ETICA E GESTIONE DEI BIG DATA

VISTA DAL PROFESSIONISTA

Roma 29 marzo 2018

NICOLA SCIARRA
Vice Presidente CI&Q ODCEC Roma

BIG DATA

La sfida più grande per la maggior parte delle aziende oggi è nella capacità di acquisire un vantaggio competitivo lavorando sui ***dati****. Tutte le aziende e tutte le Pubbliche Amministrazioni stanno diventando grandi "****fabbriche di dati****". Noi stessi contribuiamo costantemente, consapevolmente e spesso anche inconsapevolmente, alla "****produzione di dati****".*

"**Big Data**" vuol dire (letteralmente) "**Grandi dati**", ovvero grandi quantità di dati, che presi insieme occupano molto spazio di archiviazione.

Una ricerca su Google, un nostro acquisto al supermercato, una foto, un messaggio vocale, un tweet.

Tutti questi sono dati.

La maggior parte delle nostre attività quotidiane, oggi, crea dei dati, che possono essere raccolti, analizzati e monetizzati.

BIG DATA

Quando Asimov, in *Cronache della Galassia* (1951), delineò la psicostoriografia, non esisteva ancora un campo di applicazione per questa scienza.

A voler usare la definizione dell'Enciclopedia Galattica (CXVI edizione, EEA Terminus) la psicostoriografia è quella parte della matematica che studia le reazioni di un gruppo a determinati stimoli (maggiori informazioni possono essere trovate su wikipedia ma con rischio spoiler per chi non ha ancora letto i libri).

I presupposti per questa analisi però non sono da poco: il gruppo analizzato deve essere sufficientemente grande e all'oscuro dell'analisi psicostorica stessa. Sebbene secondo l'Enciclopedia galattica le dimensioni minime del gruppo possano essere trovate "facilmente" con il primo teorema di Seldon, possiamo limitarci a dire che la cifra si aggira sulle centinaia di miliardi.

Questi numeri non sono applicabili alla popolazione terrestre (al contrario della popolazione galattica a cui li applica Seldon), ma che succede **se prendiamo in considerazione i Big Data?**

Proviamo ad applicare ora la psicostoriografia ai Big Data e vediamo dove ci porta.

Psicostoriografia

Con concetti non matematici, Gaal Dornick ha definito la psicostoria come quella ***branca della matematica che studia le reazioni di un agglomerato umano a determinati stimoli sociali ed economici…***

Implicito è il fatto che ***l'agglomerato umano in questione deve essere sufficientemente grande da consentire valide elaborazioni statistiche****.*

La ***psicostoriografia*** *è la quintessenza della sociologia;* ***è la scienza del comportamento umano ridotto ad equazioni matematiche****.*

Big Data

I **big data** sono **dati** che superano i limiti degli strumenti database tradizionali, ma con questo termine si intendono anche le **tecnologie** finalizzate ad estrarre da essi conoscenze e valore.

In pratica, potremmo definire **big data** l'analisi di quantità incredibilmente grandi di informazioni.

L'analisi corretta dei **big data** ha l'obiettivo principale di **estrarre informazioni aggiuntive** rispetto a quelle che ottenibili da piccole serie di dati.

BIG DATA - DIFFUSIONE

I dati si diffondono a ritmo esponenziale. Il loro uso si sposta sempre più da marketing ed economia alla politica. I dati creano le leggi e diventano lo strumento per governare.

