fornito dal dizionario Garzanti (“preariato di chi svolge un lavoro di tipo intellettuale”¹¹) non coglie la

⁵ Per approfondimenti si rimanda a A. Fumagalli, *L'economia politica del comune*, DeriveApprodi, Roma, 2017

⁶ Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven: Yale University Press, 2006 [versione italiana a cura di Franco Carlini, *La ricchezza della rete. La produzione sociale trasforma il mercato e aumenta le libertà*, EGEA Università Bocconi, Milano, 2007]

⁷ Questo valore di rete dovrebbe essere distinto dalla definizione scientifica: secondo la legge di Metcalfe il “valore” di ogni network di telecomunicazioni è direttamente proporzionale al quadrato del numero di nodi o utenti connessi al sistema (n^2).

⁸ Gilles Deleuze e Felix Guattari, *L'Anti-Oedipe. Capitalisme et schizophrénie*, Paris: Minuit, 1972.

⁹ Matteo Pasquinelli, “L'algoritmo PageRank di Google: diagramma del capitalismo cognitivo e rentier dell'intelletto comune” in F. Chicchi and G. Roggero (eds), *Sociologia del lavoro*, Milano: Franco Angeli, 2009.

¹⁰ Franco Berardi Bifo, *La fabbrica dell'infelicità. New economy e movimento del cognitariato*, DeriveApprodi, Roma, 2002 e *Il sapiente, il mercante, il guerriero. Dal rifiuto del lavoro all'emergere del cognitariato*, Derive Approdi, Roma, 2004

¹¹ <http://www.garzantilinguistica.it/ricerca/?q=cognitariato>

complessità che ne consegue. E' infatti il concetto di lavoro intellettuale che viene messo in discussione. Se nell'ultimo decennio del secolo scorso, possiamo assistere ad una sorta di "taylorizzazione del lavoro intellettuale e di intellettualizzazione del lavoro manuale"¹², oggi tale processo è andato ben oltre la dicotomia, seppur ridefinita, tra manualità e intellettualità, sino a rendere superata tale differenza. Una differenza che oggi è stata ricompresa all'interno del termine "lavoro cognitivo" e ampliata in quello di "lavoro relazionale".

E' infatti da tale prestazione lavorativa che ha origine il *valore di rete*, che oggi tende a pervadere, in modo differenziato e diversificato, diverse attività produttive, da quelle di magazzinaggio (sempre più digitalizzate), a quello dei grandi centri commerciali sino ai settori della consulenza immateriale. Ovunque c'è un app, c'è *valore di rete*, cioè valore biopolitico.

A fronte delle recenti evoluzioni, ritorna di estrema attualità, con riferimento al contesto taylorista, l'osservazione di Romano Alquati riguardo l'*informazione valorizzante* ai tempi della fabbrica olivettiana:

"il lavoro produttivo si definisce nella qualità delle informazioni elaborate dall'operaio ai mezzi di produzione, con la mediazione del capitale costante"¹³.

Il *valore di rete* è allo stesso tempo esito di un processo di sfruttamento, di estrazione e di imprinting. E' la forma di plusvalore del *cognitivo*, a cui bisognerà aggiungere il plusvalore del *bios*.

Esso è dato dalla compenetrazione del sistema sensoriale umano (da quello percettivo dei 5 o più sensi a quello nervoso) con la rete informazionale e digitale che sempre più avvolge l'attività di produzione e accumulazione.

Da questo punto di vista, assistiamo al divenire macchinico dell'umano¹⁴, al divenire spaziale (ovvero relazionale) dell'umano¹⁵, ma allo stesso tempo al divenire umano delle macchine¹⁶.

¹² Andrea Fumagalli, *Bioeconomia e capitalismo cognitivo*, Carocci, Roma, 2007

¹³ Romano Alquati, "Composizione organica del capitale e forza-lavoro alla Olivetti", in *Quaderni Rossi*, n. 2, 1963, pag. 121, citato da Matteo Pasquinelli, "Italian Operaismo and the Information Machine", *Theory, Culture & Society*, 2015, Vol. 32(3) [vers. italiana: Matteo Pasquinelli, "Capitalismo macchinico e plusvalore di rete: note sull'economia politica della macchina di Turing": <http://www.uninomade.org/capitalismo-macchinico/>, novembre 2011]. Si veda anche Giorgio Griziotti, *Neurocapitalismo, Mediazioni tecnologiche e vie di fuga*, Mimesis, Milano, 2016, pag. 60.

