

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea in Informatica per il Management

**CITTADINI AL TEMPO DELLA
DATAFICATION: I BIG DATA POSSONO
ESSERE USATI PER IL BENE SOCIALE?**

Relatore:
Chiar.mo Prof.
EDOARDO MOLLONA

Presentata da:
MARCO BRONZINI

Sessione II
Anno Accademico 2018/2019

Un ringraziamento speciale a mia madre che mi ha sempre sostenuto in tutti i momenti difficili aiutandomi a credere in me stesso. Inoltre, vorrei dedicare questo percorso a mio nonno che spero lo avrebbe reso orgoglioso di me. Per concludere voglio ringraziare gli amici che mi sono stati vicini negli ultimi anni che mi hanno aiutato a prendermi meno sul serio.

Introduzione

Sono sempre stato interessato ad approfondire quale impatto ha avuto e può avere l'informatica in tutti i settori della società moderna, in particolare l'uso dei dati in tutti i processi decisionali. Frequentare questo corso di Laurea mi ha permesso di approfondire molti aspetti inerenti ai processi aziendali anche relativi all'analisi dei dati per implementare strategie e soluzioni per migliorare l'efficienza aziendale.

Negli ultimi anni si è assistito ad una crescita esponenziale di soluzioni commerciali basate sull'utilizzo dei dati, favorite per esempio dall'incremento di disponibilità di fonti eterogenee che generano grandi quantità di dati da cui possono essere estratte informazioni utili. Tuttavia, in questo elaborato voglio approfondire un altro aspetto inerente all'utilizzo dei dati, cioè utilizzare essi per scopi umanitari e quindi per il bene sociale con ricadute sull'intera comunità. La digitalizzazione degli ultimi vent'anni ha permesso la trasformazione di molti aspetti delle nostre vite sotto forma di dati da cui poter estrarre nuove forme di valore. Questo particolare trend prende il nome di "Datafication" di cui ho cercato di approfondire le dimensioni e le problematiche nel primo capitolo del mio elaborato. Per parlare dell'utilizzo dei dati nella nostra società bisogna introdurre il concetto di "Big data" che voglio esplicitare attraverso la definizione elaborata da Degryse (2016) ovvero che essi siano *"una combinazione e somma totale dei dati (personali, commerciali, geografici, comportamentali) disponibili sulle reti digitali (internet, telefoni cellulari, navigatori satellitari, ecc.) e utilizzabili come materia prima"*. Proprio questa definizione mi permette di introdurre e analizzare successivamente il neologismo chiamato "Colonialismo dei dati" che vuole enfatizzare lo sfruttamento degli esseri umani attraverso i propri dati (che diventano quindi una materia prima da cui estrarre valore). Questo concetto di colonialismo dei dati è inserito all'interno del più vasto fenomeno

della datafication, tuttavia quest'ultimo presenta altri pericoli ed effetti collaterali come: la presunzione di un universalismo della tecnologia, la poca trasparenza nella raccolta ed utilizzo dei dati, un incremento della deumanizzazione del lavoro e una potenziale violazione della privacy. Problematiche che ho cercato di approfondire nei primi capitoli del mio elaborato. Una possibile soluzione per ridurre queste problematiche è data da un ulteriore incremento delle tutele legislative oppure tramite una drastica riduzione dei dati che si concedono alle aziende. Un'altra soluzione complementare è quella di concedere questi dati ad organizzazioni che si pongono come obiettivo quello di utilizzare quest'ultimi per azioni umanitarie o sociali che generino benefici per i singoli cittadini e per l'intera comunità.

In questo elaborato ho deciso di approfondire quest'ultima soluzione cercando di analizzare diversi progetti ed iniziative che utilizzano un approccio basato sui dati per implementare soluzioni etiche che abbiano ricadute per il bene sociale che quindi si allontanano dal concetto negativo del colonialismo dei dati. Iniziative internazionali come "Global Pulse" delle Nazioni Unite che sviluppa progetti proprio seguendo questo tipo di obiettivo, come per esempio nel caso dell'utilizzo dei big data per studiare i pattern di soccorso nel Mar Mediterraneo. Oppure, l'iniziativa "Big Data for Social Good" dell'associazione GSMA che abilita la realizzazione di progetti etici basati sull'utilizzo dei dati mobili; come nel caso del progetto che invia un avviso a tutte le persone che si trovano in una specifica area in cui è stata segnalata una persona scomparsa. Inoltre, ho approfondito un progetto sviluppato dalla sezione interna "Science for Social Good" del dipartimento di ricerca dell'azienda IBM, che attraverso l'analisi di dati relativi a metriche socioeconomiche cerca di determinare il livello di innovazione e competitività di una nazione. Quest'ultimo progetto è un esempio concreto per utilizzare un approccio basato sui dati per determinare empiricamente un fenomeno socioeconomico e potenzialmente implementare politiche più efficaci per la crescita di una nazione.

In aggiunta a queste iniziative, ho cercato di approfondire un altro aspetto relativo all'uso sociale dei dati, cioè se i dati concessi dai cittadini alle Pubbliche amministrazioni possono generare benefici per i singoli individui e l'intera comunità oppure se presentano effetti collaterali. Per questo scopo ho analizzato "AgID - l'Agenzia per L'Italia Digitale" che mi ha permesso inoltre di approfondire la situazione italiana in riferimento all'utilizzo

degli Open data, i vantaggi della digitalizzazione e di un approccio basato sui dati da parte delle Pubbliche amministrazioni italiane. Infine, un altro aspetto che ho esaminato è relativo al trend delle “Smart City” per capire le conseguenze dell’enorme mole di dati che viene generata dal contesto urbano e se sono presenti degli effetti collaterali riconducibili per esempio al colonialismo dei dati.

Dopo aver analizzato tutti questi progetti, nelle conclusioni ho cercato di riassumere gli elementi comuni in tutti i progetti e ho cercato di ipotizzare quali caratteristiche siano necessarie per sfruttare i big data per il bene della collettività.

Indice

Introduzione	i
1 La dimensione della Datafication	1
1.1 La nostra vita in dati	1
1.2 Il concetto del Colonialismo dei dati	4
1.3 La deumanizzazione nel mondo del lavoro	7
1.4 Il ruolo degli algoritmi	10
1.5 Alla ricerca di pattern	12
2 Il concetto di Privacy	15
2.1 La definizione	15
2.2 La percezione e le problematiche	18
3 Smart City – Le città al tempo dei Big data	21
4 Analisi iniziative per il bene sociale	25
4.1 United Nations Global Pulse	25
4.1.1 REPORT: Usare i Big data per studiare i pattern di soccorso nel Mar Mediterraneo (<i>UN Global Pulse, 2017a</i>)	28
4.1.2 REPORT: Utilizzo dei dati mobili per il bene sociale: successi, sfide e sviluppi futuri (<i>UN Global Pulse, 2017b</i>)	30
4.1.3 REPORT: Monitorare l’impatto delle crisi economiche sui crimini (<i>UNODC, 2012</i>)	34
4.2 GSMA - Big Data for Social Good	37

4.2.1	REPORT: Utilizzare l'analisi mobile in tempo reale come aiuto per il salvataggio di persone scomparse (<i>GSMA BD4SG, 2019a</i>) . . .	39
4.2.2	REPORT: Costruire comunità resilienti agli estremi climatici (<i>GSMA BD4SG, 2018</i>)	41
4.2.3	REPORT: Rispondere rapidamente ed efficacemente alle catastrofi in Giappone (<i>GSMA BD4SG, 2019b</i>)	44
4.3	IBM - Science for Social Good	46
4.3.1	REPORT: Come promuovere l'innovazione: un approccio basato sui dati per misurare la competitività economica (<i>Kuhlman et al., 2017</i>)	47
5	AgID – Agenzia per l'Italia Digitale	51
5.1	Presentazione e obiettivi dell'Agenzia	51
5.2	La situazione degli Open data in Italia	54
5.3	Progetti e appalti innovativi	57
5.4	L'intelligenza artificiale al servizio del cittadino (IA-GOV)	58
	Conclusioni	63
	Bibliografia	67

Elenco delle figure

1.1	Colonialismo del Potere	3
1.2	Dimensione del Colonialismo dei Dati	7
2.1	Soggetti sollecitati in caso di violazione della Privacy	20
4.1	Strategie dell’iniziativa UN Global Pulse	27
4.2	Spostamenti delle navi durante le operazioni di soccorso	29
4.3	Correlazione tra fattori economici e criminalità	36
4.4	Tipica implemetazione di un progetto all’interno dell’iniziativa GSMA	39
4.5	Migrazione della popolazione della regione colombiana di “La Guajira”	43
4.6	Sistema di supporto decisionale intelligente utilizzato in Giappone	45
4.7	Alcuni parametri selezionati per formare l’Open Innovation Index (OII)	48
4.8	Livelli d’innovazione globali attraverso un appoggio basato sui dati	49
5.1	Infrastruttura digitale italiana	53
5.2	Punteggio europeo riguardo la maturità degli Open data	56
5.3	Maturità degli Open data - Classificazione paesi europei	56
5.4	Potenziati danni causati da sistemi AI	62

Capitolo 1

La dimensione della Datafication

1.1 La nostra vita in dati

Stiamo vivendo in un ventunesimo secolo basato sull'economia della conoscenza, cioè poter sfruttare le informazioni ricavate da decisioni, azioni o relazioni per cercare di estrapolare un valore. Sfruttando anche l'enorme mole di dati generati da sempre più fonti eterogenee che poi potranno essere inseriti all'interno di potenti e complessi modelli matematici.

Nasce infatti il concetto di “datafication” cioè un nuovo trend tecnologico che permette la “trasformazione di molti aspetti delle nostre vite in dati da cui possono essere estratte nuove forme di valore” (Cukier e Mayer-Schöenberger, 2013). Può essere visto come un nuovo potente strumento a disposizione delle compagnie ed istituzioni per gestire e analizzare i comportamenti delle persone. Un esempio è il potere attribuito agli algoritmi che elaborano la maggior parte dei dati presenti nelle piattaforme della “gig-economy” che “permettono un aumento delle già presenti disuguaglianze nel mercato del lavoro a richiesta” (Chen, 2017). Con il termine “gig-economy” si intende un “modello economico in cui certe categorie di lavoratori svolgono attività autonome, da freelance o con contratti temporanei, invece che su base continuativa e a tempo indeterminato” (Dizionario Zanichelli, 2019). Inoltre, non sono da sottovalutare gli effetti collaterali nell'utilizzo di tecniche di datafication che possono rappresentare una potenziale invasione della privacy ed iniquità. Infatti, le sofisticate tecniche di “data mining” permettono

l'estrazione di informazioni anche da dati grezzi e non significativi ad un primo esame. Tuttavia, solo negli ultimi anni queste preoccupazioni sono entrate nel dibattito letterario (Blumenstock, 2018; Zuboff, 2017; Cukier e Mayer-Schöenberger, 2013) e marginalmente anche in quello politico.

Il ruolo centrale dei dati nella società moderna si basa sostanzialmente su tre principi: riflettono la realtà, l'analisi di questi dati genera conoscenza preziosa ed accurata e il risultato di quest'analisi può essere usata per migliorare le decisioni riguardo il mondo esterno (Ricaurte, 2019). Questa visione della nostra epoca può essere analizzata anche attraverso il modello del "Colonialismo del Potere" (Fig. 1.1) che si basa sull'appropriazione della conoscenza, esistenza e sensibilità (Ricaurte, 2019). Esprimibili in diversi livelli come per esempio quello economico: attraverso l'analisi di flussi di dati economici o la valorizzazione monetaria di essi. Possono essere valorizzati anche aspetti intrinsecamente legati alla nostra persona come: genere, sessualità, caratteristiche fisiche. Inoltre, analizzando il flusso di scelte ed azioni è possibile catturare anche gli aspetti più intimi quali per esempio: gli affetti (emozioni e percezioni della realtà), il modo di comunicare, la mobilità, i nostri desideri e abitudini. Ma anche una conoscenza relativa a risorse naturali o fenomeni socioeconomici per un supporto nei sistemi politici.

Inoltre, bisognerebbe "superare il concetto dell'universalismo della tecnologia, cioè rappresentare essa come un elemento storico e asociale che opera al di fuori di specifici contesti sociopolitici, culturali o economici" (Milan e Treré, 2019). Come sostengono i due autori citati poc'anzi "il travaso tecnologico nei paesi in via di sviluppo è lontano dall'essere rigido ed immutato, poiché il contesto ne può modellare i principi secondo le forze sociali presenti localmente come istituzioni o comunità". In queste nazioni "è spesso assente o molto limitata l'agenda digitale dei governi per attuare politiche di inclusione digitale" (Milan e Treré, 2019). Tuttavia, i cittadini di questi paesi possono diventare a loro volta vittime della datafication in quanto rappresentano un enorme bacino di potenziali consumatori per i servizi delle società occidentali.

Oltre la presunzione dell'universalismo della tecnologia, il fenomeno della datafication introduce anche altre problematiche ed effetti collaterali come: il colonialismo dei dati, la poca trasparenza nella raccolta e nell'utilizzo dei dati, un incremento della deumanizzazione nel lavoro e una potenziale violazione della privacy. Infatti, con un parallelismo

logico con il termine colonialismo storico, nel quale ci si appropriava di terreni e risorse per trarne un profitto, è stato creato il termine “Colonialismo dei dati” (data colonialism) per esprimere lo sfruttamento degli esseri umani attraverso i propri dati aprendo la strada alla “capitalizzazione della vita senza limiti” (Couldry e Mejias, 2019).

Table 1. The Coloniality of Power and the Dimensions of Data Colonialism.^a

The coloniality of power: the coloniality of knowing, being, sensing	Economy	Economic flows/data flows (financial systems, markets, transport, health, mobility) Political economy of data	Data extractivism/Data capture (the process of collection, storage, access, analysis, and use of data): Emerging industries and markets Data labor The economic value of data
	Political system	Authority Power distribution	State capture Techno-corporatism Supra-national coloniality Law Institutions
	Knowledge	Data Epistemologies Geopolitics of knowledge	Algorithmic violence/algorithmic capture: Truth regimes: data-ism, privacy, digital universalism, centers of knowledge production Knowledge propriety Corporate understanding Categorization
	Being and sensing	Gender Sexuality Body Race Language Subjectivity	Capture of life: Patriarchy Servitude Mobility Self-expression Affects: emotions, perceptions, sensations Techno-surveillance Self-monitoring Profiling Desires Interactions Communication Practices Imaginary of the self Representations and narratives Norms Biodata
	Nature	Territory Natural resources	Capture of the commons: Land/space Water Air Energy Biodiversity
	Socio-technical systems	Data infrastructures and technological development	Technological capture: Software Hardware Algorithms Artificial Intelligence/Machine Learning Data centers/“The cloud”

^aAn earlier version of these charts was published in Spanish (Ricaurte 2018).