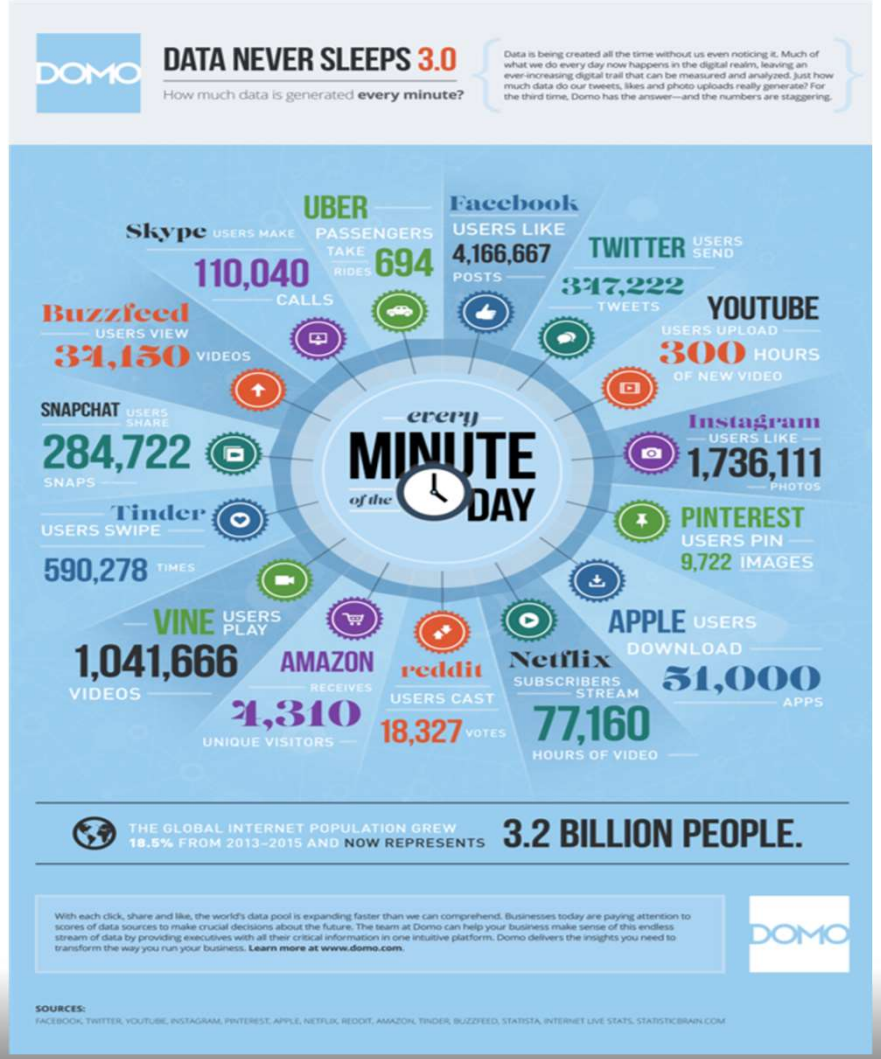
È l'era dell'impero dei Big Data: chiunque è al potere li usa ma, soprattutto, li asseconda per rimanere al comando.

*Non è davvero difficile rinvenire significative analogie tra la **psicostoria** di Asimov e lo scopo delle attuali metodologie di analisi dei **Big Data** o tra il progetto di **Galaxia** e il nostro neonato **IoT**, che implementa, giorno dopo giorno, una conoscenza (e, **forse, in futuro, anche una coscienza**) globale, da cui tutta l'umanità può trarre beneficio per il proprio sviluppo.*

I Big Data, sono scienza.

E sono la "rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà", *come li hanno definiti il docente di Oxford Viktor Mayer-Schönberg e l'analista dell'Economist Kenneth Cukier nel bestseller che gli hanno dedicato*.

L'origine è la **grande quantità** di dati che la tecnologia digitale ha messo a disposizione: il 75 per cento delle informazioni che in tutto il mondo erano immagazzinate su supporti analogici ancora nel 2000 è sceso al 7 per cento nel 2007 e al 2 per cento nel 2013, e ognuno di noi è stato investito da un "**diluvio digitale" pari a 320 volte le informazioni immagazzinate nella Biblioteca di Alessandria**.



Big Data, Futuro - 1.