¹⁴ Rosi Braidotti, *Il postumano. La vita oltre l'individuo, oltre la specie, oltre la morte*, DeriveApprodi, Roma, 2014

¹⁵ Matteo Pasquinelli "Oltre le rovine della Città Creativa: la fabbrica della cultura e il sabotaggio della rendita", in Marco Baravalle (a cura di) *L'arte della sovversione. Multiversity: pratiche artistiche contemporanee e attivismo politico*, Roma: Manifestolibri, 2009. Scrive Pasquinelli: "Il *general intellect* si presenta quindi non solo 'cristallizzato' nelle macchine ma diffuso attraverso l'intera 'fabbrica società' della metropoli. Quindi, logicamente, se la conoscenza industriale progettava e operava macchine, anche la conoscenza collettiva al di fuori della fabbrica deve essere in qualche modo macchinica. Qui dobbiamo

* * * * *

Nel processo di data-mining, i dati raccolti non sono fini a se stessi ma permettono di cogliere delle opportunità e di interpretare i fenomeni. Si tratta di processi che vanno oltre la pura constatazione e osservazione di fatti, in quanto il fine è quello di mettere in atto azioni traducibili nella creazione di valore.

E' per questo che la letteratura manageriale mainstream ha cercato di individuare i parametri che meglio consentono di catalogare i dati raccolti e di poterne individuare una misura. Si tratta eminentemente di quattro misure (dette anche le quattro V)¹⁷:

a. Volume (V1). E' la misura più facile da intuire in quanto riguarda l'accumulo dei dati.

“Ogni minuto vengono caricate sulla piattaforma di sharing Youtube 300 ore di video, il social network Facebook genera 4 petabyte di dati al giorno, il social network Twitter vanta oltre 500 milioni di Tweet ogni 24 ore, il numero di mail scambiate al giorno su scala mondiale si attesta su oltre 200 miliardi di unità”¹⁸.

“Non è facile rendersi conto di cosa significa in concreto questa grande mole di dati. Se fossero tutti racchiusi in libri cartacei, coprirebbero l'intera superficie degli Stati Uniti 52 volte.”¹⁹

b. Velocità (V2): Si riferisce alla necessità di ridurre al minimo i tempi di analisi dei dati, molte volte effettuate in *real time* o quasi. Inoltre alcuni dati potrebbero diventare obsoleti in tempi molto brevi, perciò è necessario che l'estrazione delle informazioni avvenga velocemente. La rapida elaborazione ed archiviazione può consentire di ottenere un vantaggio competitivo rispetto alla concorrenza.

guardare con attenzione alle manifestazioni del *general intellect* attraverso la metropoli per capire quando lo incontriamo 'morto' o 'vivo', già 'fissato' o potenzialmente autonomo. Per esempio, a quale livello oggi il tanto celebrato Free Software e la cosiddetta free culture sono complice delle nuove forme di accumulazione del capitalismo digitale? E a quale livello, l'ideologia della creatività e le Città Creative preparano semplicemente il terreno alla speculazione immobiliare e a nuove forme di rendita metropolitana?”

¹⁶ Gerard Raunig, *A Thousand Machines: A Concise Philosophy of the Machine as Social Movement*. Semiotexte, New York, 2010 e il più recente *Dividum: machinic capitalism and molecular devolution*, Semiotexte, New York, 2016 e A. Fumagalli, *L'economia politica del comune*, DeriveApprodi, Roma, 2017

¹⁷ Per approfondimenti, si veda R. Scaccianoce, *Big Data e creazione di valore: il caso Carrefour*, Tesi di Laurea, Università di Pavia, 2017. Si veda anche Ippolita, *Tecnologie di dominio, Lessico minimo di autodifesa digitale*, Meltemi, Milano, 2017, in particolare la voce “Big Data”, p. 27.

¹⁸ Thomas H. Davenport, *Big data @l lavoro. Sfatate miti, scoprire le opportunità*, F. Angeli, Milano, 2015, pag.11

¹⁹V. Mayer Schoenberger, K. Cukier, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, trad. it. R. Merlini, Garzanti, Milano, 2013

c. Varietà (V3): è una delle caratteristiche più importanti, in quanto introduce alcuni elementi di valore e complessità che possono incidere sul valore di rete generato. Ad un livello di base, la varietà può essere suddivisa in tre classi:

Dati strutturati: si tratta di dati organizzati in tabelle e in schemi rigidi, raccolti all'interno di un database e che hanno una lunghezza definita e un formato. Questa tipologia viene utilizzata per i servizi tradizionali di *Business Intelligence* e nelle moderne piattaforme di *Analytics*. I dati strutturati costituiscono solo il 20% di tutti quelli esistenti.