Figura 1.1: Colonialismo del Potere

FONTE: (Ricaurte, 2019)

1.2 Il concetto del Colonialismo dei dati

Per introdurre questo fenomeno devo prima definire il concetto di “big data” che voglio esplicitare attraverso la definizione elaborata da Degryse (2016) ovvero che essi siano *“una combinazione e somma totale dei dati (personali, commerciali, geografici, comportamentali) disponibili sulle reti digitali (internet, telefoni cellulari, navigatori satellitari, ecc.) e utilizzabili come materia prima”*. Il concetto di big data è abbastanza generico e aperto a molte definizioni possibili, tuttavia ho scelto questa in particolare perché mi permette di introdurre il neologismo del “Colonialismo dei dati”. Infatti, recentemente nel contesto del trend della datafication è stato effettuato un parallelismo logico con il termine colonialismo storico. In una declinazione moderna, questo fenomeno combina le pratiche estrattive predatorie con i metodi di quantificazione astratta dei moderni sistemi informatici (Couldry e Mejias, 2019). Per questa ragione è stato coniato il termine “colonialismo dei dati” per esprimere lo sfruttamento degli esseri umani attraverso i propri dati (che diventano quindi una materia prima da cui estrarre valore). Non è possibile delimitare questo trend in specifiche aree geografiche o tematiche in quanto risulta un fenomeno globale che coinvolge potenzialmente qualsiasi persona o cosa che si trovino all’interno delle odierne infrastrutture connesse. In questo ambiente virtuale la nostra vita sociale diventa una risorsa facilmente accessibile e potenzialmente sfruttabile per estrarre del valore da essa.

La tabella (Fig. 1.2) elaborata da Ricaurte (2019) dimostra come il processo del Colonialismo dei dati può essere inserito in uno scenario suddiviso in 3 fasi distinte ma complementari:

1. **L'estrazione dei dati** dalle sorgenti;
2. **L'elaborazione dei dati** attraverso l'utilizzo di sempre più complessi algoritmi;
3. Lo **sfruttamento delle informazioni ricavate** da parte degli agenti di mercato.

Inoltre, a loro volta ogni fase può essere vista come una sequenza di processi che determinano la collezione, memorizzazione, accesso, analisi e l'uso delle informazioni estratte durante una specifica fase. Per esempio, si possono utilizzare i principali provider di dati come sorgenti da inserire in sempre più complessi algoritmi anche utilizzando strumenti

avanzati come l'intelligenza artificiale o l'apprendimento automatico. Tipicamente con l'utilizzo di questi algoritmi, che si basano sostanzialmente su modelli matematici, si cerca di evidenziare dei pattern che consentano la creazione di modelli predittivi che genereranno a loro volta preziose informazioni.

L'economia digitale ha infatti permesso la nascita di un nuovo modello di business basato sulla raccolta di grandi quantità di dati, i quali genereranno un valore dopo essere analizzati o rivenduti a soggetti terzi. Questi processi di raccolta e analisi dei dati sono tipicamente concentrati in poche aziende con grandi disponibilità di capitali che, tramite un elevato livello tecnologico e di competenze, permette loro di possedere “una capacità d'analisi a volte anche maggiore rispetto ai singoli stati nazionali” (Ricaurte, 2019). Le operazioni a livello internazionale di queste società costringono i legislatori a rispondere a diversi livelli di giurisdizione, che raramente riescono a tenere il passo del dinamismo digitale. L'autrice Ricaurte (2019) sostiene che “una politica futura dovrebbe cercare di contenere e limitare questo enorme potere in mano ad aziende”; tuttavia non si tratta di nostalgia al luddismo, ma piuttosto “cercare di riportare la sovranità alla nazione anche a livello tecnologico per evitare un eccessivo controllo da parte del mercato, una maggior giustizia sociale e difesa dei diritti”. Nella nostra epoca i principali attori che esercitano questo potere sono legati a paesi come gli Stati Uniti d'America e Cina, che possono agire sia a livello locale che globale, in quanto l'infrastruttura connessa e la globalizzazione hanno permesso di avere quasi l'intera popolazione mondiale come potenziali consumatori di servizi offerti da imprese quali per citarne alcune: Google, Facebook, Amazon, Alibaba, Huawei e Tencent. Infatti, i principali agenti del fenomeno del colonialismo dei dati “sono imprese che sono coinvolte nella raccolta di azioni sociali quotidiane che possono essere analizzate e usate per generare un profitto” (Couldry e Mejias, 2019). Tuttavia, non bisognerebbe sottovalutare anche il potenziale inespresso di aziende che non hanno ancora colto l'opportunità di estrarre del valore dai dati in loro possesso; infatti molto spesso si è investito solo nella gestione e memorizzazione di essi e non piuttosto sull'analisi e valorizzazione di questi dati.

Non si può dimenticare, tuttavia, che colonialismo tra il XVI e XX secolo ha generato anche l'uccisione di milioni di nativi e lo sfruttamento di risorse naturali del pianeta. Il parallelismo che è stato effettuato dagli autori (Couldry e Mejias, 2019) di questo

neologismo ha lo scopo di enfatizzare l'aspetto di appropriazioni di risorse con una connotazione predatoria. Tale raccolta di dati è effettuata anche cercando di plasmare gli aspetti della nostra vita, come le relazioni sociali, rendendo “il più naturale possibile questo processo di estrazione di valore da essi” (Couldry e Mejias, 2019). Le piattaforme digitali hanno infatti permesso di creare un mezzo ideale per la nascita di questo nuovo paradigma sociale, per il quale ogni aspetto può essere continuamente tracciato, strutturato e analizzato sotto forma di dato. Già Arvidsson (2016) aveva analizzato queste tecniche di estrazione di valore da parte delle piattaforme digitali dei propri contenuti come forma di autofinanziamento, che rende opaco e poco trasparente il meccanismo di appropriazione dei dati degli utenti; controbilanciato dal fatto che spesso la fruizione di questi servizi è resa gratuita. Ma è centrale questo principio di appropriazione degli aspetti sociali come elementi di valore, anche incoraggiato dalla continua condivisione ed esposizione dei nostri pensieri ed attività (Couldry e Mejias, 2019).

L'estrazione di informazioni non solo riguarda gli aspetti sociali, ma anche moltissime aree della società moderna: per esempio, in tutte le aree di produzione dove spesso è presente una combinazione di lavoratori umani e macchinari, nei quali una raccolta di dati aggregati e l'elaborazione su larga scala di essi permettono di utilizzare queste informazioni per una sempre maggiore ottimizzazione di tempi, costi, produzione e logistica (Couldry e Mejias, 2019). Cioè la cosiddetta industria 4.0 che contamina molti settori come anche l'agricoltura, dove per esempio tramite l'utilizzo di vari sensori e l'automatizzazione è possibile aumentare la produttività di terreni, diminuire l'impatto ambientale e potenzialmente aumentare la qualità dei prodotti. Un altro aspetto non secondario in questa visione di centralità dei dati è la possibilità di tracciamento delle attività di un singolo individuo; volontariamente, per esempio attraverso i Fitness Tracker, oppure vincolata da un contratto di lavoro durante lo svolgimento di mansioni lavorative. Quindi mettendo insieme tutte queste trasformazioni “esse rappresentano una possibile appropriazione dell'intero dominio della vita individuale e sociale” (Couldry e Mejias, 2019). Il tutto inserito in una società che si basa sostanzialmente sul sistema economico capitalista, dove vi è una ricerca indiscriminata di un valore monetario seguendo i principi base dell'accumulazione privata di capitale e della libera iniziativa. Il trend della datafication sta cambiando la società nel suo complesso, “trasformando la vita umana in

una nuova forma astratta utilizzata come merce di scambio sotto forma di dati” (Couldry e Mejiias, 2019). “Come storicamente successe con il capitalismo industriale, che secondo il pensiero di Karl Marx, trasformò l’attività universale del lavoro (mercificandola) in una forma sociale con una dimensione astratta” (Couldry e Mejiias, 2019).

Table 2. Data Colonialism as a Multilayered Process.

	Collect	Storage and access	Analyze	Use/value/implications
Data extractivism	Unstructured data flows Data in motion Data aggregation Real-time, systematic, and pervasive data collection	Connectivity Capacity Velocity Infrastructure-as-a-service (IaaS)	Data processing and transformation Data analysis (statistics and prediction capabilities) Software-as-a-service (SaaS) Crowdsourced markets of workers Human intelligence	Monetization of data: making sense of data for different contexts and purposes (Marketing, Politics, Health, Banking, Insurance, Labour, Weather) Alliances among different actors: corporations, governments, centers of knowledge production Epistemic dominance
Algorithmic violence	Dominant Platforms Walled gardens Algorithms and black boxing Data-ism Spyware Data Brokers	Land and resources Data centers Cloud services	Artificial Intelligence Machine Learning Data Science Neuroscience Linguistics	Prediction models Profiling Reproduction of structural violence through prediction models and automatization: Inequality Discrimination Surveillance Control Exploitation Domination Manipulation Death
Agents/emergent industries and markets	Facebook Apple NSO Group Gamma International Equifax Cambridge Analytica	Amazon Web services Google Cloud	Facebook Amazon Google IBM Palantir	Clients: Governments Companies Agencies Universities

Figura 1.2: Dimensione del Colonialismo dei Dati

FONTE: (Ricaurte, 2019)

1.3 La deumanizzazione nel mondo del lavoro

Come scritto in precedenza, il fenomeno della datafication e quindi lo sfruttamento dei dati non può essere delimitato in un unico settore o area tematica, poiché questo trend è abilitato dall’ormai onnipresenza della tecnologia all’interno di tutte le sfere della società moderna. La digitalizzazione degli aspetti organizzativi, produttivi e decisionali all’interno delle aziende ha causato e causerà anche un enorme cambiamento nel mercato del lavoro, per esempio anche attraverso un monitoraggio sempre più completo dei processi che coinvolgono i lavoratori. Il mercato globale richiede alle aziende una sempre

maggior efficienza produttiva/organizzativa e i big data rappresentano una preziosa fonte per ricavare informazioni utili. Questi dati potrebbero essere utilizzati come supporto al modello di business (per esempio l'analisi dei consumi dei clienti), migliorare gli aspetti organizzativi oppure per la gestione del personale (Dagnino, 2017a). Quest'ultimo, cioè l'utilizzo dei big data per la gestione dei dipendenti prende il nome di "People analytics" (o HR analytics) e ha l'obiettivo di supportare i dipartimenti delle risorse umane di un'azienda nelle attività di selezione e valutazione del personale. Infatti, una maggior efficienza organizzativa è permessa anche da tecniche di questo tipo che potenzialmente permettono di "costruire un team più funzionale" attraverso una "selezione dei migliori candidati, promozione dei migliori dipendenti e l'individuazione dei lavoratori improduttivi" (Dagnino, 2017b). Applicazioni di questo genere sono già realtà, per esempio l'azienda britannica "Cognisess" ha realizzato una piattaforma completa di analisi predittiva delle persone che è utilizzata dai loro clienti per valutare, reclutare e monitorare i dipendenti. Un esempio concreto d'utilizzo di essa lo descrive Altemeyer (2019) dove analizza come la piattaforma sviluppata da Cognisess sia stata usata da una importante catena alberghiera per classificare le prestazioni dei candidati durante i colloqui (attraverso la tecnica di "Computer Vision") e prevedere le loro potenziali performance nelle successive fasi del reclutamento. Uno dei motivi di utilizzo di questi strumenti è quello di cercare di ridurre la componente soggettiva durante questi processi, tuttavia l'utilizzo di questi dati e il ricorso a modelli matematici/statistici non garantisce una piena oggettività (O' Neil, 2017).

L'autore Dagnino (2017a) evidenzia alcune aree problematiche relative allo sfruttamento dei dati relativi ad un dipendente: l'origine di questi dati (spesso collezionati in contesti sensibili, come l'attività lavorativa oppure dai social network), le finalità decisionali nell'utilizzo di essi (rendere trasparenti le decisioni scaturite dall'analisi di questi dati) e infine i possibili effetti discriminatori (dovuti alle distorsioni delle stesse piattaforme oppure intenzionalmente utilizzando le informazioni ricavate dall'analisi). In questo contesto si inserisce anche una recente risoluzione del Parlamento europeo ("Implicazioni dei Big Data in termini di diritti fondamentali" - 2016/2225-INI) nella quale viene sottolineato che i "Big Data possono condurre non solo a violazioni dei diritti fondamentali dei singoli, ma anche ad una disparità di trattamento e ad una discriminazione indiretta

nei confronti di gruppi di persone con caratteristiche simili, in particolare per quanto concerne l'equità e le pari opportunità di accesso all'istruzione e all'occupazione, quando si offre un lavoro alla persona o la si valuta”.

Nel contesto italiano per questo tipo di problematiche il lavoratore è tutelato attraverso due rami legislativi distinti che si collocano su differenti livelli: attraverso lo Statuto dei Lavoratori che è subordinato al Codice della Privacy (Dagnino, 2017a). Quest'ultimo è un testo unico che riunisce tutte le normative italiane vigenti in materia di protezione dei dati personali, che è stato recentemente aggiornato con maggiori tutele grazie al regolamento europeo (UE 2016/679) conosciuto come GDPR (General Data Protection Regulation). Inoltre, negli ultimi anni anche lo Statuto dei lavoratori ha subito modifiche in alcune sue parti attraverso per esempio la riforma chiamata “Jobs Act” del 2015, in particolare è stato anche aggiornato l'articolo 4 relativo agli apparati audiovisivi e il controllo a distanza (Dagnino, 2017a). “Il comma 1 di quest'articolo vieta anche se con una diversa formulazione l'utilizzo di software di analisi dei big data con il fine esplicito di controllo dei lavoratori” (Dagnino, 2017a). Tuttavia, come spiega sempre Dagnino (2017a) le attività di People analytics potrebbero rientrare nel comma 3 del medesimo articolo “che prevede l'utilizzabilità dei dati provenienti dai controlli preterintenzionali o dagli strumenti di lavoro «a tutti i fini connessi al rapporto di lavoro»”. Il tutto comunque subordinato alle tutele previste riguardo la protezione dei dati personali e ad un eventuale intervento del Garante della Privacy. Inoltre, ad essi si aggiungono decreti e norme ad-hoc per colmare alcune lacune normative create dalla “gig-economy” come nel caso del decreto attualmente in discussione in parlamento relativo al rafforzamento delle tutele dei lavoratori delle piattaforme digitali (cioè per esempio i cosiddetti rider delle piattaforme come Just Eat, Deliveroo e Glovo).

Per quanto riguarda la questione discriminatoria, Rota (2017) evidenzia dei limiti nella legislazione attualmente vigente, sottolineando la necessità di “introdurre una forma giuridica più articolata di discriminazione che possa fronteggiare alcuni effetti della rivoluzione tecnologica” e in particolare che siano “maturi i tempi per ipotizzare una discriminazione da algoritmo (o algorithm-oriented)”. Sempre l'autrice sopracitata evidenzia anche un problema d'identificazione di eventuali responsabili nel caso di violazioni derivate dall'utilizzo di piattaforme d'analisi dati che sempre più spesso utilizzano tecni-

che avanzate di intelligenza artificiale per replicare una condotta umana. Quindi, come scritto poc'anzi la legislazione italiana (e non solo) prevede diverse tutele al lavoratore nel contesto della datafication, che tuttavia non potrà fermare questo approccio basato sui dati che rischia di produrre una deumanizzazione del lavoro.

1.4 Il ruolo degli algoritmi

La digitalizzazione e il trend della datafication permettono di generare un'enorme mole di dati da sempre più fonti eterogenee che tuttavia perde di valore se ci si ferma ad analizzare solo i singoli dati grezzi, che spesso risultano poco significativi. Molto spesso i dati vengono inseriti all'interno di algoritmi, che tuttavia si rivelano solo gli esecutori materiali di azioni predeterminate a monte. Infatti, dietro ogni algoritmo si nasconde un modello matematico o statistico che ne determina i risultati. Esso sarà eseguito all'interno di un ambiente protetto che spesso cela meccanismi opachi ed invisibili all'esterno, tranne ai loro diretti creatori (O' Neil, 2017). Tipicamente ogni modello matematico o statistico nasce da una ricerca di ottimizzazione che è strettamente legato al concetto di merito: "è ottimale premiare i più bravi e relegare nelle ultime posizioni i meno capaci" (O' Neil, 2017). Inoltre, giusto per citare un esempio, la meritocrazia è anche alla base del pensiero economico liberista che fonda le proprie idee proprio sul merito e l'individualismo.