Secondo molti "*in tema di* ***previsioni per il 2018****, una cosa è certa:* ***sarà l'anno dei Big Data****. Un'esagerazione? No, se pensiamo che BCC Research ha stimato che il valore di mercato della* **smart city** *raggiungerà i 775 miliardi di dollari nel 2021, con un tasso di crescita annuo del 18%. Buona parte di questa cifra monstre passerà dalla capacità o meno di rendere davvero "intelligente" il tessuto urbano, in modo che sia rapidamente adattabile agli scenari in continua evoluzione e risponda prontamente alle esigenze dei cittadini. In questo senso* ***l'Internet of Things*** *gioca un ruolo cruciale nel permettere di "leggere" ciò che avviene così da affrontarlo in modo rapido, efficiente ed efficace,* ***oltre a prevedere determinati aspetti*** *e intervenire adeguatamente".*

Big Data e data analysis consentono, infatti, di gestire ed interpretare correttamente una mole sempre più imponente di informazioni (c.d. *dataset*), generati sia da dati strutturati (cioè presenti nei *database*) sia da dati non strutturati, generalmente sotto forma di flussi di messaggi, documenti di testo, foto, video, *file* audio e *social media*.

Big Data, Futuro - 2.

Le previsioni ed i modelli basati sui *Big Data* trovano, poi, nello sviluppo dell'*Internet of Things* (IoT) un "nutrimento privilegiato", dato che la miriade di sensori e di *chip*, inseriti in modo sempre più massiccio in qualsiasi oggetto destinato a connettersi con la rete, **accresce in modo esponenziale, giorno dopo giorno, la quantità e varietà delle fonti di dati a disposizione per la successiva analisi.**

La combinazione di tutte le informazioni generate da oggetti e sistemi costantemente interconnessi tra loro ci permetterà (perlomeno, così ci promettono i tecnici)**, ben presto, di avere a disposizione sistemi di analisi assolutamente raffinati**, consentendoci di fare un salto qualitativo straordinario in brevissimo tempo.

Più dati del nostro mondo fisico trasferiremo nel mondo digitale, più i nostri algoritmi potranno essere precisi e affidabili e potranno aiutarci a migliorare in ogni campo: con i dati corretti ed in numero sufficientemente significativo e dettagliato, **miglioreremo le nostre capacità di previsione in ogni settore della vita**, costruendo **modelli sempre più utili** e **validi**, anche **economicamente**.

Big Data, Futuro – 3.

Per poter raggiungere questa meta ideale, **ci servono le tecnologie e le metodologie di *Big Data analytics*** , nella loro (per ora) triplice funzione:

1. **descrittiva**, cioè finalizzata a ricevere le giuste informazioni per poter trasformare le conoscenze in vantaggio competitivo;
2. **predittiva**, consistente in modelli di analisi di dati massivi finalizzati ad ottenere la "comprensione del futuro", fornendo alle organizzazioni **informazioni intuitive basate su dati e stime sulla probabilità di un futuro risultato** (si potrà predire, ad esempio, la propensione all'acquisto di un prodotto di un particolare *target* di clienti oppure identificare le caratteristiche comuni nei crimini come le frodi, per evitarle più facilmente);
3. **prescrittiva,** attraverso analisi che siano in grado di **quantificare l'effetto delle future decisioni per consigliare i possibili risultati prima ancora che le azioni suggerite siano effettivamente adottate**. L'analisi prescrittiva ambisce, in pratica, non solo a prevedere che cosa accadrà, ma anche a spiegare perché accadrà e, quindi, cerca di fornire le raccomandazioni migliori in merito alle azioni che sfruttano queste previsioni.

Il problema non sono i Big Data

Il problema non sono i Big Data, le tracce di noi che lasciamo in Rete, ma come sono usati e da chi.

I grandi dati, infatti, ci fanno trovare l'acqua preferita al supermercato, un'auto o un taxi per andare da dove vogliamo a dove vogliamo, verificando prima, grazie a informazioni raccolte a nostra insaputa, se c'è traffico sul percorso.

I Big Data sanno trasformarsi in Small Data, piccoli dati di grande utilità, e ci fanno sapere se il nostro volo è in orario, dopo che lo abbiamo prenotato risparmiando grazie a un sito che con i dati prepara la soluzione più adatta per un fine settimana fuori. E l'alloggio lo abbiamo ovviamente scovato e pagato, trovando il più economico, grazie all'invisibile lavoro dei Big Data.

Per non parlare delle carte fedeltà che ci regala prezzi buoni e premi vari?