Dati non strutturati: sono i dati conservati senza alcuno schema, composti da un elevato numero di metadati, ossia informazioni che specificano il contenuto e il contesto di una pagina web. Si tratta di file audio, video o linguaggio umano. In questo caso vengono utilizzati sistemi di *information retrieval*, insieme di tecniche finalizzate a gestire la rappresentazione, l'organizzazione e l'accesso ad oggetti contenenti informazioni quali documenti, pagine web e cataloghi online.

Dati semi-strutturati: in essi si incontrano le caratteristiche delle prime due classi. Questa tipologia non è conforme alla struttura formale dei modelli di dato associato con le basi di dati relazionali o altre forme di tabelle di dati, ma comunque contiene etichette o altri marcatori per separare gli elementi semantici. Un esempio sono i dati in XML. Insieme ai dati *non strutturati* essi rappresentano l'80% dei dati esistenti.

Il livello di varietà di dati da cui trarre informazioni valorizzanti è quindi molto elevato. Al riguardo possiamo parlare di una sorta di *divisione cognitiva dei dati*. Tale divisione non è solo verticale ma si delinea anche sul piano dell'orizzontalità degli stessi dati, a seconda se sono prodotti direttamente dalle macchine (grazie all'evoluzione tecnologica) (*Computer or machine generated*)²⁰ o dall'interazione con l'intervento umano (*Human generated*)²¹.

²⁰ I dati strutturati generati dalle macchine includono:

Sensor data: esempi sono la tecnologia RFID (Radio Frequency Identification) basata sulla propagazione nell'aria di onde elettromagnetiche che consentono la rilevazione a distanza di oggetti, animali e persone; i dispositivi medici e i dati del GPS (Global Positioning System), che se analizzati permettono di capire i comportamenti dei consumatori.

Web Data: sono i dati catturati dai server e dalla rete.

Dati finanziari: dati generati dai sistemi di trading.

²¹ I dati strutturati generati dall'uomo includono invece:

Input data: sono i dati che possono essere immessi nel computer da qualsiasi persona, riguardo nome, età, reddito ecc.

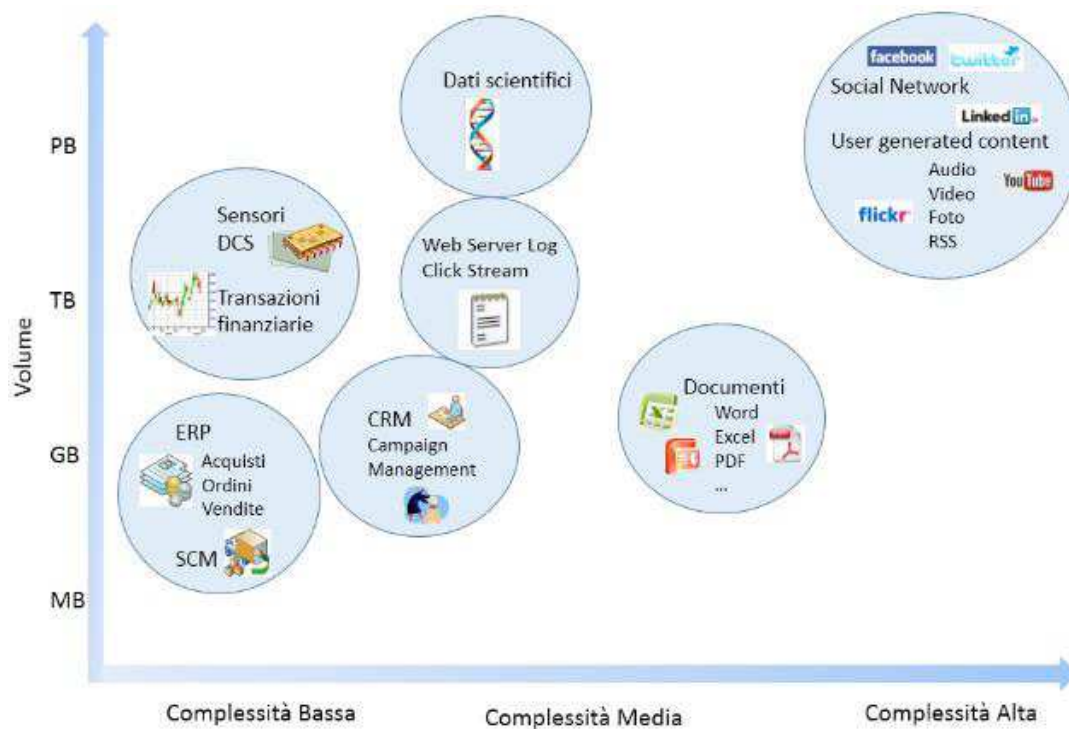
Click-stream data: dati generati da qualsiasi click da parte degli utenti. Questo tipo di dati permette di analizzare il comportamento del cliente e di realizzare nuovi modelli di acquisto.