Molto spesso questi modelli definiscono una loro realtà interna che poi utilizzano per giustificare i propri risultati, talvolta anche perdendo di vista il contesto esterno. "Per determinare se un modello funziona, in genere si cerca di eseguire un confronto con un risultato atteso scaturito da un modello informale largamente accettato, il che non garantisce una piena oggettività" (O' Neil, 2017). Non di rado infatti essi hanno bisogno di feedback esterni al sistema per cercare di adeguarsi e migliorare il proprio operato, per esempio attraverso l'analisi di errori ed eccezioni. Un esempio lampante è stato quello che ha colpito Google con il suo algoritmo per riconoscimento e catalogazione automatica d'immagini, che nel 2015 aveva catalogato per errore una foto con due persone di colore come "Gorilla" generando una profonda indignazione generale. L'azienda al tempo si era immediatamente scusata e rassicurato che il problema all'algoritmo sarebbe stato risolto

immediatamente. Tuttavia, se “si utilizzano questi cicli di feedback in modelli distorti si rischia di creare un circolo vizioso nel sistema che rafforzerà la sua realtà distopica” (O’ Neil, 2017). Un esempio che viene descritto da O’Neil (2017) sono le classifiche delle Università americane che sono stilate attraverso un modello di classificazione basato su una moltitudine di parametri, quali per esempio: percentuali di laureati, qualità delle infrastrutture, numero di immatricolazioni, reputazione e pubblicazioni dei docenti. Questi fattori sono spesso resi pubblici, il che “li rende facilmente manipolabili dai singoli atenei che attueranno delle politiche, più o meno legali, per aumentare il proprio punteggio” (O’ Neil, 2017). Soprattutto dato dal fatto che la posizione in classifica possiede una spirale auto rafforzativa, continuando a premiare i primi e penalizzare le Università nelle posizioni inferiori. “Questi tipi di modelli, se presi come riferimento e usati su larga scala, hanno l’effetto collaterale di modificare il comportamento dell’intero sistema” (O’ Neil, 2017). In precedenza, la reputazione di un’Università si basava sostanzialmente sul passaparola, che lasciava spazio alla diversità d’opinione con valide argomentazioni da più parti. Infatti, anche dove è presente questa consapevolezza di una possibile distorsione della realtà questi modelli vengono comunque usati fintanto che produrranno risultati soddisfacenti per l’utilizzatore, come servizio o profitto monetario (O’ Neil, 2017). Tuttavia, spesso si perdono di vista le persone che stanno dietro i freddi numeri.

Il principale aspetto dannoso e più criticabile nell’utilizzo dei Big data o più in generale del fenomeno della Datafication è che essi hanno la presunzione di voler avere una visione totalitaria; quantificando ogni aspetto dell’esistenza umana e non, sotto forma di numeri. Quest’ultimi verranno inseriti in sempre più complessi algoritmi, che tuttavia non sono altro che una sequenza di passi finiti scaturiti da una mente umana.

1.5 Alla ricerca di pattern

Se la datafication permette di trasformare molti aspetti della nostra vita sotto forma di dati, gli algoritmi sono uno strumento ideale per analizzare queste enormi collezioni di dati. Quest'ultime permettono infatti di avere a disposizione una preziosa fonte cronologica e storica relativa per esempio all'evoluzione di un determinato fenomeno socioeconomico. L'uso di modelli matematici o statistici permettono agli algoritmi per esempio di identificare eventuali pattern, cioè degli schemi ricorrenti. Questi pattern possono essere usati per scopi nobili come per esempio per determinare i luoghi più probabili di soccorso nel Mar Mediterraneo (UN Global Pulse, 2017a); oppure anche come descrive l'autrice O'Neil (2016) attraverso lo storico e la posizione dei crimini passati, è possibile utilizzare dei modelli statistici per calcolare la probabilità che un tipo di crimine avvenga in una determinata area geografica. Questi tipi di utilizzi sono già realtà in diversi paesi e hanno mostrato dei risultati abbastanza soddisfacenti (per esempio dal 2010 alcune forze di polizia britanniche hanno sperimentato il sistema "CRUS - Criminal Reduction Utilising Statistical" sviluppato da IBM). Tuttavia, almeno per ora non siamo in presenza di modelli che cercano di predire crimini futuri, il loro risultato sono semplici valori probabilistici basati sullo storico, che possono rappresentare un utile strumento per ottimizzare le sempre insufficienti risorse di polizia (O'Neil, 2016).

Quando tuttavia si cerca di determinare dei pattern comportamentali a livello individuale questi nobili scopi vengono a mancare. Un esempio lampante sempre secondo O'Neil (2016) sono i test di personalità che sempre più spesso vengono somministrati nei colloqui di lavoro. Questi test tipicamente sono strutturati come un elenco non troppo breve di domande a risposta chiusa. "Queste domande spesso si presentano come innocue, banali o fuori luogo ma celano un significato ben più profondo, infatti esse sono formulate secondo specifici modelli psicologici; le risposte prese in collettivo ed elaborate permettono di tracciare un profilo comportamentale e/o psicologico del soggetto" (O'Neil, 2017). Infatti, il principale scopo d'utilizzo nei colloqui di lavoro è individuare gruppi di persone che possiedono determinate caratteristiche o pattern comportamentali. Una cosa legittima, soprattutto per determinati tipi di lavori, ma che tuttavia non essendo esplicitato lo scopo del test o il valore delle domande rendono questo processo opaco e poco trasparente. Essi produrranno un risultato soddisfacente per l'azienda, ma

con il rovescio della medaglia di produrre una sofferenza nelle vittime di questo sistema, che si vedranno escluse per il solo fatto che il test li abbia catalogati all'interno di un determinato profilo comportamentale che magari li rappresentava solo in quel periodo della loro vita. In precedenza, questo processo era effettuato parzialmente dal reclutatore, che spesso aveva alle spalle studi di psicologia, quindi ancora una volta “si ricorre ai modelli formali con lo scopo di eliminare una componente soggettiva” (O' Neil, 2017). Negli ultimi anni si è anche ricorso a strumenti e procedure automatizzate per l'analisi e selezione dei curriculum vitae anche tramite l'intelligenza artificiale, il che potrà permettere un notevole risparmio di tempo e risorse escludendo chi non possiede degli specifici requisiti, ma rischia di perdere di vista la componente umana.

Parallelamente anche “durante la navigazione web possiamo ritrovarci senza saperlo all'interno di modelli che analizzano ogni nostro click, azione e scelta per determinare un pattern comportamentale o inserirci in determinate catalogazioni” (O' Neil, 2017). È proprio questa mancanza di trasparenza che determina la connotazione predatoria descritta dal neologismo “Colonialismo dei dati”. Le pubblicità mirate funzionano proprio in questo modo, il provider pubblicitario (per esempio: Google Ads o Facebook Ads) attraverso il tracciamento delle nostre attività esegue una nostra profilazione e saremo destinatari di pubblicità inerente ai nostri interessi. Visualizzeremo degli annunci in linea con i nostri interessi; niente di negativo quindi? Alcuni potrebbero non aver nulla in contrario o anzi essere soddisfatti di questo meccanismo, ma è impossibile non sollevare l'enorme questione di privacy individuale.

Capitolo 2

Il concetto di Privacy

Uno degli effetti collaterali più importanti del fenomeno della Datafication è relativo alla potenziale violazione della privacy individuale, inoltre quest'ultima è anche strettamente legata al concetto analizzato precedentemente del colonialismo dei dati. Infatti, per capire come i dati possono essere usati per il bene sociale è necessario approfondire il concetto stesso di privacy, le problematiche che possono insorgere e la percezione da parte degli utenti stessi.

2.1 La definizione

Molto spesso sia la legislazione che la letteratura si sono concentrati per lo più sulla difesa dei diritti e delle tutele riguardo la privacy individuale, tuttavia “attualmente non esiste un riconoscimento internazionale sul concetto stesso di privacy” (Cannataci, 2016). Una possibile definizione è stata esplicitata in una relazione alle Nazioni Unite (Frank La Rue, 2013) durante il “Consiglio per i diritti umani” che definisce la privacy come: *“la presunzione che gli individui dovrebbero avere un’area di sviluppo autonomo, d’integrazione e libertà personale. Possedere una propria “sfera privata” con o senza interazione con gli altri, liberi da interventi statali o da eccessivi interventi indesiderati da parte di altre persone”*.

Il gruppo di esperti del “Global Pulse Privacy Advisory” all’interno dell’innovativa iniziativa “UN Global Pulse” ha elaborato (per conto del “UN Development Group”)

alcuni principi base per la protezione dei dati e della privacy come linee guida da seguire nell'uso operativo etico dei big data nel processo di raggiungimento dei vari obiettivi di sviluppo sostenibili (SDG) prefissati nell'Agenda 2030 (United Nations Development Group, 2017). Esprimibili in questi nove punti:

1. **Utilizzo legale, legittimo ed equo:** l'accesso o la raccolta dei dati deve essere conforme alla legislazione vigente nonché ai più alti standard di riservatezza e condotta etica e morale.
2. **Utilizzo limitato allo scopo dichiarato:** il trattamento e l'utilizzo dei dati deve essere coerente e non superare lo scopo dichiarato durante la loro raccolta. Inoltre, questo scopo deve essere legittimo e definito in modo preciso e restrittivo.
3. **Riduzione dei rischi e valutazione dei benefici:** la valutazione dei rischi, dei danni e dei benefici devono essere fatte ogni qualvolta vi è un nuovo uso dei dati, tenendo in considerazione il contesto dell'utilizzo dei dati come: fattori geografici, politici e religiosi. Applicare tutte le misure disponibili per abbassare il potenziale danno fisico, emotivo o economico causati da un'eventuale violazione dei diritti della persona.
4. **Contesti e dati sensibili:** dovrebbero essere usati standard più rigorosi per la protezione dei dati personali o sensibili, anche se essi non risultano esplicitamente riconoscibili. Infatti, il contesto di raccolta dei dati può influenzare la catalogazione sul tipo di dato, come nel caso in cui vengano ricavati da una popolazione vulnerabile, persone a rischio, bambini o adolescenti.
5. **Sicurezza dei dati:** essa è fondamentale per garantire la riservatezza e la protezione durante l'intero ciclo di vita dei dati. Devono essere inoltre presenti solide garanzie tecniche, organizzative e procedurali, tenendo conto della tecnologia disponibile e del costo di attuazione di esse. I dati personali devono essere identificati e utilizzare dove possibile metodi di aggregazione, pseudonimizzazione o mascheramento per ridurre qualsiasi rischio potenziale per la privacy.
6. **Riduzione e minimizzazione dei dati:** l'accesso ai dati, l'analisi o altri usi dovrebbero essere ridotti al minimo per adempiere allo scopo indicato nella rac-

colta. La quantità di dati, inclusa la sua granularità, dovrebbe essere limitata al minimo necessario. Qualsiasi conservazione dei dati deve rispettare la legittimità dello scopo e non aggiungere componenti non necessarie. Al termine del periodo di tempo necessario per adempiere allo scopo i dati dovrebbero essere eliminati definitivamente in modo adeguato.

7. **Qualità dei dati:** tutte le attività sui dati dovrebbero essere progettate, svolte, segnalate e documentate con un adeguato livello di qualità e trasparenza. La qualità dei dati deve essere misurata attraverso parametri come: accuratezza, pertinenza, sufficienza, integrità, completezza, usabilità, validità e coerenza.
8. **Responsabilità, trasparenza e Open data:** dovrebbero essere adottati adeguati meccanismi di governance e responsabilità per monitorare la conformità con le leggi pertinenti, incluse le tutele di privacy e i più alti standard di riservatezza, morali e di condotta etica in relazione all'uso dei dati. Inoltre, gli open data sono un motore importante per l'innovazione, responsabilità e trasparenza. Pertanto, quando possibile, i dati dovrebbero essere resi accessibili, a meno che i rischi di renderli pubblici superino i benefici o ci siano altre basi legittime per non farlo.
9. **Utilizzo da parte di terzi:** l'utilizzo dei dati da parte di collaboratori di terze parti dovrebbero ricalcare le protezioni, tutele e finalità della raccolta originale. È raccomandato un processo di valutazione nell'uso dei dati concessi a terzi anche tramite accordi giuridicamente vincolanti che definiscono i parametri per l'accesso e gestione dei dati. Inoltre, il trasferimento a terzi dei dati dovrebbe essere adeguatamente segnalato e accettato durante la raccolta di essi.

Questi nove principi sono il punto di partenza per utilizzare i dati in modo etico e come supporto al bene sociale.

2.2 La percezione e le problematiche

Come analizzato precedentemente il trend della datafication può essere sfruttato per ricavare preziose informazioni, tuttavia questo processo può potenzialmente aumentare i rischi di violazione della privacy. Nello specifico si può definire questa violazione come *“un’esposizione, perdita o inferenza indesiderata riguardante informazioni private o riservate”* (Shamsi e Khojaye, 2018).

Gli autori Shamsi e Khojaye (2018) hanno elaborato i seguenti scenari in relazione ad una possibile violazione della privacy nei sistemi basati sui big data, che ho riassunto nei seguenti punti:

- **Monitoraggio dei governi:** essi attueranno delle misure di sorveglianza e monitoraggio dei dati per incrementare il livello di sicurezza (per esempio, il sistema PRISM in dotazione al NSA dal 2007) oppure per migliorare i servizi pubblici erogati;
- **Raccolta informazioni dai service provider:** queste aziende tengono traccia e profilano le azioni degli utenti per migliorare il loro modello di business con conseguenze sui profitti e collateralmente sull’esperienza degli utenti. Estrapolare valore da questi dati si rivela alquanto profittevole per queste imprese, infatti come evidenziava Arvidsson (2016) la maggior parte delle piattaforme digitali utilizza il valore ricavato dai propri contenuti come forma di autofinanziamento;
- **Attacchi di re-identificazione:** qualsiasi informazione che distingua una persona da un’altra può essere utilizzata per identificare nuovamente i dati anonimi, principalmente attraverso tre tipi di attacchi: di correlazione (formare una collezione più completa ricavata da più sorgenti di dati), arbitrari di identificazione (individuare un soggetto specifico all’interno di un set di dati) oppure con un attacco di identificazione del target (correlare diversi set di dati relativi ad un determinato individuo);
- **Violazione dei dati:** scenario relativo in caso di attacchi informatici che compromettano la sicurezza dei sistemi che possono portare alla distruzione, perdita,

alterazione, divulgazione o accesso non autorizzato ai dati archiviati. Questi attacchi possono provenire internamente attraverso un accesso fisico oppure con molta più probabilità dall'esterno attraverso una violazione remota del sistema permesso dalle moderne infrastrutture connesse. Diventa quindi fondamentale implementare misure adeguate a prevenire le intrusioni, sia a livello del sistema d'archiviazione sia per quanto riguarda le comunicazioni verso l'esterno.

Nell'ultimo decennio le connessioni tramite la rete mobile sono cresciute esponenzialmente, rendendo i dispositivi cellulari onnipresenti nella nostra quotidianità; essi conterranno una moltitudine di dati personali ricavati anche da una sempre maggiore presenza di sensori (in primis il GPS).