Cercare lavoro? Trovare casa? Andare a un concerto?

Provarci oggi senza Big Data è impresa complicata.

Il problema non sono i Big Data

Evidentemente no, vedi il caso di Cambridge Analytica.

Tutto sta, come sempre nei confronti degli strumenti, a come si usano i mezzi a disposizione, perché e con quale livello di trasparenza. Per evitare che i Big Data da carta per rinnovare l'economia e innalzare la qualità della vita diventino armi di propaganda capaci di mettere a rischio i sistemi democratici bisogna per prima cosa rispettare le regole professate. Se dico a un social network che non voglio che i miei dati siano usati per determinati scopi, così deve essere. Se no, si rompe il rapporto di fiducia e questo dovrebbe preoccupare in primo luogo i creatori e i manager di questi nuovi colossi. Sarebbe bello se funzionassero forme di autoregolamentazione, altrimenti tocca sperare in regole e sanzioni efficaci.

Il secondo problema, più politico, è garantire che le moderne forme di sondaggio dello spirito degli elettori e dei tempi non siano deviate verso la creazione di bolle e onde emotive. Esempio: siccome so che tu tieni molto al tema X, per esempio la gestione dei flussi migratori, ti preconfeziono un flusso di notizie non vere e tendenziose con commenti forti per fomentare la rabbia a scopo di raccolta del consenso. Anche qui le autoregolamentazioni sarebbero benvenute, altrimenti tocca confidare in normative giuste e in lettori (ed elettori) saggi, capaci di andare a cercare informazioni in luoghi più sicuri.

Il problema non sono i Big Data

Sì, ma la politica è un'altra cosa, si dice. La politica è il limite?

Per esempio, Barack Obama. Le campagne elettorali dell'ex presidente democratico sono state universalmente elogiate anche perché hanno portato la rivoluzione digitale nella politica.

Obama ha vinto le primarie e poi le presidenziali anche "**misurando tutto**", come diceva il suo stratega elettorale Jim Messina

Incrociando i dati degli elettori (negli Stati Uniti ci si registra come "democratici", "repubblicani" o "indipendenti") e quelli raccolti in Rete e sui social network si potevano e si possono fare campagne elettorali mirate fino al punto di far arrivare all'elettore credente il messaggio, la lettera o l'e-mail che sottolinea la probità religiosa del candidato X oppure recapitare all'elettore pensionato le proposte di riforma avanzate dalla candidata Y per rafforzare il sistema previdenziale.

Lo staff di Obama, con questo fine: "***misurare tutto***", faceva più di 100 analisi di dati con oltre 66.000 simulazioni al computer ogni giorno. E ha vinto anche grazie a questa moderna applicazione su scala più ampia e interconnessa di quella che era l'arte dei sondaggi e della statistica. Tutto bene, dunque?

REGOLAMENTAZIONE – GDPR

Le regole minime da rispettare:

I dati personali devono essere trattati in modo lecito, corretto e trasparente nei confronti dell'interessato («**liceità, correttezza e trasparenza**») e quindi occorre sempre valutare se l'utilizzo dei dati personali nei Big Data sia nelle ragionevoli aspettative delle persone considerando che la complessità dei metodi di Big Data può talvolta rendere difficile per le organizzazioni di essere trasparenti sul trattamento dei dati personali.

La legittimità degli interessi non è un'opzione.

I Big Data devono sempre bilanciare gli interessi del titolare/responsabile con quelli degli interessati.

- **limitazione della finalità**
- **minimizzazione dei dati**
- **esattezza**

Sono tutti dati personali, e tutti devono essere trattati conformemente a normativa.

GDPR – PRIVACY

Un nuovo concetto di privacy

Nel momento in cui si entra nell'ambito della Data Monetization ecco che occorre affrontare il tema normativo e nella fattispecie l'autorizzazione del trattamento per le finalità dichiarate.

Non più codice della privacy ma **GDPR 2018**, che letteralmente sta per "**General Data Protection Regulation**" e dal maggio 2018 sarà il nuovo **Regolamento Europeo Privacy**.