Game related data: altri tipi di dati possono provenire dai giochi, all'interno dei quali ogni mossa dell'utente viene registrata.

A tali dati, occorre poi aggiungere quelli derivanti dai processi di automazione di varie attività umane (ad esempio, le immagini satellitari, i dati scientifici e le fotografie e video degli strumenti di sicurezza e sorveglianza) e quelli invece generati dall'attività relazionale umana (*Social media data*: ad esempio Facebook, Youtube, Twitter, LinkedIn, Flickr., *Mobile data*: ad esempio i messaggi di testo o le informazioni sulla posizione geografica e, più in generale, i dati *Contenuti dei siti web*: ad esempio, Youtube o da Instagram.²²

Tale tassonomia può essere riassunta nel seguente grafico, dalla quale si evince che i dati più difficili da analizzare sono quelli non strutturati provenienti dalle piattaforme di social media. E non è un caso che tali dati siano quelli a più alto valore aggiunto.

Fig. 1:



Fonte: <http://www.dataskills.it/caratteristiche-dei-big-data/>

d. Veridicità (veracity) (V4): indica il grado di accuratezza e di attendibilità dei dati. E' condizione necessaria (anche se non sufficiente) per poter estrarre valore dai dati. E'

²² Per approfondimenti, si veda J. Hurwitz, A. Nugent, F. Halper, M. Kaufman, *Big Data for Dummies*. John Wiley&Sons, Hoboken, New Jersey, 2013.

anche il parametro che definisce una sorta di divisione cognitiva dei dati, a seconda del loro grado di replicabilità più o meno automatizzata. I dati sono conoscenza e come la conoscenza, è possibile distinguere tra conoscenza (dati) standardizzata e conoscenza (dati) tacita. Tale distinzione opera all'interno di un ciclo di vita dei dati. Più un dato è, nella sua specificità, attendibile e accurato (in quanto in grado di fornire informazioni più dettagliate), tanto più tale dato è sottoposto a diritti di proprietà intellettuale che ne limitano la diffusione. Un nuovo dato è per definizione fonte di conoscenza tacita che solo successivamente può socializzarsi e divenire patrimonio comune, tramite un processo di standardizzazione nella sua diffusione, sino a diventare prima o poi obsoleto, nel momento in cui tale dato viene sostituito da nuovi dati più accurati. Maggiore è l'accuratezza e la specificità del dato, più le sue fonti sono segrete e più acquista potenzialmente valore.

I quattro parametri, velocemente descritti, sono legati da processi di interdipendenza. Ad esempio, la varietà implica maggior accuratezza e specificità ma minor volume. Maggiore è la velocità computazionale di elaborazione di un insieme di dati, maggiore è la versatilità e la varietà del suo utilizzo ma, allo stesso tempo, minore è la sua durata nel tempo e più precoce la sua possibile obsolescenza.

La società di consulenza Gartner, proprio partendo da questi parametri, definisce i Big Data nel seguente modo:

“I Big Data sono un patrimonio informativo caratterizzato da velocità, volume e variabilità elevati, che richiede forme innovative di analisi e gestione finalizzate a ottenere una migliore comprensione nei processi decisionali”.²³

Tale definizione risulta essere più convincente di tante altre²⁴ in quanto unisce al suo interno sia le caratteristiche tecniche sia la necessità di nuove tecnologie di elaborazione e gestione diverse da quelle tradizionali.

Lo scopo è infatti quello di creare “valore di rete”. Le quattro V, precedentemente descritte, infatti, confluiscono e convergono nel definire una quinta V, quella più importante: il valore di scambio che ne deriva.

A tal fine si attua un processo di trasformazione (immateriale) dal dato come valore d'uso al dato come valore di scambio.

²³ L. Camiciotti, C. Racca, *Creare valore con i Big Data. Gli strumenti, i processi, le applicazioni*, Edizioni LSWR, Milano, 2015

²⁴ Ad esempio IBM definisce i Big Data “come un set di dati la cui dimensione va al di là della capacità dei tipici strumenti di gestione, memorizzazione e analisi,” mentre McKinsey considera i Big Data come “dati che possono essere riuniti e analizzati per individuare nuovi modelli e prendere decisioni migliori” (citazioni tratte da L. Camiciotti, C. Racca, op. cit.)