In questo contesto, l'associazione GSMA ha commissionato una ricerca globale (coinvolgendo 11.500 utenti mobile provenienti da 7 diverse nazioni) per esaminare le problematiche relative alla privacy degli utenti mobile e il modo in cui influenzano il loro atteggiamento verso la connettività e l'utilizzo delle applicazioni sugli smartphone (GSMA, 2014). Il primo dato che emerge da questa ricerca (GSMA, 2014) è che più dell'80% degli utenti si dicono preoccupati di condividere le proprie informazioni personali durante l'accesso a servizi o applicazioni; infatti prima dell'installazione di un'applicazione su smartphone la maggior parte degli utenti (65%) vorrebbe scoprire quali informazioni personali quest'ultima richiede e poter esercitare un'autorizzazione o un'eventuale revoca d'accesso ai dati. Quest'ultima eventualità è abilitata dalle ultime versioni dei principali sistemi operativi mobile (Android e iOS) che hanno aumentato il livello di controllo dei dati da parte degli utenti stessi. Inoltre, poco meno della metà degli utenti (48%) che si dichiarano preoccupati limiterebbero l'uso dell'applicazione a meno che non vi siano maggiori rassicurazioni sulla tutela dei loro dati. Un altro dato interessante che emerge da questo studio (GSMA, 2014) è che 8 utenti su 10 che accettano la privacy policy (informativa sul trattamento dei dati) di una specifica applicazione/servizio, dichiara di non aver letto il testo perché "troppo lunga" o con "un lessico troppo legale". A mio parere questo è un punto centrale per quanto concerne la trasparenza nel trattamento dei dati, infatti anche se è presente un obbligo legale per le aziende di redigere quest'informativa risulta nella quasi totalità dei casi poco "User-friendly". Infine, è interessante analizzare a quali soggetti l'utente si rivolge in caso vi sia stata una violazione della propria privacy,

dalla ricerca emerge infatti che gli utenti (con variazioni fra i diversi paesi) si rivolgono sia direttamente agli sviluppatori delle applicazioni, ma anche tramite i service provider, le autorità di regolamentazione, avvocati o direttamente alla polizia (Fig. 2.1).

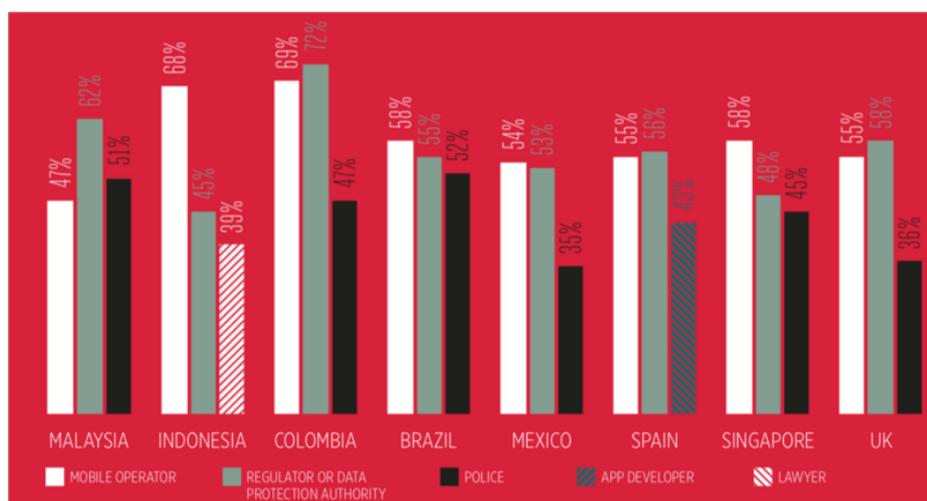


Figura 2.1: Soggetti sollecitati in caso di violazione della Privacy

FONTE: (GSMA, 2014)

Capitolo 3

Smart City – Le città al tempo dei Big data

Dopo aver introdotto il concetto della datafication è utile capire le ricadute dell'enorme mole di dati che viene generata dal contesto urbano (dispositivi IoT, smartphone, sensori, ecc.) e se sono presenti degli effetti collaterali riconducibili per esempio al Colonialismo dei dati. Per questa ragione ho deciso di approfondire il tema delle “Smart City” a partire da quali caratteristiche debba avere, le eventuali problematiche e alcune iniziative per uno sviluppo urbano più sostenibile. Infatti, la crescente urbanizzazione delle città pone delle problematiche da dover affrontare necessariamente attraverso uno sviluppo più efficiente tramite per esempio un modello di maggior sostenibilità. Problematiche quali per esempio: l'ordine pubblico, qualità dei servizi pubblici (trasporto, sanità, rifiuti, ecc.), manutenzione infrastrutture o viabilità. Nell'articolo “IBM - Building sustainable cities through partnerships and integrated approaches” del 2016 l'autore Philippe Sajhau (vicepresidente di IBM Francia e responsabile del programma IBM “Smarter Cities”) cerca di argomentare l'utilizzo tecnologico come mezzo fondamentale per questa ricerca di sostenibilità e non un utilizzo fine a sé stesso. Utilizzare la tecnologia per una sostenibilità sul lungo periodo, inoltre può comportare anche benefici visibili sul breve-medio periodo (Philippe Sajhau, 2016) come per esempio: una maggiore crescita economica regionale (investimenti locali e impiego di forza lavoro), risparmi economici per la comunità (allocazione efficiente di risorse e monitoraggio in tempo reale) e permettere benefici

tangibili per i residenti delle città (qualità della vita, servizi pubblici migliorati, ecc.).

Negli ultimi decenni è stato coniato il termine “Smart City” che definisce un concetto abbastanza generale di una città intelligente, digitale e collegata ad internet. La società IBM, più specificamente nell’ambito del programma “Smarter Cities”, ha cercato di definire tre criteri (Philippe Sajhau, 2016) per definire una città che rientri nella categoria di Smart City:

- (a) La presenza di piattaforme di dati in grado di raccogliere e aggregare una vasta quantità di dati da diverse fonti, i cosiddetti “big data”;
- (b) Avere a disposizione informazioni a livello regionale, cioè queste piattaforme di dati devono consentire di redistribuire le informazioni in modo coerente e chiaro, anche attraverso l’uso di Open data o previsioni fatte attraverso modelli predittivi;
- (c) Partecipazione (passiva o attiva) dei cittadini al processo. L’utente sarà la fonte primaria per ricavare i dati per capire le dinamiche della città, attivamente attraverso segnalazioni dirette tramite piattaforme specifiche oppure indirettamente attraverso dati estrapolati dalle loro azioni o comportamenti (utilizzo di servizi pubblici, traffico, ecc.)

Il trend per delle città sempre più smart e interconnesse è tracciato, ma non è del tutto scontato. Attualmente in Europa poche città possono contare di rientrare in questi tre criteri per ricadere in questa specifica definizione di “Smart City”. "Uno dei motivi può essere la poca consapevolezza dei funzionari eletti sui vantaggi di questo approccio che secondo la loro opinione non giustificherebbe gli investimenti necessari" (Philippe Sajhau, 2016). In contrapposizione ci sono molte città asiatiche, costruite quasi da zero a partire dagli anni ‘80/’90, che fanno uso massiccio di tecnologia e dati anche per una sorveglianza di massa. Per implementare la tecnologia all’interno della città è fondamentale anche il supporto di aziende che offrono servizi OTT (Over-the-top) come Microsoft, Vodafone, A2A Smart City o IBM, che a loro volta si appoggiano ad apparati di rete prodotti da aziende come Nokia, Huawei, Cisco o Qualcomm. La futura rete mobile 5G e i dispositivi IoT (Internet of Things) possono creare ulteriori modelli di business ed innescare un’ulteriore accelerazione nella interconnessione urbana.

Le città italiane come le loro analoghe europee hanno cercato di seguire almeno parzialmente questo trend, secondo la ricerca “ICity Rate 2018” di FPA, Milano risulta la città italiana più smart seguita da Firenze e Bologna con un ottimo posizionamento nel ranking nazionale anche per città di medie dimensioni come Trento, Bergamo, Parma e Pisa. Invece a livello europeo, secondo lo studio “The Smart City Breakaway 2019”, le città considerate più smart sono Vienna e Londra.

Inoltre, è interessante approfondire il recente progetto sperimentale “WeGovNow” realizzato nella città di Torino, Londra (Southwark) e nel comune di San Donà di Piave (VE) finanziato all’interno del Programma europeo di Ricerca e Innovazione (Horizon 2020). Questo progetto si pone come obiettivo primario quello di esplorare come il settore ICT possa essere usato per la progettazione e la costruzione di politiche e servizi pubblici in un contesto urbano che abbia caratteristiche di sostenibilità e inclusione sociale. In particolare, adottando un paradigma “we-government” per riorganizzare la pubblica amministrazione abilitando una partecipazione consapevole e condivisa sia da parte dei cittadini che dagli attori pubblici o privati (Visentin, 2019). L’obiettivo concreto di questo progetto è quello di realizzare una piattaforma che permetta al cittadino d’interfacciarsi attivamente con la pubblica amministrazione, per esempio: per creare/discutere/votare proposte o iniziative di pubblico interesse, segnalare malfunzionamenti/problematiche sul territorio o visualizzare attraverso una mappa i servizi, iniziative o eventi di pubblico interesse, ecc. In questo contesto si inserisce anche l’iniziativa “European Partnership on Smart City and Communities” (EIP-SCC) nata nel 2012 con lo scopo di creare dei cluster di città europee, aziende e soggetti della società civile per avviare una trasformazione digitale che proponga delle soluzioni sostenibili per l’ambiente, la società e la salute. Per queste ragioni l’iniziativa ha contribuito alla realizzazione di più di 370 progetti (dato aggiornato a fine 2017) nei settori dell’energia, dell’ICT e della mobilità in 10 città europee (Copenaghen, Amsterdam, Vienna, Barcellona, Parigi, Stoccolma, Londra, Amburgo, Berlino, Helsinki) (AgID, 2019).

Infine, un approccio più critico è stato sviluppato da Matthew Crawford (2017), uno scrittore americano e ricercatore presso l’Università della Virginia, dove enfatizza il pericolo che offre la possibilità di tracciare i movimenti delle persone all’interno del contesto urbano legata soprattutto alla pubblicità mirata. "Con la tecnologia onnipresente e sem-

pre più persone ultra-connesse la pubblicità urbana diventa sempre più sofisticata (e spudorata) per catturare l'attenzione dei cittadini" (Matthew Crawford, 2017). Secondo lo scrittore il problema non risiede nella tecnologia in sé, ma più ad una mera questione di economia politica dove al momento la maggior parte delle smart city non sono progettate per il bene pubblico ma per lo più sviluppate e controllate da aziende del settore ICT.

Secondo la mia opinione, questo aspetto evidenziato da Crawford (2017) risulta abbastanza vero (sia al giorno d'oggi sia per prospettive future) ma non rappresenta un vero e proprio problema o limite nello sviluppo della società. Come scritto precedentemente il modello data-driven ha creato effettivamente una vasta gamma di nuovi modelli di business (in questo caso il marketing mirato) che spesso trattano informazioni personali e sensibili; tuttavia le tutele previste dalla legislazione e procedure tecniche come l'aggregazione o l'anonimizzazione permettono di proteggere, anche se non totalmente, i cittadini.

Capitolo 4

Analisi iniziative per il bene sociale

Dopo aver analizzato il fenomeno della datafication e i relativi effetti collaterali, ora mi appresto ad analizzare alcuni progetti realizzati da iniziative internazionali che si pongono come scopo quello di un utilizzo etico dei dati per azioni umanitarie per l'intera comunità. Il colonialismo dei dati analizzato precedentemente ci ha fornito un riferimento negativo da cui queste iniziative prendono la distanza, esse infatti si pongono l'obiettivo di utilizzare l'approccio basato sui dati per implementare soluzioni etiche che abbiano ricadute per il bene sociale. La prima iniziativa che voglio analizzare prende il nome di "Global Pulse" ed è stata creata un decennio fa dalle Nazioni Unite con lo specifico scopo di esplorare scenari d'utilizzo dei big data per supportare il raggiungimento degli "Obiettivi di Sviluppo Sostenibile" (SDG) contenuti nell'Agenda 2030.

4.1 United Nations Global Pulse

Il Global Pulse è un'iniziativa di forte carattere innovativo delle Nazioni Unite con oggetto i Big data, creata nel 2009 dall'allora Segretario generale Ban Ki-moon. La visione di questo progetto è un futuro in cui i dati vengono sfruttati in modo sicuro e responsabile come bene pubblico. A tal fine, quest'iniziativa sta lavorando per promuovere la consapevolezza delle opportunità dei big data per uno sviluppo sostenibile e per scopi umanitari attraverso per esempio partnership per la condivisione di



dati pubblici-privati per generare strumenti e approcci analitici generati dalla sua rete di laboratori sparsi nel mondo (Pulse Lab). Quest'iniziativa infatti si struttura come una rete di laboratori sperimentali (attualmente a New York, Jakarta e Kampala) in cui la ricerca sui big data è coordinata da diversi esperti del settore (e non) presenti all'interno del mondo accademico, dei governi nazionali, del settore privato o in altre agenzie delle Nazioni Unite.

Una visione attuata attraverso una strategia d'azione su due binari paralleli: l'innovazione e la catalizzazione dell'ecosistema (Fig. 4.1). Un'innovazione che passa per l'utilizzo di nuovi approcci, strumenti e competenze per scoprire nuovi scenari d'utilizzo dei big data per lo sviluppo o come supporto al processo decisionale. Infine, cercare d'influenzare l'intero ecosistema attraverso il contributo nello sviluppo di quadri normativi e standard tecnici per affrontare le sfide relative alla condivisione dei dati e alla protezione della privacy; oppure fornendo alle organizzazioni del settore pubblico assistenza politica e tecnica per rafforzare le loro capacità d'analisi in tempo reale delle politiche socioeconomiche attuate.

Un possibile partner dell'iniziativa UN Global Pulse è un qualsiasi soggetto (governo, entità all'interno delle Nazioni Unite, mondo accademico o privato) che voglia esplorare scenari d'utilizzo dei big data (o più in generale del paradigma "data-driven") per supportare il raggiungimento degli "Obiettivi di Sviluppo Sostenibile" (SDG). Quest'ultimi racchiudono una serie di 17 obiettivi (raggiungibili attraverso dei target prestabiliti) concordati dall'Organizzazione delle Nazioni Unite che mirano a risolvere molteplici problematiche di sviluppo economico e sociale quali per esempio: la povertà, la fame, la salute, l'istruzione, il cambiamento climatico, l'uguaglianza sociale e di genere, l'urbanizzazione, ecc. Sono presenti 169 target da cercare di raggiungere entro la fine del prossimo decennio (2030) e sono raccolti nel documento "Trasformare il nostro mondo. L'Agenda 2030 per lo sviluppo sostenibile" che ha sostituito gli obiettivi scaduti nel 2015 che erano stati prestabiliti nel 2000 come "Obiettivi di sviluppo del Millennio".

Il ruolo del Global Pulse in una collaborazione con un partner per un progetto data-driven avviene con tre possibili livelli di coinvolgimento: implementazione su commissione (esecutore), guida e coordinamento (facilitatore) o fornendo assistenza riguardo problematiche sui dati, privacy, etica o metodologie (consulenza tecnica). Inoltre, i risultati e

la conoscenza accumulata dai laboratori sperimentali vengono condivisi attraverso delle relazioni oppure tramite documenti tecnici/accademici.