Stiamo parlando dell'*Articolo* 6 del **General Data Protection Regulation** (**GDPR**) che delibera proprio sulle condizioni che permettono o non permettono il trattamento.

Il principio normativo in base al quale si lavora sui dati solo se l'interessato ha espresso il consenso al trattamento dei propri dati personali per una o più specifiche finalità è un vero e proprio "faro" nell'attività del Data Scientist.

CLOUD COMPUTING

Con il termine cloud computing o semplicemente cloud nell'ambito di quesa trattazione ci si riferisce a un insieme di tecnologie e di modelli di servizio che:

- **favoriscono la fruizione e l'erogazione di applicazioni informatiche, di capacità elaborativa e di stoccaggio via web;**
- **promuovono a seconda dei casi il trasferimento dell'elaborazione o della sola conservazione dei dati dai computer degli utenti ai sistemi del fornitore dei servizi.**

PROFESSIONISTI

Smart Working – Social Collaboration

Cosa sono, a cosa servono e quali sono i vantaggi degli strumenti di collaborazione di nuova generazione: i nuovi software in Cloud offrono l'esperienza dei social, mettono le persone al centro, realizzano la **extended enterprise** e rendono semplice lavorare collaborando a distanza.

Extended enterprise significa mettere i collaboratori al centro di una organizzazione flessibile, agevolare lo svolgimento del lavoro ovunque (**smart working**) garantendo al tempo stesso **usabilità** e soprattutto **sicurezza**: significa **condivisione dei documenti, strumenti di collaborazione, accesso ai sistemi centrali, nuove architetture e applicazioni,** il tutto coordinato attraverso nuovi modelli organizzativi

Servono anche per adeguarsi al GDPR e non mettere a rischio dati sensibili.

PROFESSIONISTI

Cos'è la Social Collaboration e come agevola il lavoro

*«È un tema caldo, e continuerà a esserlo», conferma in un'intervista a ComputerWorld **Wayne Kurtzman, Research Director di IDC**, «in quanto le piattaforme di **Social Collaboration** incarnano il concetto stesso di **digital disruption**: sono alimentate dalle persone, usano nuove metriche comportamentali e hanno ricadute positive sui KPI aziendali».*

Gli strumenti che, anche grazie all'intelligenza artificiale, riescono a coniugare tutti questi aspetti e a cui le aziende stanno delegando sempre più il compito di trasformare l'aspirazione dello **smart working** in una realtà consolidata sono le piattaforme di **Social Collaboration**.

*Un trend che coinvolge anche molte aziende italiane, tra cui pure le PMI», commenta **Daniele Vistalli CEO di FACTOR-Y.***