3. Verso una teoria del valore-data: il processo di trasformazione e le fonti di valorizzazione

La creazione del valore di rete, tramite l'elaborazione dei Big Data, avviene prevalentemente in alcuni settori. Il dato, in sé e per sé, è caratterizzato da valore d'uso, come la forza-lavoro o il comune (al singolare)²⁵. In qualità di input produttivo in un contesto di produzione immateriale si trasforma in valore di scambio, all'interno di contesti di produzione in grado di utilizzare la tecnologia algoritmica appropriata.

Tale processo tuttavia è lungi dall'essere omogeneo e preciso. Di fatto nella gestione delle *clouds* di Big Data regna la confusione, l'approssimazione e l'eterogeneità, anche per l'imperfezione delle tecnologie utilizzate, continuamente in fase di miglioramento, anche grazie al coinvolgimento degli stessi fornitori e utenti di dati.

Si consideri ad esempio il servizio di traduzione offerto da Google: la differenza rispetto agli altri sistemi riguarda proprio l'utilizzo di un dataset più vasto e molto caotico. Non si tratta di una traduzione parola per parola, ma di una analisi di milioni di testi ufficiali che arrivano da fonti come l'Onu e che forniscono una grande quantità di dati:

«Nonostante la confusione dell'input il sistema di Google funziona meglio. Le sue traduzioni sono più accurate di quelle offerte da altri sistemi. Ed è molto, molto più ricco. A metà del 2012 il suo dataset copriva oltre 60 lingue. Era persino in grado di accettare input vocali in 14 lingue per rendere più fluide le traduzioni. E siccome tratta il linguaggio semplicemente come una massa caotica di dati a cui applicare il calcolo delle probabilità, può persino tradurre tra due lingue come l'hindi e il catalano».²⁶

L'esattezza quindi passa in secondo piano nel momento in cui la vastità permette di identificare il trend generale e di catturare l'intero fenomeno. La tassonomia a cui si è da sempre abituati viene rimpiazzata da meccanismi più disordinati e flessibili come il *tagging*, ampiamente utilizzato su Internet. Questo sistema permette di etichettare principalmente foto o video, in modo da rendere rintracciabili i numerosi contenuti presenti in rete, attraverso tag creati dagli utenti. La confusione in questo caso può essere dovuta alla scrittura errata dei tag e nel modo in cui essi sono organizzati. Tutto questo però permette di avere un accesso più ampio ai vari dati con un universo più ricco di etichette.

L'eterogeneità dei dati e le possibili varianti dei big data possono essere evidenziati in questa tabella

²⁵ Vedi C. Vercellone, A. Giuliani, F. Brancaccio, P. Vattimo (a cura di), *Il comune come modo di produzione*, Ombre Corte, Verona, 2017 e A. Fumagalli, *L'economia politica del comune*, DeriveApprodi, Roma, 2017.

²⁶V. Mayer Schoenberger, K. Cukier, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, trad. it. R. Merlini, Garzanti, Milano, 2013

Tab. 1: Possibili varianti di big data

Stile dei dati	Fonte dei dati	Settore interessato	Funzione interessata
Grande volume	Online	Servizi finanziari	Marketing
Non strutturati	Video	Sanità	Logistica
Flusso continuo	Sensori	Manifattura	Risorse umane
Formati multipli	Genomica	Turismo/trasporti	Finanza

Fonte: T.H. Davenport, *Big data @l lavoro. Sfatate miti, scoprire le opportunità*, FrancoAngeli, Milano, 2015

Tale ripartizione, opinabile come ogni tassonomia, pone alla ribalta un settore trasversale che possiamo definire *Business Intelligence*. Si tratta di una funzione aziendale che ha la funzione di estrarre valore per le diverse finalità produttive dall'elaborazione e dalla ripartizione dei dati.

Essa si riferisce

“all’insieme di processi aziendali per raccogliere dati e analizzare informazioni strategiche, alla tecnologia utilizzata per realizzare questi processi e alle informazioni ottenute come risultato di questi processi”.²⁷

La Business Intelligence quindi è un sistema di modelli, metodi, processi, persone e strumenti che rendono possibile la raccolta regolare e organizzata del patrimonio dati generato da un'azienda e attraverso elaborazioni, analisi ed aggregazioni, ne permette la trasformazione in informazioni, la conservazione, la reperibilità e la presentazione in forma semplice, flessibile ed efficace tale da costituire un supporto alle decisioni strategiche, tattiche ed operative.