Nel 2018, i laboratori dell'iniziativa hanno organizzato più di 100 eventi in tutto il mondo tra workshop, interventi in conferenze e corsi di formazione. Sempre nello stesso anno, come dichiarato nel loro report annuale (UN Global Pulse, 2018) è stato ulteriormente aumentato il portafoglio di strumenti e metodi sviluppati per e insieme ai partner. Come per esempio "Qatalog", uno strumento di data mining che permette di estrarre informazioni dai social media e dai programmi radiofonici, già in uso in alcuni progetti in sviluppo (come per esempio in "Usare l'apprendimento automatico per analizzare i contenuti radio in Uganda"). Oppure "PulseSatellite" uno strumento prototipale ad alta precisione per l'automatizzazione del processo di rilevazione delle strutture nelle immagini satellitari per supportare le operazioni di migranti e rifugiati, sviluppato in collaborazione con l'istituto delle Nazioni Unite per la formazione e ricerca (UNITAR). Per dare un esempio del tipo di progetti seguiti da UN Global Pulse, nello scorso anno (2018) sono stati sviluppati, solo per citarne alcuni: "Esplorare gli effetti della violenza estremista nei discorsi dell'odio online" (in collaborazione con IBM Science for Social Good e l'Università di Pompeu Fabra), "Utilizzare i registri delle chiamate per comprendere l'integrazione dei rifugiati in Turchia" (in collaborazione con UNHCR) o ancora il progetto "L'utilizzo di dati della rete mobile per esplorare la risposta al disastro in Asia" (in collaborazione con Digicel); quest'ultimo è un esempio concreto di come i dati mobili possono essere utilizzati per gestire e monitorare le evacuazioni in tempo reale (UN Global Pulse, 2018).

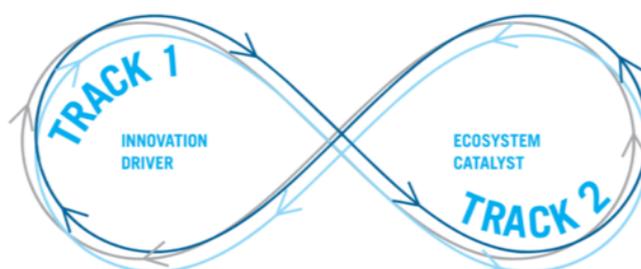


Figura 4.1: Strategie dell'iniziativa UN Global Pulse

FONTE: (UN Global Pulse, 2018)

4.1.1 REPORT: Usare i Big data per studiare i pattern di soccorso nel Mar Mediterraneo (*UN Global Pulse, 2017a*)

Il tema dei fenomeni migratori è sempre più spesso al centro dell'attenzione mediatica italiana ed internazionale. Per questo motivo e perché analizza un fenomeno che avviene a ridosso delle coste del nostro paese ho deciso di esaminare questo particolare progetto. È un importante esempio perché dimostra come l'incrocio dei dati già disponibili permetta di comprendere con più efficacia alcuni fenomeni socioeconomici come quello migratorio in questo specifico caso. Il progetto si discosta completamente dal fenomeno del colonialismo dei dati, in quanto non tratta nessun dato inerente alle persone, ma si focalizza sugli spostamenti e segnali radio generati dalle navi che sono pubblicamente accessibili.

IL PROGETTO

Il grande fenomeno migratorio che ha interessato le coste dell'Europa meridionale ha generato un significativo aumento degli sforzi di ricerca e di salvataggio nel Mar Mediterraneo, che tuttavia “nel 2016 ha fatto segnare il record negativo per il numero di morti di rifugiati e migranti che attraversavano queste acque” (IOM, 2017). Global Pulse ha collaborato con l'Alto Commissariato delle Nazioni Unite per i Rifugiati (UNHCR) riguardo un progetto che ha analizzato nuove fonti di big data per fornire una migliore comprensione del contesto di ricerca e operazioni di salvataggio nel Mar Mediterraneo.

Nello specifico il progetto ha utilizzato i dati di localizzazione delle navi (AIS) per determinare le loro rotte dall'Italia e da Malta verso le zone di salvataggio per poi successivamente incrociarli con i dati d'allarme trasmessi dalle chiamate di soccorso delle navi bloccate in mare. Attraverso il sistema automatico di comunicazione marittimo (AIS) ogni nave trasmette regolarmente una serie di informazioni sulla sua posizione (latitudine e longitudine), il tipo di nave, l'identificativo, velocità, la rotta e la sua destinazione. Questi dati sono accessibili pubblicamente da qualsiasi persona, come succede anche per l'analogo sistema aeronautico. Le chiamate di soccorso invece sono dei segnali radio in broadcast gestite tramite il “World-Wide Navigational Warnings Service” e vengono utilizzate per avvisare o richiedere aiuto verso le navi (o porti) nelle

vicinanze riguardo eventuali problemi di sicurezza o emergenze a bordo. Questi avvisi tipicamente contengono oltre la ragione dell'avviso, una stima delle persone a bordo e una posizione approssimativa della nave.

Per esempio, nell'immagine sottostante (Fig. 4.2) si può vedere le tracce del salvataggio da parte di due navi di "Medici Senza Frontiere - MSF" (Bourbon Argos e Aquarius) che tra l'1 e 6 ottobre 2016 hanno condotto otto salvataggi soccorrendo oltre 1.000 persone (secondo fonti MSF).



Figura 4.2: Spostamenti delle navi durante le operazioni di soccorso
FONTE: (UN Global Pulse, 2017a)

Il risultato di quest'analisi ha mostrato un cambiamento nel modello dei segnali di soccorso nel tempo. Le segnalazioni nel periodo esaminato furono registrate sempre più vicino al confine libico, costringendo le operazioni di salvataggio ad avvicinarsi sempre di più alle coste libiche ed estendere le zone di soccorso prestabilite inizialmente (UN Global Pulse, 2017a). Questo progetto ha mostrato l'importanza di avere a disposizione dati accessibili pubblicamente (open data) per poi poterli utilizzare per studi o analisi che possano essere usati anche per scopi umanitari come per esempio in questo caso migliorare l'efficienza e il coordinamento dei soccorsi in mare.

4.1.2 REPORT: Utilizzo dei dati mobili per il bene sociale: successi, sfide e sviluppi futuri (*UN Global Pulse, 2017b*)

Ho scelto questo progetto perché trattava una tematica che è cresciuta esponenzialmente nell'ultimo decennio, cioè l'utilizzo dei dati generati dalla rete cellulare. Data la dimensione di questo fenomeno risulta quindi importante capire come utilizzare in modo etico questi dati e individuare le eventuali problematiche e sfide future. Anche in relazione ad un'ulteriore accelerazione futura di questo settore dovute all'introduzione di nuove tecnologie come la rete 5G e i dispositivi IoT. È importante analizzare l'utilizzo di questi "mobile big data" perché potenzialmente presentano aspetti del colonialismo dei dati, infatti ogni nostra interazione con la rete cellulare è registrata ed è possibile estrarre delle informazioni all'insaputa dell'utente stesso. Per questa ragione lo studio analizza questo tipo di problematiche ed espone degli scenari d'utilizzo etici e sociali per questi dati generati dai dispositivi presenti sulla rete cellulare.

LO STUDIO

Nel 2017 il Global Pulse in collaborazione con GSMA (con il supporto di "Vital Wave Consulting" e "Vodafone Americas Foundation") ha cercato di esaminare il potenziale valore nell'utilizzo dei dati della rete mobile per il bene sociale. Questo studio (UN Global Pulse, 2017b) denominato "La situazione dei dati mobili per il bene sociale" aveva l'obiettivo di esaminare lo scenario attuale, valutare gli eventuali ostacoli e delineare delle raccomandazioni per gli sviluppi futuri. Un'analisi per esempio effettuata attraverso una serie di interviste di esperti del settore, alcuni progetti ad-hoc oltre che basarsi sull'attuale letteratura presente (ricerche scientifiche, studi di fattibilità, ecc.). Capire il reale potenziale (e problematiche) nell'utilizzo dei dati mobili è la chiave della rivoluzione basata sui dati; infatti la diffusione, ormai totale, della rete mobile "ha permesso un profondo cambio di paradigma di come le persone accedono alle informazioni, utilizzano un servizio o comunicano fra di loro" (UN Global Pulse, 2017b). Inoltre, l'utilizzo di questa rete mobile genera (anche passivamente) una grande quantità di informazioni (mobile big data) che possono avere un enorme valore sia monetario ma anche per un bene sociale. Tuttavia, questo potenziale valore economico ha fatto sì che nascessero

modelli di business basati sulla collezione e l'analisi di questi preziosi dati mobili che determinano l'aumento dei rischi nell'uso improprio di essi.

Per questa analisi il Global Pulse definisce e prende in esame i dati mobili come elementi di informazioni contenuti nell'elenco dei dettagli delle chiamate e i CDR (Call detail record). Questi CDR vengono generati ad ogni comunicazione con la torre radio e contengono preziose informazioni come: numero di telefono del mittente e ricevente, data e ora della chiamata, la stazione radio d'entrata e d'uscita e il tipo di trasmissione (es: chiamata, SMS). Anche se tutta la parte di connettività Internet ricopre una notevole importanza come sorgente di dati di valore, essa non è stata trattata in questo studio dell'iniziativa. La ragione principale, secondo la mia opinione, è data dal fatto che il traffico internet generato in mobilità è inserito alla pari di tutto il traffico web e quindi meritevole di un'analisi con metodologie e strumenti diversi. Infatti, già con i soli dati CDR, dopo che sono stati anonimizzati e aggregati, possono fornire una serie di informazioni utili (per organizzazioni con fini umanitari o di sviluppo) quali relative a:

1. *Mobilità*: poter tracciare gli spostamenti attraverso le diverse torri cellulari. Essi possono rivelare i movimenti a livello di comunità/popolazione che risultano utili per esempio a seguito di catastrofi naturali, epidemie o anche utilizzati per una pianificazione urbana più efficiente. Come per esempio, nel progetto analizzato da Elias ed altri (2016) che utilizza l'analisi dei dati mobili per migliorare la pianificazione del trasporto pubblico;
2. *Interazione sociale*: ricavare le informazioni su come gruppi di individui interagiscono con la loro comunità (chi contattano, con quale frequenza, ecc.) per comprendere il comportamento e le tendenze socioeconomiche (per esempio nel progetto UN Global Pulse: "Using call detail records to understand refugee integration in turkey");
3. *Attività economica*: utilizzare lo schema delle ricariche telefoniche e il consumo di servizi a pagamento per estrapolare informazioni sulla salute dell'economia locale e della sua resilienza. Con quest'ultima si intende "la capacità di paesi, comunità e famiglie di gestire il cambiamento mantenendo o trasformando gli standard di vita

di fronte a shock o stress senza comprometterne le prospettive di lungo termine” (Sturgess, Sparrey e DFID, 2016).

Lo studio analizza in particolare il caso d’uso relativo al monitoraggio della migrazione umana a seguito del terremoto del 25 Aprile 2015 in Nepal (magnitudo 7.8). In questo caso è stata fatta la stima dei movimenti della popolazione nazionale derivati dall’analisi dei dati CDR anonimizzati dei 12 milioni di clienti dell’operatore telefonico locale “N-Cell”. Questo ha permesso un approfondimento in tempo reale sulla mobilità della popolazione consentendo una più efficace organizzazione dei soccorsi e per il dislocamento di risorse umanitarie.

Uno degli aspetti evidenziati da UN Global Pulse (2017b) in questo studio è il fatto che “nonostante l’effettiva dimostrazione (tramite ricerche e progetti pilota) di capacità di estrazione di conoscenza dai dati mobili, essi al momento siano poco integrati nei programmi dei relativi stakeholder (istituzioni pubbliche, organizzazioni internazionali, università e il settore privato)” perché questi soggetti sono ancora in attesa di risultati e prove più concrete che dimostrino vantaggi misurabili (con caratteristiche di sostenibilità e scalabilità) per giustificare un eventuale investimento.

Iniziative come la già citata “UN Global Pulse” e quelle più recenti come “GSMA Big Data for Social Good” e “IBM Science for Social Good” sono un fondamentale stimolo e punto di partenza necessario per comprendere il reale potenziale nell’utilizzo dei dati come supporto per il bene sociale e non solo come modello di business. Inoltre, la privacy e l’applicazione di standard etici dovrebbero essere una priorità per tutti i soggetti coinvolti. Infatti, senza queste tutele si rischia poca trasparenza e mancate responsabilità da parte dei soggetti che forniscono, analizzano e usano i dati generati dagli utenti (UN Global Pulse, 2017b).

Sempre secondo questo studio il fattore privacy è quello che ha contribuito maggiormente alla riluttanza nello sfruttamento dei dati mobili, sia per rischi legati direttamente ai consumatori (possono subire danni emotivi, fisici o economici nel caso la loro identità venga esposta); ma anche gli operatori di rete (MNO) saranno cauti nello sfruttamento (o cessione) di questi dati per timore di possibili azioni legali o danni reputazionali nel caso di violazione della riservatezza dei propri clienti (UN Global Pulse, 2017b).

Le pratiche di anonimizzazione e di aggregazione diventano quindi strumenti essenziali per lo sfruttamento etico ed efficiente di questi dati per scopi sociali e umanitari. Un'altra problematica legata all'analisi dei dati mobili sono le possibili distorsioni date dal fatto che la base clienti di uno specifico operatore di rete può non rappresentare con accuratezza l'intera popolazione, per esempio a livello demografico, geografico o economico (UN Global Pulse, 2017b).

Nel medesimo studio il Global Pulse (2017b) redige un elenco di raccomandazioni e consigli per le sfide e sviluppi futuri relative all'uso dei dati mobili per il bene sociale. Vengono identificati i seguenti punti:

- (a) **Individuare e costruire modelli di business sostenibili:** attualmente la maggior parte dei progetti sviluppati si basano su dati concessi senza alcun costo (in forma di donazione o per scopi sperimentali);
- (b) **Mitigare i potenziali rischi e affrontare le lacune relative alla protezione dei dati:** seguendo specifiche tutele e norme oppure utilizzando strumenti e metodologie per ridurre i rischi al minimo;
- (c) **Sviluppare capacità nei posti giusti:** formare una forza lavoro competente nelle questioni tecniche relative al mondo dei dati e formare un gruppo dirigente che sappia avere un ruolo di collegamento tra le tematiche tecniche, gli operatori di rete e i responsabili delle decisioni a livello locale;
- (d) **Creare strumenti globali per il bene pubblico:** attraverso piattaforme modulari e adattabili oppure utilizzando strumenti open-source per l'analisi dei dati e per valutare il rapporto tra costi e benefici.

Questo studio sull'utilizzo dei dati mobili conclude sostenendo che dopo le prove concrete di fattibilità ottenute nell'uso di questi dati tramite progetti ad-hoc o a breve termine, ci si stia preparando per un'implementazione più sistematica e istituzionalizzata. Un'implementazione che si concentrerà sull'impatto concreto nello sviluppo sociale e che dovrà essere scalabile, replicabile e sostenibile.

4.1.3 REPORT: Monitorare l'impatto delle crisi economiche sui crimini (UNODC, 2012)

Utilizzare un approccio basato sui dati si rivela utile anche per capire con più efficacia le relazioni tra vari fenomeni socioeconomici che si presentano a livello nazionale e internazionale. Questo progetto in particolare cerca di incrociare diversi set di dati per ricercare una correlazione tra la crisi economica e la criminalità all'interno di determinati paesi. Anche se intuitivamente si potrebbero collegare questi due fenomeni, questa correlazione non è facilmente dimostrabile; inoltre essi risultano fenomeni molto complessi e formati da una moltitudine di fattori diversi. L'utilizzo dei big data permette di affrontare queste sfide con metodi empirici, anche cercando di quantificare con una certa precisione questi fenomeni. Questo studio risulta un esempio virtuoso di analisi dati, poiché sfrutta prettamente statistiche e parametri a livello macro-regionale e quindi non presenta nessun elemento di colonialismo dei dati. Infine, un'altra motivazione che mi ha portato a selezionare questo particolare progetto è che nei paesi presi in esame era presente anche l'Italia.