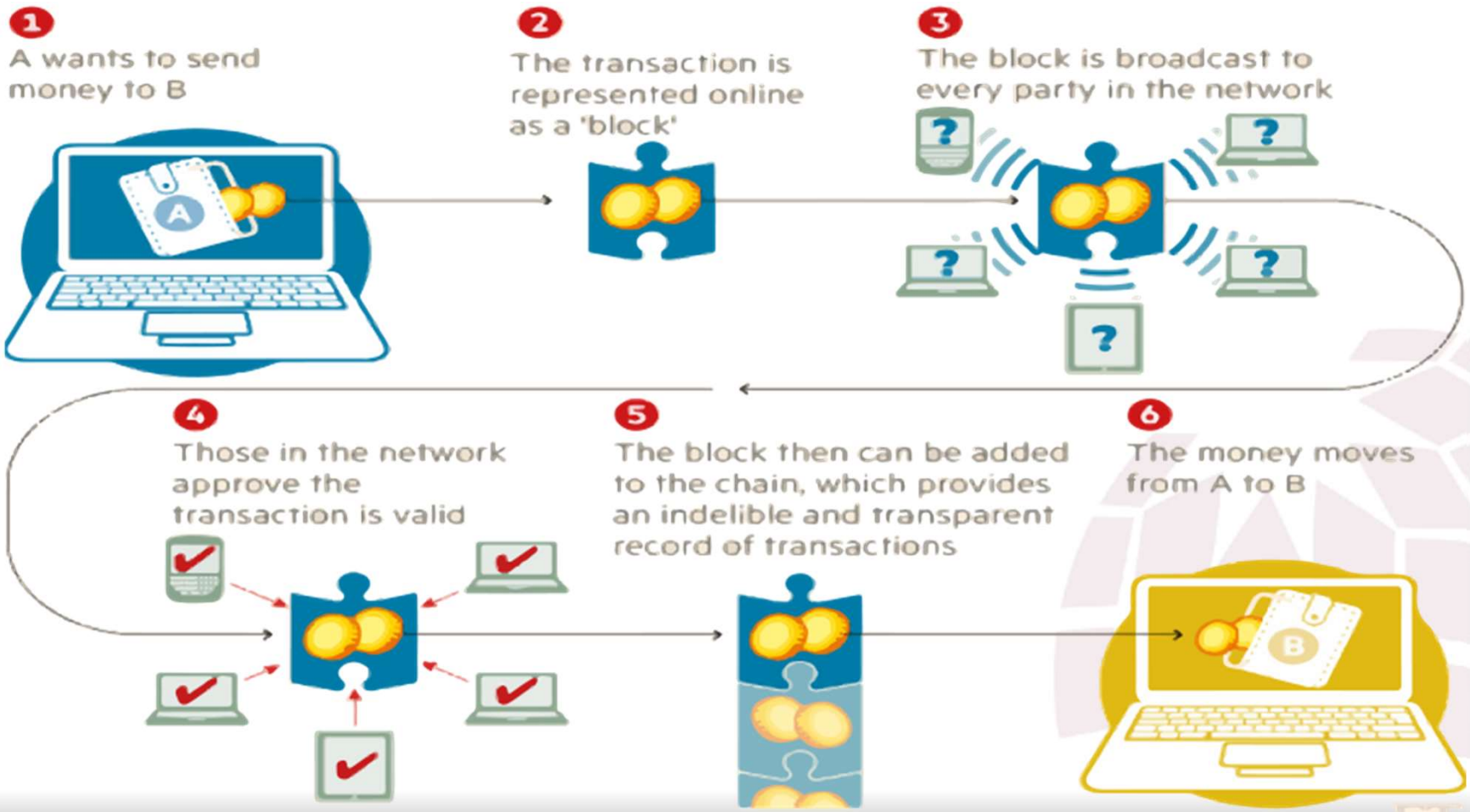
Se nelle organizzazioni di dimensioni maggiori le Intranet sono diventate ben più di semplici punti di accesso alle news aziendali, trasformandosi in veri e propri forum dove competenze e attività di formazione vengono condivise attraverso meccanismi di gamification, le piattaforme di Social Collaboration propriamente dette stanno prendendo piede grazie agli indiscutibili vantaggi che offrono: uniformità nella user experience, sicurezza, flessibilità e misurabilità dei risultati prodotti dalle interazion.

BLOCKCHAIN

Possiamo definire la blockchain come un immodificabile registro digitale di transazioni economiche, che può essere programmato per registrare non solo transazioni finanziarie ma, in teoria, ogni tipo di informazione. **(cit. Diego Barberi).**

Non a casa i big data, insieme alla tecnologia blockchain, sono considerati dal World Economic Forum tra le innovazioni che stanno cambiando il mondo.

Esempio di Blockchain gestione money transfer



BLOCKCHAIN

Con la blockchain il registro del consenso diventa pubblico, in modo da poter permettere di compiere transazioni in tutta sicurezza, con più soggetti.

Ma senza la necessità di intermediari, la **blockchain sopperisce con la matematica e la tecnologia alla mancanza di fiducia reciproca**

Il denaro elettronico non esiste.

Esiste solo un registro sul quale c'è scritto a chi appartiene e chi lo tiene in custodia. **La gestione di quel registro è il punto chiave.** Per mezzo della trascrizione di proprietà sul registro si conferiscono i titoli ai soggetti e tutto ciò consente ai mercati complessi di funzionare.

Queste primitive necessità, la moneta e la trascrizione, hanno percorso tutta la nostra storia economica **congiungendosi per mezzo delle banche**.