Il sistema di *Business Intelligence* comporta dunque:

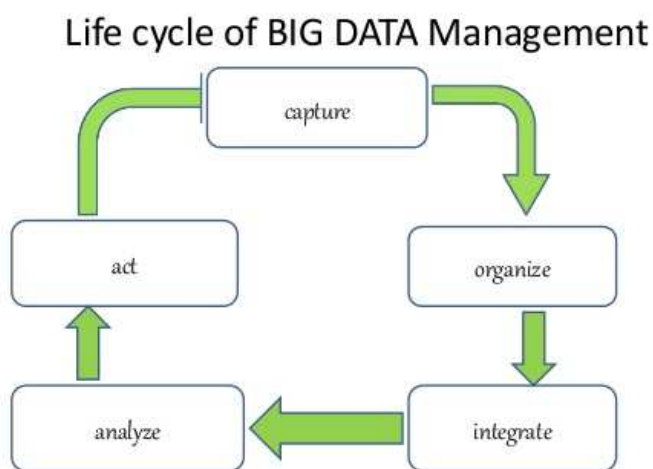
- la raccolta dei dati del patrimonio dell'azienda
- la loro pulizia, validazione e integrazione
- la successiva elaborazione, aggregazione e analisi
- l'utilizzo fondamentale di questa mole di informazioni nei processi strategici e di valorizzazione.²⁸

²⁷L. Camiciotti, C Racca, *Creare valore con i Big Data*, op. cit. .

²⁸“Cos’è la Business Intelligence”, Dataskills, Understanding the world. <http://www.dataskills.it/business-intelligence/cos-e-business-intelligence/>.

Si struttura in tal modo un vero e proprio ciclo di vita e di valorizzazione del sistema dei big data che può essere descritto nella figura seguente, sulla base di una successione di operazioni che iniziano con la “cattura/appropriazione dei dati”, la loro “organizzazione”, “integrazione”, “analisi”, “azione”.

Nella stragrande maggioranza dei casi, soprattutto per quanto riguarda i dati non strutturati (circa l’80% del totale), tali dati si presentano come valore d’uso, prodotti e socializzati dagli utenti/consumatori nell’espletazione degli atti di cooperazione e relazione che vengono svolti quotidianamente. Non a caso si parla di cattura o meglio di espropriazione più o meno forzata o volontaria.



Fonte: <https://www.slideshare.net/TakrimulLaskar/big-data1st-presentation>

Tale ciclo di vita descrive, in estrema sintesi, il processo di valorizzazione dei big data. Vale la pena soffermarsi sulle due operazioni di “organizzare” e “integrare”. Si tratta di due operazioni che solo negli anni recenti hanno potuto raggiungere un certo grado di sofisticazione, grazie all’evoluzione tecnologica degli algoritmi di II° generazione. L’organizzazione e l’integrazione dei dati è alla base della produzione del valore di rete. E’ l’aspetto produttivo di valore di scambio, mentre l’“analisi” e l’“azione” ne rappresentano la commercializzazione, ovvero la realizzazione monetaria sui mercati di sbocco.

E’ infatti in queste due fasi che comincia a strutturarsi il “capitalismo delle piattaforme”²⁹, Con il termine capitalismo delle piattaforme, facciamo riferimento alla capacità da parte delle imprese di definire una nuova composizione del capitale in grado di gestire in modo sempre più automatizzato un *processo di divisione di dati* in funzione dell’utilizzo commerciale che ne può derivare. Esso si basa sulla partecipazione, più o

²⁹ B. Vecchi, *Il capitalismo delle piattaforme*, Manifestolibri, Roma, 2017.

meno consapevole, dei singoli utenti, trasformati ora in *prosumer*. Sono infatti gli utenti delle diverse piattaforme, siano esse finalizzate a fornire informazioni per soddisfare i propri desideri o spazi virtuali per la comunicazione, il gioco e lo sviluppo di relazioni, a fornire la materia prima che viene poi sussunta nell'organizzazione capitalistica produttiva.

Possiamo affermare che se oggi le relazioni umane, la cooperazione sociale, la produzione di intelligenza collettiva, la riproduzione sociale sono espressione di comune come modo di produzione³⁰, al momento attuale esse sono alla base del comunismo del capitale, ovvero della capacità del capitale di sussumere e catturare quelle che sono le istanze di vita degli esseri umani³¹.

Lo strumento principale di questa capacità di catturare il comune umano è costituito dalle *machine learning*.