IL PROGETTO

Nel 2012 il team del laboratorio del Global Pulse di New York ha stilato una sintesi dei risultati conseguiti in diversi progetti nel periodo 2010-2011 che ha coinvolto il lavoro di undici agenzie delle Nazioni Unite. Ho voluto selezionare e riassumere uno di questi che mi ha colpito particolarmente. Questo progetto sviluppato dall'ufficio delle Nazioni Unite "Drugs and Crime" (UNODC) ha l'obiettivo di provare a ricercare correlazione (tra i dati disponibili) tra crisi economiche e criminalità, tenendo comunque in considerazione che questi due fenomeni sono estremamente complessi con una moltitudine di fattori diversi. Le teorie sulla motivazione criminale, compresa la teoria della deformazione, propongono che "comportamenti illeciti siano causati, almeno in parte, da frustrazioni indotte strutturalmente dal divario tra aspirazioni e aspettative e il loro raggiungimento nella pratica" (UNODC, 2012). "Laddove la crisi finanziaria si manifesta attraverso una crescita economica negativa e una disoccupazione diffusa, un gran numero di persone può subire gravi e improvvise riduzioni del reddito o della stabilità economica" (UNODC,

2012). Questo, a sua volta, ha il potenziale per causare un aumento della percentuale della popolazione con una motivazione più alta per identificare soluzioni illecite ai loro problemi immediati (UNODC, 2012). L'UNODC raccoglie regolarmente dati statistici sulla criminalità e giustizia penale a livello globale; per questo progetto essi hanno preso in considerazione le statistiche di criminalità e i dati economici degli ultimi 20 anni di quindici paesi (Argentina, Brasile, Canada, Costa Rica, El Salvador, Italia, Giamaica, Lettonia, Mauritius, Messico, Filippine, Polonia, Thailandia, Trinidad-Tobago e Uruguay) scelti in base ad una serie di fattori tra i quali il livello di criminalità complessivo, l'impatto della crisi finanziaria del 2008/2009 e sull'effettiva disponibilità di statistiche criminali aggiornate ad alta frequenza.

Il rapporto evidenzia delle prove di una relazione tra fattori economici e criminalità: quando le condizioni economiche peggiorano, il crimine aumenta. Infatti, l'82% dei paesi analizzati (Fig. 4.3) mostrano una relazione tra almeno un fattore economico ed un reato di omicidio intenzionale, rapina o furto d'auto (UNODC, 2012). Inoltre, 11 dei 15 paesi esaminati mostrano cambiamenti economici significativi nel periodo della crisi economica del 2008/09. È stato anche notato che in un numero significativo di paesi (7/8 tra gli 11) il cambiamento economico durante la crisi è coinciso con un aumento dei crimini più violenti (in alcuni casi il numero di rapine sono raddoppiate, in altri sono aumentati gli omicidi o furti). Da segnalare che in alcuni casi non è stato possibile collegare un aumento di criminalità ad un fattore economico, ma non c'è stato nessuno caso in cui si è osservata una diminuzione del crimine durante un periodo di stress economico. È interessante anche analizzare l'andamento dell'Italia: lo studio evidenzia un aumento nel nostro paese delle rapine e furti d'auto in relazione al periodo della crisi economica del 2008/09 la quale aveva comportato una diminuzione del reddito reale della popolazione.

Questo modello statistico nei contesti esaminati ha anche evidenziato un intervallo medio pari a circa quattro mesi e mezzo tra i cambiamenti della variabile economica e il conseguente impatto sui livelli di criminalità. L'utilizzo di questo strumento risulta utile nello stimare (con una certa precisione) una possibile variazione del crimine con un preavviso di qualche mese, utile per esempio nel dispiegamento preventivo di forze di polizia.

Country	Geographic unit	Economic crisis (on visualization)	Crime type affected by economic crisis (on visualization)	Economic indicator identified as predictor of crime change (by statistical model)
Argentina	Buenos Aires	✓ (in 2002)	-	Share price index
Brazil	National	✓	n/a	Share price index, unemployment rate
	Rio de Janeiro		Robbery, motor vehicle theft	Treasury bill rate
	Sao Paulo		Homicide, robbery	Male unemployment rate, currency per SDR
Canada	National	✓	n/a	Treasury bill rate, unemployment rate, share price index, deposit rate
Costa Rica	National	✓	Robbery	-
El Salvador	National	✓	Homicide	-
Italy	National	✓	Robbery, motor vehicle theft	Real income
Jamaica	National	✓	Homicide, robbery	-
Latvia	National	✓	-	Youth unemployment rate
Mauritius	National	✓	-	Real income, currency per SDR
Mexico	National	✓	Robbery, motor vehicle theft	Male unemployment
Philippines	National	-	-	Deposit rate, share price
Poland	National	-	-	Treasury bill rate
Thailand	National	✓	Motor vehicle theft	Unemployment rate, real income
Trinidad and Tobago	National	✓	-	Real income, lending rate
Uruguay	Montevideo	-	-	GDP

Figura 4.3: Correlazione tra fattori economici e criminalità
FONTE: (UNODC, 2012)

4.2 GSMA - Big Data for Social Good

Come esposto nel capitolo precedente un approccio basato sui big data è estremamente utile per analizzare dei fenomeni socioeconomici attraverso dei metodi empirici, come nel caso dell'incrocio dei dati generati dalle navi presenti nell'Mar Mediterraneo che permette, oltre una migliore comprensione del fenomeno migratorio, di ottimizzare le operazioni di salvataggio dei naufraghi e quindi potenzialmente di incrementare il numero di persone salvate.

In questa sezione ho scelto di analizzare alcuni progetti all'interno dell'iniziativa "Big Data for Social Good" creata nel 2017 dall'associazione GSMA. Quest'ultima è un'organizzazione nata nel 1995 che rappresenta gli interessi degli operatori di rete mobile (MNO) in tutto il mondo, con più di 750 operatori mobili membri GSMA ed altre 300 aziende associate operanti nell'ecosistema mobile. Come descritto



precedentemente, approfondire l'utilizzo dei dati generati dalla rete cellulare è di fondamentale importanza per capire i benefici o eventuali problematiche. Soprattutto dato dal fatto che i "mobile big data" potenzialmente presentano aspetti del colonialismo dei dati. Proprio per questo motivo ho deciso di approfondire i progetti di questa iniziativa, che cercano di utilizzare questi dati come supporto per azioni sociali e umanitarie, quindi allontanandosi dal riferimento negativo del colonialismo dei dati.

L'iniziativa "Big Data for Social Good" mira a sfruttare il potere dei "mobile big data" per aiutare le agenzie pubbliche e le organizzazioni non governative (ONG) a combattere epidemie, catastrofi naturali e l'impatto ambientale. Sfide come malattie infettive, inquinamento, terremoti, inondazioni e altri disastri sono tra le più critiche che il mondo deve affrontare oggi e gli operatori della rete mobile sono in una posizione unica per aiutare le istituzioni a rispondere in modo efficace ed efficiente alle emergenze sfruttando le preziose informazioni in loro possesso. Infatti, i provider mobili raccolgono una vasta quantità di dati complessi e grezzi ed attraverso un'analisi avanzata possono fornire preziose informazioni e strumenti per delle soluzioni a problemi concreti. Inoltre i "mobile big data" possono rafforzare e accelerare i progressi rispetto agli "Obiettivi di sviluppo sostenibile" (SDG).

L'associazione GSMA ha il ruolo di intermediario e offre una piattaforma unica per: collaborare attraverso l'incrocio dell'offerta e domanda, stabilire un quadro comune, delineare un approccio sostenibile oltre che supervisionare il rispetto e la protezione della privacy individuale (e collettiva). Una supervisione della privacy attuata per esempio tramite il documento "I principi etici dell'AI" che rappresenta una linea guida nell'utilizzo dell'intelligenza artificiale per processare i dati nei vari progetti in collaborazione. Il documento delinea principi come: rispettare i diritti umani, non arrecare danni agli individui, garantire equità, accessibilità e benefici alla comunità, non utilizzare dati distorti o discriminatori, garantire trasparenza (in tutte le fasi del processo), tutelare la privacy e l'etica come design.

Attualmente (2019) aderiscono all'iniziativa "Big Data for Social Good" circa 20 operatori di rete operanti in più di 124 mercati, quali per esempio: Orange, SK Telecom, Telefónica, Vodafone e Telenor e T-Mobile. Con l'assistenza, il supporto e la consulenza di organizzazione come: UN Globale Pulse, UNICEF, OCHA e Digital Impact Alliance. Come altre iniziative di questo genere, essa mira a creare progetti ad-hoc per determinati contesti, tuttavia non è da dimenticare che l'obiettivo finale è quello di mostrare le potenzialità di soluzioni di questo genere che successivamente, se avranno caratteristiche di scalabilità, sostenibilità e modularità, potranno essere implementate su larga scala. L'associazione GSMA infatti sostiene che l'utilizzo dei mobile big data, ma personalmente allargato anche all'approccio basato sui dati (data-driven), che permetta un doppio vantaggio: migliorare le capacità governative (uso efficiente di budget e migliorare la capacità di misurare l'impatto e risultati di politiche socioeconomiche) e incrementare le opzioni offerte alla popolazione (migliore salute ed efficienza sanitaria, gestione emergenze con analisi in tempo reale).

Nello schema sottostante (Fig. 4.4) vengono mostrate le principali milestone nel processo d'implementazione di un tipico progetto seguito dall'iniziativa: si parte dalle fasi iniziali (definizione del problema e degli obiettivi, approfondimento di aspetti etici, legali ed economici) per poi proseguire con la progettazione tecnica (accordi formali, sviluppo tecnico e analisi funzionamento) per passare infine alla validazione dell'intero progetto con la successiva implementazione funzionale e attiva come strumento inserito nel contesto identificato nelle fasi preliminari.



Figura 4.4: Tipica implemetazione di un progetto all'interno dell'iniziativa GSMA

FONTE: (GSMA Digital Toolkit - The Ecosystem)

bigdatatoolkit.gsma.com

4.2.1 REPORT: Utilizzare l'analisi mobile in tempo reale come aiuto per il salvataggio di persone scomparse (*GSMA BD4SG, 2019a*)

Uno degli aspetti più delicati nell'utilizzo dei dati provenienti dalla rete cellulare è data dalla natura stessa dei dispositivi da cui sono generati. Nell'ultimo decennio le connessioni tramite la rete mobile sono cresciute esponenzialmente, rendendo i dispositivi cellulari onnipresenti nella nostra quotidianità; essi contengono una moltitudine di dati personali ricavati anche da una sempre maggiore presenza di sensori. La posizione geografica del dispositivo è uno di questi dati sensibili, ma al tempo stesso risulta uno fra i più utili. Questo particolare progetto utilizza la posizione dei vari dispositivi presenti in una determinata area geografica come supporto per le operazioni di ricerca di una persona scomparsa. Esso utilizza un approccio etico e lontano dal concetto del colonialismo dei dati, in quanto utilizza dati aggregati e solo per lo scopo finale di aiutare l'intera comunità.

IL PROGETTO

Le probabilità di ritrovamento di una persona scomparsa dopo 48 ore si riducono drasticamente (anche del 90%); quindi un intervento tempestivo ed efficiente risulta essenziale soprattutto in territori molto vasti e variegati come per esempio la Russia.

Proprio in questo contesto, grazie all'iniziativa "Big Data for Social Good", nasce l'accordo tra l'operatore di rete russo "MegaFon" e l'associazione sempre russa "LizaAlert" (GSMA BD4SG, 2019a). Quest'ultima, nata nel 2010, si



pone come obiettivo la ricerca e salvataggio di persone scomparse, coordinando centinaia di volontari durante le ricerche (a supporto delle forze di polizia locali). Solo lo scorso anno (2018) l'associazione "LizaAlert" ha ricevuto quasi 14.000 domande di assistenza e in ben 79% di questi casi si è risolto con il ritrovamento della persona scomparsa (GSMA BD4SG, 2019a).

Per supportare le attività di ricerca l'azienda "MegaFon" ha sviluppato un apposito algoritmo basato sull'analisi dei mobile big data per allertare rapidamente un gruppo di potenziali testimoni via SMS (o MMS), invitandoli a condividere ogni possibile informazione utile con l'associazione "LizaAlert". L'algoritmo sfrutta i dati già in possesso dell'azienda per determinare, tramite l'identificazione della torre radio, le persone che si trovavano nelle vicinanze o frequentavano spesso il luogo dove è scomparsa la persona. In aggiunta, vengono anche inclusi gli ultimi numeri contattati di recente dalla persona scomparsa. L'operatore di rete per tutelare maggiormente la privacy individuale memorizza solo il testo del messaggio inviato e il numero aggregato degli SMS spediti, eliminando definitivamente l'elenco delle persone selezionate dopo l'invio. Inoltre, potranno essere contattati solo i clienti che precedentemente avevano rilasciato il consenso per ricevere comunicazioni dirette dalla società. Per sfruttare questo algoritmo l'azienda ha realizzato una piattaforma web dove una persona autorizzata dell'associazione può creare una segnalazione contenente i dati (ed eventualmente una foto) della persona scomparsa (ultima posizione conosciuta, descrizione, età, nome, ecc.) che serviranno anche per formare il testo del messaggio da recapitare.

L'associazione "LizaAlert" dichiara che in questi primi sei mesi d'utilizzo il sistema è stato usato per 250 ricerche di persone scomparse, avvisando in media circa 2.000 persone per ogni segnalazione e che in più del 30% di questi casi è stata ricevuta almeno una chiamata con informazioni utile nella ricerca (GSMA BD4SG, 2019a).

4.2.2 REPORT: Costruire comunità resilienti agli estremi climatici (GSMA BD4SG, 2018)

Il cambiamento climatico è uno dei temi più importanti del nostro secolo, l'accelerazione negli ultimi 40 anni di questo fenomeno ha innescato un aumento di fenomeni naturali estremi, come tempeste, cicloni e frane. Risulta quindi importante capire le conseguenze di questi fenomeni sui territori e le implicazioni sulla vita delle popolazioni residenti. Questo progetto cerca di dimostrare come l'analisi dei dati mobili possa aiutare a comprendere meglio le migrazioni interne dovute ai cambiamenti climatici che sono caratterizzati da fenomeni naturali sempre più estremi. In questo caso vengono utilizzati dati macro-regionali aggregati e anonimizzati con il fine ultimo di comprendere e quantificare questa migrazione e aiutare il territorio colpito da questi fenomeni estremi. Per cui, siamo in presenza di un utilizzo etico di questi dati, che si allontana quindi dal colonialismo dei dati.

IL PROGETTO

Gli spostamenti all'interno dei territori nazionali causati dalla variabilità climatica stanno diventando un problema silenzioso su scala globale; nell'ultimo decennio si stima che ogni anno in media 25,5 milioni di persone siano costrette a sfollare a causa di catastrofi naturali (frane, tempeste, cicloni, inondazioni, ecc.). In questo scenario, l'America Latina risulta una delle regioni più fragili e vulnerabili riguardo l'impatto della variabilità climatica; si stima per esempio che entro il 2050 circa 17 milioni di cittadini potrebbero essere costretti a lasciare le loro case a causa della lenta ma costante evoluzione dei cambiamenti climatici caratterizzati da fenomeni sempre più estremi (GSMA BD4SG, 2018). I dati riguardanti questo fenomeno di migrazione interna sono difficili da quantificare o collezionare con precisione, soprattutto in paesi a basso-medio reddi-

to dove con tutta probabilità questi sfollati vivranno in località remote e/o rurali con scarse infrastrutture e situazioni socioeconomiche instabili (GSMA BD4SG, 2018). La difficoltà nel reperire questi dati fa sì che questo fenomeno risulti per lo più invisibile allo Stato o alle organizzazioni internazionali rendendo molto complesso l'utilizzo di modelli migratori per comprendere il fenomeno.