Finora il sistema bancario ha mantenuto il potere per le monete *fiat*, da qui in avanti esiste anche un altro registro, la **blockchain**.

BLOCKCHAIN

Si tratta di un registro permanente del consenso che abbiamo raggiunto sullo status di un asset. È un marcatore temporale, quindi è solo una conseguenza del precedente, e ben più importante, consenso. Qui c'è scritta la relazione tra il proprietario e l'asset (se lo possiede o controlla, se ha l'accesso, etc.)

Vediamo alcuni casi possibili:

- **Se il registro è nelle sole mani del proprietario, o dell'azienda, non lo deve confrontare con altri.**
- **Se è tra due persone, il consenso o lo si trova o non lo si trova.**
- **Se è tra più di due persone, e il registro resta privato, nasce il problema del *double spending:* per questo fino ad ora è stato condiviso e gestito da un intermediario**. *Per esempio, il catasto.*

L'UOMO HA LA POSSIBILITÀ DI TRUCCARE IL SISTEMA PERCHÉ I SUOI INCENTIVI SONO ALLINEATI AI SUOI INTERESSI, NON AL SISTEMA.

BLOCKCHAIN

L'essere umano, può farlo perché non è difficile, può inventare qualsiasi stratagemma, fino a quando gli altri soggetti o enti regolatori non lo scoprono.

Solo se lo scoprono, inizia un percorso di erosione della fiducia e eventuale sanzione e esclusione. A quel punto scatta un meccanismo perverso di reazione uguale e contraria, così si rende il sistema sempre più complicato e insicuro. Da qui le dispute nei tribunali.

Con la blockchain il registro del consenso si rende pubblico, per esempio fuori dal perimetro aziendale, per transare in sicurezza con più soggetti senza intermediari.

Dato che all'interno dell'azienda il meccanismo di coordinamento è la gerarchia, non è possibile "**esportare la gerarchia**" in una rete tra partner quando impieghiamo una blockchain.

Per questo c'è bisogno di un consenso che richiede un **lavoro lungo** ma **facilmente verificabil**e per accettarsi che nessuno tenti di modificarlo a suo vantaggio.

BLOCKCHAIN

Non è uno spreco di risorse, perché anche le chiavi per le serrature e le inferriate per il mondo fisico, e il firewall e gli antivirus per il mondo digitale, sarebbero uno spreco. Per non parlare dell'immenso mondo assicurativo per la copertura del rischio di furti.

In realtà l'umanità ha speso sempre enormi risorse economiche, e che concorrono a far crescere il Pil, solo per proteggersi da se stessa. Tutte queste ingenti risorse sono dovute solo alla mancanza di fiducia reciproca.

Con l'uso della blockchain ce ne priveremo per sempre.

SAREMO OBBLIGATI A COMPORTARCI IN MODO ONESTO.

La blockchain è sottrazione di volere e potere all'uomo per darli alla matematica e alla tecnologia, perché non abbiamo fiducia tra noi.

BLOCKCHAIN

C'è però da fare un'ultima considerazione, di carattere strutturale e sociale. Una **svolta tecnologica** è possibile solo in presenza di una **società fortemente "digitalizzata"**, che abbia competenze ed eque possibilità d'accesso. E questo dipende anche dalle scelte e dalle **politiche** dei governi.

In Italia, secondo i dati **Istat** relativi al 2016, si è connesso alla rete il 67,4% delle famiglie. Più in generale, nella popolazione da sei anni in su, il 36% non usa internet. In uno scenario simile il rischio sarebbe quello di vedere **sovradimensionate le istanze di alcuni gruppi più "tecnologici",** a scapito di altri che utilizzano di meno internet e suoi strumenti.

MA LA TECNOLOGIA NON È NECESSARIAMENTE UN ELEMENTO DA DEMONIZZARE.

COME TUTTI GLI STRUMENTI ESSA NON È NEGATIVA DI PER SÉ.

PIUTTOSTO PUÒ ESSERE NEGATIVO, QUANDO NON DISASTROSO

L'USO CHE SE NE FA.