Fino a qualche tempo fa erano in pochissimi a sapere cosa si intendesse per *algoritmi*, mentre oggi essi sono presenti nella quotidianità come un aspetto fondamentale della società odierna:

“Non sono solo nel vostro cellulare o nel laptop, ma anche nelle automobili, a casa vostra, nei vostri elettrodomestici e nei giocattoli. La vostra banca è un intreccio gigantesco di algoritmi e gli esseri umani si limitano a girare qualche manopola qua e là. Gli algoritmi decidono l'orario dei voli e governano gli areoplani. Fanno funzionare le fabbriche, comprano e spediscono merci, ne incassano i proventi e tengono la contabilità. Se all'improvviso tutti gli algoritmi smettessero di funzionare, sarebbe la fine del mondo così come lo conosciamo”.³²

Gli algoritmi lavorano senza che noi ci rendiamo conto della loro presenza e del loro funzionamento. Per algoritmo si intende “un procedimento di calcolo” o ancora un metodo per la soluzione di un problema, “una sequenza di istruzioni che dice ad un computer cosa fare”³³.

Gli algoritmi sono alla base delle *machine learning*. Per capire la potenza del *machine learning* si può ricorrere ad una analogia riportata da Pedro Domingos³⁴, in cui questo tipo di tecnologia viene paragonato all'agricoltura: gli algoritmi di apprendimento sono

³⁰ C. Vercellone, A. Giuliani, F. Brancaccio, P. Vattimo, *Il comune come modo di produzione*, Ombre Corte, Verona, 2017; T. Negri, “Il comune come mezzo di produzione”, 10 giugno 2016, <http://www.euronmade.info/?p=7331>; A. Fumagalli, *Economia politica del comune*, Derive Approdi, Roma, 2017.

³¹ Al riguardo, molto interessante è il concetto di Bioipermedia, coniato da G. Griziotti: “Il Bioipermedia si può definire in modo ampio come l'ambito in cui il corpo nella sua integralità si connette ai dispositivi di rete in modo talmente intimo da entrare in una simbiosi in cui avvengono modificazioni e simulazioni reciproche”. Vedi. G. Griziotti, *Neurocapitalismo*, Mimesis, Milano, 2016, p. 120

³² P. Domingos, *L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo*, Trad. it. A. Migliori. 2016 Bollati Boringhieri editore.

³³ P. Domingos, *L'algoritmo definitivo*, op. cit.

³⁴ P. Domingos, *L'algoritmo definitivo*, op. cit.

i semi, i dati sono il terreno e i programmi appresi sono le piante adulte. L'esperto di *machine learning* è l'agricoltore il quale pianta i semi, irriga e concima il terreno senza interferire ulteriormente. Grazie a questa metafora emergono due aspetti. Il primo è legato alla grande quantità di dati, poiché più ne possediamo più possiamo imparare. La diffusione del *machine learning* è strettamente legata alla comparsa dei Big Data. Il secondo aspetto invece evidenzia come con una montagna di dati a disposizione possa ridurre la complessità che caratterizza questi processi. Con il *machine learning* il processo subisce una forte accelerazione:

“La Rivoluzione industriale ha automatizzato il lavoro manuale, e la Rivoluzione dell'informazione ha fatto lo stesso con quello intellettuale. Il *machine learning*, invece, automatizza l'automazione stessa: se non ci fosse, i programmatori diventerebbero i colli di bottiglia che frenano il progresso”.³⁵

Gli algoritmi ci aiutano a navigare tra la vastità di dati presenti in rete, ma soprattutto sono in grado di influenzare le nostre decisioni e il contesto culturale. Quelli che svolgono maggiormente questa funzione sono gli algoritmi delle piattaforme di social networking: ogni volta che le utilizziamo lasciamo informazioni che vengono registrate, elaborate e usate per gli altri utenti. Si tratta di una raccolta di informazioni individuali destinate successivamente alla collettività. Nel momento in cui gli algoritmi ci suggeriscono ciò che ci piace e ci aiutano nelle nostre relazioni, iniziano a plasmare la nostra identità e ad influire nelle nostre scelte. Nella società dell'informazione il problema principale risiede nella quantità o meglio, nella scelta illimitata che i Big Data creano: tra la molteplicità di prodotti da scegliere o di occasioni da cogliere, quale può essere quella più adatta per noi? Gli algoritmi e il *machine learning* offrono una soluzione. Nelle stesse aziende le operazioni da svolgere aumentano con il passare del tempo a dismisura, così come può crescere sensibilmente il numero dei suoi clienti. Di conseguenza il *machine learning* diventa fondamentale:

“Amazon non può codificare adeguatamente i gusti di tutti i suoi clienti in un programma, e Facebook non è in grado di scrivere un programma che scelga gli aggiornamenti migliori da mostrare a ogni suo utente. Walmart, il colosso della distribuzione, vende milioni di prodotti e deve prendere miliardi di decisioni al giorno: se i suoi programmatori cercassero di scrivere un programma dedicato, non finirebbero mai. La soluzione adottata da aziende del genere, invece, è sguinzagliare gli algoritmi di apprendimento sulle montagne di dati che hanno accumulato e lasciarli indovinare cosa vogliono i clienti”.³⁶

Questi algoritmi non sono perfetti, ma nel fornire i loro risultati influenzano l'utente e le sue decisioni. Essi sono gli intermediari tra i dati e il consumatore e concentrano al loro interno il potere e il controllo. Sono le moderne catene di montaggio.