Proprio per cercare di rispondere a questa lacuna d'informazioni nasce la collaborazione, permessa dall'iniziativa creata da GSMA, tra l'organizzazione delle Nazioni Unite per l'alimentazione e l'agricoltura (FAO) e l'operatore mobile "Telefónica" (attraverso la sua sezione interna "LUCA" specializzata in soluzioni basate sui Big data e AI). L'obiettivo principale di questo progetto è quello di sfruttare l'enorme mole di dati della rete mobile (mobile big data) per misurare e mappare questa migrazione interna attraverso la creazione di modelli di mobilità, utilizzati come proxy data per capire il comportamento umano. Con "proxy data" si intende l'utilizzo di dati indiretti per effettuare una correlazione con il dato d'interesse originale, che si presume non conosciuto o non determinabile. Un'informazione più completa e di qualità soprattutto come supporto alle agenzie e responsabili politici, può garantire una più efficacia e tempestività nell'adottare misure di protezione sociale, come per esempio sostegno alle economie locali, aiuti ai cittadini e gestione delle emergenze. Questa collaborazione si è concentrata sullo stato della Colombia, che tra i paesi del Sud America risulta quello con il tasso più elevato di insorgenza di importanti eventi naturali; si stima infatti che l'84% della popolazione sia vulnerabile all'esperienza di due o più eventi climatici estremi (GSMA BD4SG, 2018).

In particolare, l'operatore di rete "Telefónica" ha utilizzato i dati (aggregati e anonimizzati) del dettaglio delle chiamate (CDR) dei suoi 11 milioni di clienti della Colombia combinati ad una serie di dati provenienti da fonti pubbliche o governative, come per esempio attraverso la mappa elaborata dal "Sistema d'Informazione Ambientale della Colombia - SIAC" che calcola il potenziale impatto dei cambiamenti climatici sul territorio nazionale fino all'anno 2040. Per questo scopo l'azienda ha utilizzato la loro piattaforma interna "SmartSteps" per analizzare la lista delle attività mobili per creare un profilo di mobilità per ogni linea telefonica. Nello specifico il progetto prende in esame i dati storici del 2017 riguardo la popolazione della regione colombiana di "La Guajira" per evidenziare i loro spostamenti durante il periodo di siccità che ha caratterizzato la regione in

quello stesso anno. Il risultato di quest'analisi evidenzia che circa 12.000 persone hanno dovuto abbandonare le proprie abitazioni per dirigersi verso molte zone urbane all'interno del paese (Fig. 4.5). Commentando il progetto, l'organizzazione FAO evidenzia come queste informazioni complete e dettagliate possano, oltre che migliorare la comprensione del fenomeno migratorio, aiutare la comunità per attuare politiche agricole più sostenibili (sviluppo e gestione più efficace delle risorse naturali) e sostenere pratiche d'inclusione finanziaria per i cittadini vulnerabili nelle zone rurali (GSMA BD4SG, 2018).

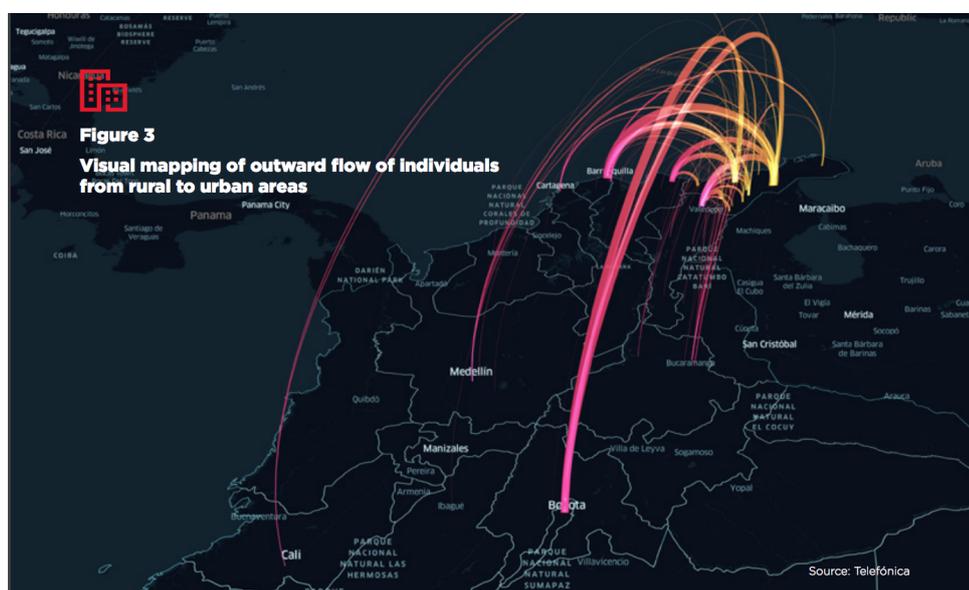


Figura 4.5: Migrazione della popolazione della regione colombiana di “La Guajira”
FONTE: (GSMA BD4SG, 2018)

4.2.3 REPORT: Rispondere rapidamente ed efficacemente alle catastrofi in Giappone (*GSMA BD4SG, 2019b*)

Un approccio basato sui dati risulta un pratico strumento di supporto per gestire le emergenze nel caso di catastrofi. Questo particolare progetto si pone l'obiettivo di realizzare una piattaforma basata sull'intelligenza artificiale per aggregare molte sorgenti di dati eterogenei generati da un contesto urbano come tramite i cellulari, sensori o dispositivi IoT. Infatti, al giorno d'oggi sono presenti una moltitudine di dispositivi che sono in grado di generare dati, che tuttavia spesso risultano difficili da integrare fra loro. Ho selezionato questo particolare progetto poiché è un esempio pratico di come i big data e l'aggregazione di tipi eterogenei di dati possono rivelarsi utili per gestire situazioni d'emergenza. Un altro aspetto importante di questo progetto è dato dal fatto che è uno dei pochi progetti per il bene sociale che si appresta ad essere adottato su larga scala.

IL PROGETTO

L'isola giapponese risulta compresa in una zona caratterizzata da un'alta frequenza di terremoti ed eruzioni vulcaniche rispetto ad altre parti del mondo, cioè è maggiormente incline a catastrofi naturali che si manifestano con sempre più violenza (*GSMA BD4SG, 2019b*). Tuttavia, le risorse per la gestione di queste emergenze sono le medesime o addirittura inferiori rispetto al passato, risulta quindi essenziale trovare dei metodi economici ed efficienti per rafforzare la risposta in caso di catastrofi naturali e proteggere i cittadini (*GSMA BD4SG, 2019b*).

Proprio con questo obiettivo, grazie all'iniziativa GSMA "Big data for Social Good" è nata una collaborazione intersettoriale tra l'operatore telefonico giapponese KDDI, Toyota e OYO che mira a dimostrare l'uso dei dati mobili e dell'intelligenza artificiale come supporto per una risposta efficiente del governo giapponese nel caso di emergenze o catastrofi naturali. Nello specifico, KDDI supportata dalle altre due aziende ha sviluppato un sistema di supporto decisionale intelligente (IDSS) che tramite l'uso dell'intelligenza artificiale integra e riassume all'interno di una piattaforma le informazioni provenienti da un'ampia gamma di fonti eterogenee come sensori e dispositivi IoT connessi alla rete

mobile. Il sistema basato sull'utilizzo di molti dati eterogenei ed elaborati dall'intelligenza artificiale permette di aumentare la precisione e velocità nel processo decisionale di organi governativi durante le situazioni emergenziali, per esempio diramando avvisi immediati, dislocare efficacemente risorse ed aiuti, regolamentare la viabilità durante le evacuazioni o identificare i luoghi sicuri per i cittadini.

Nell'immagine sottostante (Fig. 4.6) viene mostrata una schermata d'esempio del sistema, in questo particolare caso si può vedere (in alto a destra) un messaggio del bot (basato sull'intelligenza artificiale) che notifica il pericolo imminente riguardo una zona designata come centro d'evacuazione. Questo tipo d'informazione può supportare il governo a decidere se avviare preventivamente una fase di evacuazione e dispiegare una squadra di soccorso.

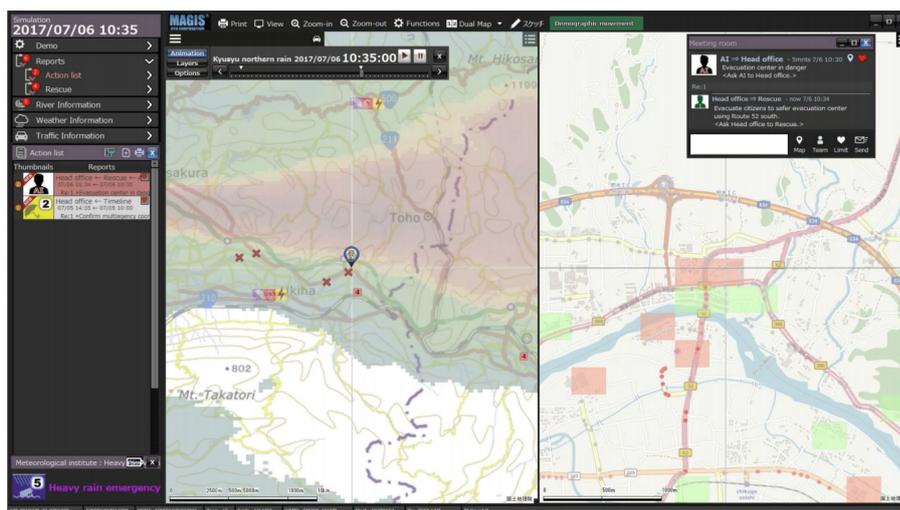


Figura 4.6: Sistema di supporto decisionale intelligente utilizzato in Giappone
FONTE: (GSMA BD4SG, 2019b)

Questa collaborazione è nata nel 2018 con lo scopo di realizzare un progetto pilota e nel corso di quest'anno (2019) sarà ulteriormente migliorato ed integrato in circa 1.700 città giapponesi per rafforzare la resilienza locale e nazionale (GSMA BD4SG, 2019b). Come altri progetti di questo tipo, l'utilizzo dei big data o più nello specifico dei dati della rete mobile permette un'accelerazione e miglioramento nelle capacità dei governi nell'attuare misure di prevenzione per i cittadini o contenimento nelle fasi emergenziali.

4.3 IBM - Science for Social Good

Come analizzato nella sezione precedente i dati generati dai dispositivi presenti sulla rete cellulare se usati in modo etico e per il bene sociale possono rivelarsi un valido strumento per comprendere e gestire con più efficacia situazioni emergenziali.

Durante l'approfondimento dell'organizzazione "UN Global Pulse" sono venute a conoscenza di un'altra iniziativa leggermente diversa, ma con i medesimi obiettivi finali. All'interno dei laboratori di ricerca (IBM Research) dell'azienda IBM è stata infatti creata una speciale divisione interna chiamata "Science for Social Good": essa si pone l'obiettivo di utilizzare un approccio scientifico attraverso tecnologie avanzate come l'intelligenza artificiale per creare degli strumenti concreti per il bene sociale della comunità. L'azienda IBM non penso abbia bisogno di presentazione, tuttavia è da sottolineare come il dipartimento "IBM Research" attualmente risulta la più grande organizzazione di ricerca industriale del mondo, con laboratori distribuiti in tutti i sei continenti. Di seguito ho analizzato un progetto realizzato da loro che utilizza un approccio basato sui dati per cercare di scorporare e quantificare il livello di innovazione dei paesi. Esso risulta interessante perché dimostra come una metodologia empirica tramite l'analisi dei dati permetta di avere una prospettiva diversa rispetto alle classiche analisi prettamente teoriche. Come in altri progetti analizzati, quest'ultimo utilizza metriche e statistiche aggregate a livello macro-regionale e pubblicamente accessibili, per cui non presenta nessun elemento riconducibile al colonialismo dei dati. Anzi, rappresenta un ottimo strumento di supporto per comprendere al meglio i fattori socioeconomici che determinano la competitività economica di un paese.



Ad oggi (2019) oltre il progetto da me analizzato, la divisione ha realizzato altri 27 progetti, quali per citarne alcuni: "Comprensione delle epidemie", "Modellazione del discorso dell'odio" e "Caccia al virus Zika con l'apprendimento automatico". Nello sviluppo di questi progetti tipicamente l'azienda si pone come obiettivo finale quello di realizzare degli strumenti (o piattaforme) concreti come supporto per un determinato problema, come per esempio nel progetto "Neurology-as-a-Service" dove il team coinvolto sta sviluppando un sistema cognitivo cloud per interpretare automaticamente gli

elettroencefalogrammi (EEG) e apprendere continuamente dai dati.

Inoltre, l'azienda già a partire dal 2013 ha iniziato lo sviluppo del rivoluzionario ecosistema "IBM Watson" che rappresenta una piattaforma modulare, aperta e innovativa di intelligenza artificiale in grado di gestire ed elaborare grandi quantità di dati non strutturati per poi ricavarne autonomamente delle informazioni (o decisioni). Watson utilizzata in contesti (e applicazioni) aziendali permette di creare potenti modelli ad-hoc ed automatizzare l'intero ciclo di vita dell'apprendimento automatico.

4.3.1 REPORT: Come promuovere l'innovazione: un approccio basato sui dati per misurare la competitività economica *(Kuhlman et al., 2017)*

Questo progetto si basa sul presupposto che l'innovazione è un fattore chiave per la crescita economica (e sociale) nei paesi di tutto il mondo, che tuttavia risulta difficile da definire e ancor di più da misurare (Kuhlman et al., 2017). Quindi diventa fondamentale comprendere meglio l'innovazione e avere a disposizione dei metodi per poterla quantificare come supporto per decisioni politiche o aziendali.

Per questo scopo il team coinvolto ha realizzato, in collaborazione con il dipartimento "IBM Research", un modello predittivo che analizzando (attraverso un'analisi della regressione "Group lasso") una moltitudine di metriche a livello nazionale cerca di scoprire automaticamente le "leve" attuabili dell'innovazione, cioè stabilire empiricamente delle relazioni tra i singoli parametri nazionali e il livello d'innovazione. Il modello è stato anche perfezionato e allenato utilizzando i dati già disponibili (ricavati tramite sondaggi) del livello percepito d'innovazione in 150 paesi forniti dal "World Economic Forum"; tra l'altro quest'ultimo in uno studio del 2016 ("Valutazione dei principali indicatori d'innovazione") mostra che "sebbene molte relazioni precedenti abbiano tentato di quantificare l'innovazione, vi è stato uno scarso consenso sulle misure appropriate sul quale basare quest'analisi". La difficoltà nel definire l'innovazione è data dal fatto che essa sia un concetto strettamente immateriale e può assumere molti significati in diversi contesti (Kuhlman et al., 2017).

Il risultato finale di questo modello è l'indicatore "Open Innovation Index" (OII), cioè

un sistema automatico per misurare (e prevedere) l'innovazione utilizzando gli open data e algoritmi di apprendimento automatico. Gli input per il modello di questo studio sono il set di dati WDI (World Development Indicators) pubblicati annualmente dalla Banca Mondiale, che comprendono stime nazionali, regionali e globali che misurano il livello di sviluppo in un'ampia gamma di ambiti. Queste statistiche sono internazionalmente riconosciute e comprendono circa 1.500 metriche riguardo più di 200 economie di paesi e regioni in un arco temporale di 56 anni. Questi dati sono accessibili pubblicamente (open data) da tutti gli utenti attraverso il sito web “data.worldbank.org”. Dopo una preelaborazione, il set di dati finali preso in considerazione comprende circa 700 metriche annuali (più del 50% escluse per dati parziali) relativi a 142 paesi per una serie temporale di 7 anni (periodo 2009-2015).