³⁵P. Domingos, *L'algoritmo definitivo*, op. cit.

³⁶P. Domingos, *L'algoritmo definitivo*, op. cit.

Conclusioni: gli effetti sul lavoro

L'industria dei Big Data crea valore sulla base di un processo di produzione la cui materia prima è costituita dalla vita degli individui³⁷. Tale "materia prima" è in buona parte fornita gratuitamente, in quanto finalizzata alla produzione di valore d'uso.

Il "segreto"³⁸ dell'accumulazione sta nella trasformazione del valore d'uso in valore di scambio. Oppure, detto in altri termini, la trasformazione del lavoro concreto, che sta alla base delle attività quotidiane di vita che generano dati via relazioni sociali e necessità di informazioni, in lavoro astratto.

Secondo Marx, il lavoro concreto, qualitativamente definito, è volto a produrre valore d'uso; il lavoro astratto è invece pura estrinsecazione di forza-lavoro umana, che prescinde dagli aspetti qualitativi e dalle determinazioni specifiche riferite all'utilità dei singoli lavori e la cui quantità determina il valore creato. Nel sistema capitalistico di produzione, il lavoro astratto è il lavoro socialmente necessario per produrre una merce che si realizza sul mercato finale, ovvero valore di scambio, sulla base della tecnologia disponibile.

Nell'industria dei big data, il lavoro astratto è costituito dall'organizzazione e dall'integrazione dei dati. Tale attività presuppone un rapporto salariale con gli addetti assunti per tale scopo. La materia prima è invece lavoro concreto e non materia in senso stretto: sono i dati grezzi della vita quotidiana, da cui si estrae valore. Per questo parliamo di "valore-rete", un valore che si somma al valore-lavoro necessario perché tale valore-rete, che appare inizialmente come valore d'uso, possa trasformarsi in valore di scambio.

Nella valorizzazione dei big data, il processo di sussunzione si scompone quindi in due parti e cambia pelle.

Nella prima fase si attua un processo di accumulazione originaria come estensione della base produttiva sino a inglobare il tempo di vita, che non viene tuttavia salarizzato

³⁷ Al riguardo, rimando al saggio di Giorgio Griziotti "Big emotional data. Cybermarcatori somatici del neuro capitalismo" in questo stesso volume e a Giorgio Griziotti, *Megamacchine del neuro capitalismo. Genesi delle piattaforme globali*: <http://effimera.org/megamecchine-del-neurcapitalismo-genesi-delle-piattaforme-globali-giorgio-griziotti>.

³⁸ Il riferimento è K. Marx, *Das Kapital - Bd. I, VII. Der Akkumulationsprozeß des Kapitals*, 24. *Die sogenannte ursprüngliche Akkumulation*: <http://www.textlog.de/kapital-geheimnis.html>. "Geheimnis" significa "segreto" in italiano.

ovvero remunerato: nella maggior parte dei casi è *partecipazione passiva non soggettivata*. Al riguardo, non possiamo quindi parlare di vera e propria *sussunzione formale*³⁹.

Nella seconda fase, subentra l'utilizzo di forza lavoro organizzata (e salariata) che procede all'attività di *processing*, secondo i canoni più di tradizionali della *sussunzione reale*.

Per questo, possiamo concludere che il processo di valorizzazione dei big data è un ottimo esempio di *sussunzione vitale*⁴⁰ dell'uomo al capitale.

³⁹ Non è un caso che negli Stati Uniti siano sorti dei movimenti per chiedere a Facebook che la partecipazione alla piattaforma venga in qualche modo remunerata. Tra i tanti, <https://www.ft.com/content/5103204e-7b5b-11e7-ab01-a13271d1ee9c>

⁴⁰ Per un'approfondimento del concetto di "sussunzione vitale" si rinanda a: A. Fumagalli, *L'economia politica del comune*, DeriveApprodi, Roma, 2017, specie cap. I, pp. 39-68.,