Durante l'elaborazione, il modello creato seleziona in media 26 gruppi contenenti 257 parametri che rientrano nelle categorie; solo per citarne alcune: politica economica, formazione del debito, ambiente, settore finanziario, infrastrutture, salute, ecc. (Fig. 4.7).

World Development Indicators	Global Competitiveness Index	Pillar 12: Innovation
<ul style="list-style-type: none"> • Economic policy & debt • Education • Environment • Financial sector • Health • Infrastructure • Poverty • Private sector & trade • Public sector • Social protection & labor 	<ul style="list-style-type: none"> • Institutions • Infrastructure • Macroeconomic environment • Health & primary education • Higher education & training • Goods market efficiency • Labor market efficiency • Financial market development • Technological readiness • Market size • Business sophistication • Innovation 	<ul style="list-style-type: none"> • Capacity for innovation • Company spending on research and development • University-industry collaboration in research and development • Government procurement of advanced tech products • Availability of scientists and engineers • Number of applications filed under the Patent Cooperation Treaty (PCT) per million people in population

Figura 4.7: Alcuni parametri selezionati per formare l'Open Innovation Index (OII)

FONTE: (Kuhlman et al., 2017)

Nell'immagine sottostante (Fig 4.8) vengono visualizzati i risultati previsti (tramite i punteggi “Open Innovation Index”) del modello per il triennio 2013-2015 attraverso una mappa globale tematica. L'indicatore varia da un valore di 1 a 7, con un punteggio minimo effettivo previsto di 2,50 ed uno massimo di 5,66.

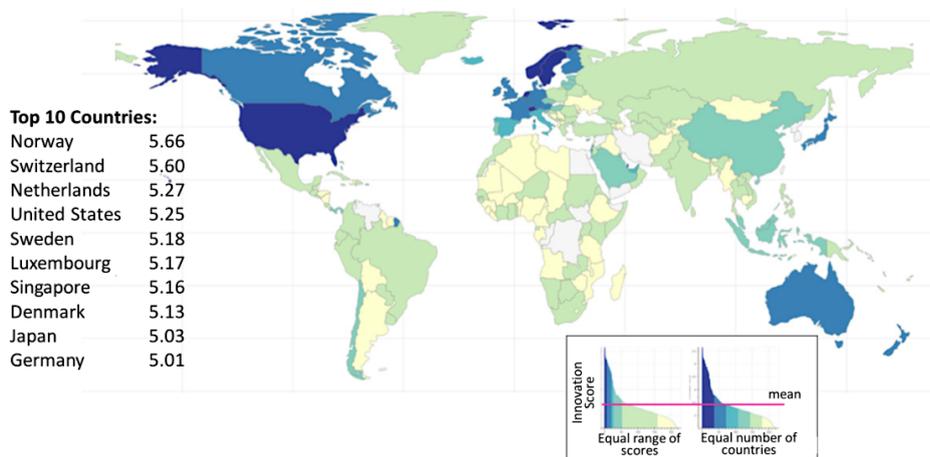


Figura 4.8: Livelli d'innovazione globali attraverso un approccio basato sui dati
 FONTE: (Kuhlman et al., 2017)

Secondo Kuhlman e altri (2017) questo “approccio basato sui dati e l'apprendimento automatico permette di offrire una prospettiva diversa rispetto alla tradizionale analisi economica che tipicamente partendo da un'ipotesi cerca di confermarla o escluderla attraverso l'applicazione di metodi statistici”. “Un approccio data-driven è focalizzato appunto sui dati stessi, ignorando (o ridimensionando) le ipotesi precedenti e le idee convenzionali sul fenomeno, permettendo di esplorare empiricamente nuove interpretazioni e idee che potrebbero essere state trascurate in precedenza” (Kuhlman et al., 2017).

Capitolo 5

AgID – Agenzia per l’Italia Digitale

Dopo aver analizzato queste tre iniziative internazionali ho approfondito un altro aspetto relativo all’uso dei dati per il bene sociale, cioè se il trend della datafication può essere utilizzato dalle Pubbliche amministrazioni per generare benefici per i singoli cittadini e l’intera comunità oppure se sono presenti degli effetti collaterali. Infatti, la posizione nettamente superiore dello Stato nei confronti dei singoli cittadini potenzialmente può contenere degli elementi di colonialismo dei dati. Per il mio obiettivo ho approfondito la situazione italiana attraverso “L’Agenzia per l’Italia Digitale” che si occupa di attuare l’agenda digitale italiana. Essa mi ha permesso di approfondire la situazione del nostro paese in riferimento all’utilizzo e benefici degli open data, i vantaggi della digitalizzazione e le prospettive future delle Pubbliche amministrazioni italiane riguardo un approccio basato sui dati.

5.1 Presentazione e obiettivi dell’Agenzia

L’iniziativa nazionale AgID (Agenzia per l’Italia Digitale) è nata nel 2012 con lo scopo di realizzare gli obiettivi contenuti all’interno dell’agenda digitale italiana e quindi contribuire alla diffusione dell’utilizzo della tecnologia e della comunicazione per favorire l’innovazione e la crescita economica. I contenuti presenti in quest’agenda nazionale ricalcano gli obiettivi e principi contenuti nella sua analogo versione europea redatta dalla Commissione europea per favorire l’innovazione all’interno dei paesi europei. Infatti,

l'agenda digitale italiana è coerente con il Piano di azione europeo per la gestione digitale della pubblica amministrazione (eGovernment) che impegna gli stati membri a definire le proprie politiche digitali sulla base di alcuni principi, per citarne alcuni: *digital by default* (servizi digitali come opzione predefinita), *once only* (non richiedere informazioni già fornite da cittadini o imprese), *cloud first* (favorire e implementare l'utilizzo di servizi cloud) e l'apertura (e trasparenza) dei dati e dei processi amministrativi.

In concreto, l'Agenzia per l'Italia Digitale (AgID) ha il compito d'attuazione tecnica del "Piano Triennale per l'informatica della Pubblica amministrazione" approvato dal governo italiano. Questo Piano Triennale è un documento di indirizzo strategico ed economico che definisce un modello di riferimento per lo sviluppo informatico nella Pubblica amministrazione, fissando i principi architetturali fondamentali, le regole di usabilità e la logica nelle spese per il settore ICT. Per esempio, in una sezione del Piano Triennale 2019-2021 (AgID, 2019) viene enfatizzato il processo d'identificazione e creazione di ecosistemi, cioè aree settoriali ed omogenee, dove l'erogazione dei servizi delle amministrazioni verso i cittadini e imprese possa avvenire completamente in modo telematico, cioè il digitale come mezzo predefinito (seguendo il principio "*digital by default*"). Una dimostrazione di ecosistema è il Servizio Sanitario Nazionale (SSN), dove al centro si posiziona il Fascicolo Sanitario Elettronico (FSE) al quale sono collegati i servizi digitali che migliorano la fruibilità delle cure, dei servizi di diagnosi e assistenza (AgID, 2019). Attualmente le principali piattaforme realizzate dall'agenzia sono: il sistema di pagamento elettronico verso le Pubbliche amministrazioni (pagoPA), il Sistema Pubblico di Identità Digitale (SPID), il Fascicolo Sanitario Elettronico (FSE) e in progressivo sviluppo l'Anagrafe Nazionale della Popolazione Residente (ANPR). Inoltre, sono presenti molte basi di dati d'interesse nazionale utilizzate come strumento essenziale per uno svolgimento più efficiente (seguendo il principio "*once only*") nelle funzioni delle amministrazioni, quali per citarne alcune più rilevanti: l'Anagrafe Nazionale della Popolazione Residente (ANPR), il Registro delle imprese, il Casellario giudiziario e il Repertorio Nazionale dei Dati Territoriali (RNDDT).

Infine, sempre all'interno dell'ultimo Piano Triennale (2019-2021) è citato tra le altre cose anche il progetto "Web Analytics Italia" che si basa sul presupposto che i "siti web degli enti pubblici sono ormai il principale vettore che conduce i cittadini verso le

informazioni e i servizi digitali erogati dalle amministrazioni” (AgID, 2019). In questo contesto è stata creata un’infrastruttura nazionale che fornisce alle pubbliche amministrazioni un punto di raccolta centrale (e standardizzato) dei dati analitici sull’uso dei siti e dei servizi digitali degli enti pubblici. Quest’infrastruttura raccoglie dati generati da una combinazione eterogenea di metriche, quali per esempio: durata delle visite, provenienza geografica, tecnologia utilizzata, ricerche interne e contenuti visualizzati.

Un’ulteriore prova rispetto al fatto che la digitalizzazione sia un fattore abilitante nell’uso dei dati e che apra nuove frontiere sulla comprensione delle azioni dei cittadini. Tuttavia, nel rapporto “Digital Economy and Society Index - DESI 2017” viene rilevato che, nonostante sia presente un aumento dei servizi digitali erogati dalle amministrazioni italiane, vi è un divario tra l’offerta di questi servizi e il loro effettivo utilizzo da parte dei cittadini.

Voglio concludere analizzando alcuni dati (aggiornati a fine 2019) relativi al progresso della digitalizzazione nelle amministrazioni italiane: più di 5 milioni di identità digitali (SPID) rilasciate, 141 milioni di fatture elettroniche verso la PA, circa 28.000 open data disponibili e l’adesione di 14 regioni al Fascicolo Sanitario Elettronico (FSE), tuttavia con un tasso d’attivazione da parte del cittadino pari al 21% (avanzamentodigitale.italia.it).

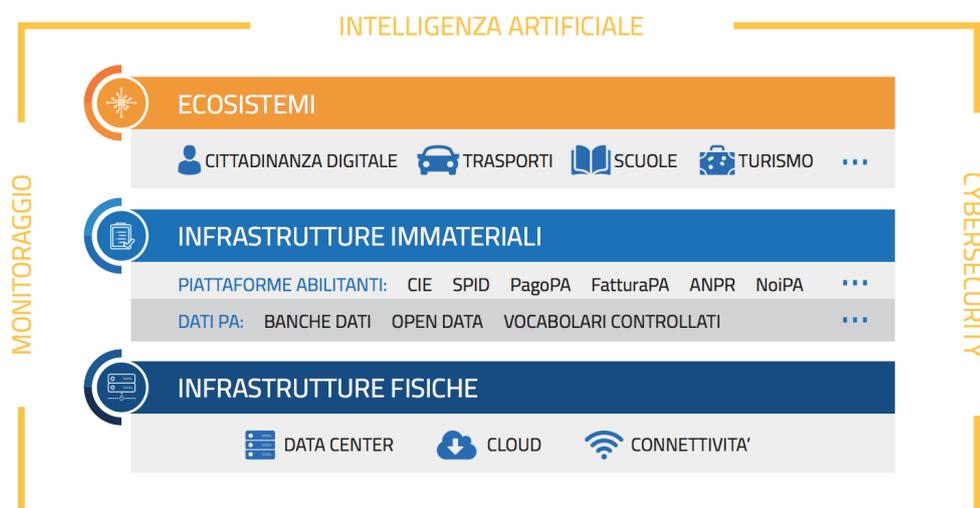


Figura 5.1: Infrastruttura digitale italiana

FONTE: (AgID, 2018b)

5.2 La situazione degli Open data in Italia

Come riportato precedentemente (nel capitolo “Il concetto di Privacy”) il gruppo di esperti “Global Pulse Privacy Advisory” nel documento da loro redatto (United Nations Development Group, 2017), in riferimento agli open data sostengono che essi siano *“un motore importante per l’innovazione, responsabilità e trasparenza”* o ancora *“quando possibile i dati dovrebbero essere resi accessibili, a meno che i rischi di renderli pubblici superino i benefici o ci siano altre basi legittime per non farlo”*.

Proprio in questo contesto e nell’ambito dell’iniziativa internazionale “Open Government Partnership - OGP” il governo italiano (tramite la propria agenda digitale) incoraggia l’utilizzo degli open data per “promuovere la cultura della trasparenza della pubblica amministrazione”. L’iniziativa OGP ha lo scopo di incoraggiare impegni concreti da parte dei singoli governi nazionali riguardo tematiche come la trasparenza, contrasto alla corruzione, sostegno alla partecipazione civica e l’utilizzo di nuove tecnologie per l’innovazione. In particolare, l’Agenda per l’Italia Digitale (AgID) elabora le linee guida nazionali per valorizzare il patrimonio informativo, specificando gli standard tecnici riguardo gli open data e gestisce il catalogo nazionale dei dataset resi disponibili dalle Pubbliche amministrazioni. Quest’ultimi dati vengono resi accessibili pubblicamente attraverso il sito “dati.gov.it” oppure “geodati.gov.it” per quanto riguarda i dati territoriali (RNDDT).

Inoltre, in adempimento della Direttiva europea sul riutilizzo dell’informazione del settore pubblico (PSI), l’agenzia ha condotto anche per l’anno 2018 un’indagine attraverso questionari somministrati alle Pubbliche amministrazioni (regioni, comuni, città metropolitane, ecc.) riguardo la situazione, l’impatto e gli ostacoli nell’implementazione degli open data. La situazione emersa da questa indagine (AgID, 2018a) evidenzia come il principale ostacolo nell’implementazione e la diffusione degli open data sia quello di “una mancanza di una cultura solida e diffusa, che permetta a cittadini e imprese di comprendere il valore dei dati aperti” (AgID, 2018a). Ciò nonostante viene constatato che rispetto all’anno precedente (2017), gli approcci delle amministrazioni “appaiono più consolidati e orientati al coinvolgimento di attori esterni alla Pubblica amministrazione, attraverso la promozione di seminari, convegni ed eventi legati al tema dei dati aperti e alla trasparenza della PA” oppure anche predisponendo percorsi di formazione per i

propri dipendenti. Inoltre, nella valutazione sull'impatto nell'uso di questi dati da parte delle amministrazioni viene rilevato che “sono in aumento i casi d'uso dei dati aperti per la progettazione di politiche pubbliche (pianificazione urbanistica, gestione territorio, definizioni piani sanitari, ecc.)” (AgID, 2018a). Tuttavia, sempre in questo rapporto è stato anche evidenziato una scarsa attenzione delle amministrazioni nel “predisporre d'azioni strutturate di analisi delle ricadute economiche e sociali che i dati aperti generano una volta esposti” che rende attualmente “poco evidente l'impatto generato attraverso il riutilizzo dei dati resi disponibili dalle pubbliche amministrazioni” (AgID, 2018a).

Questa indagine italiana è successivamente inserita, insieme alle corrispettive europee, nel più ampio contesto del report annuale “Open Data Maturity in Europe” redatto dall'European Data Portal che si pone l'obiettivo di analizzare la situazione degli open data all'interno dell'Unione europea. Nel loro ultimo report (European Data Portal, 2018) viene evidenziato un tasso di maturità complessiva del 65% a livello europeo, con un picco del 82% raggiunto dalla maturità della componente “Policy” (Fig. 5.2). Quest'ultima dimostra “una solida base in termini di obiettivi strategici” riguardo gli open data, in contrapposizione con il tasso di maturità del 50% relativo all'impatto nell'uso di questi dati (Impact) che “fornisce la prova del ritmo molto lento con cui si muovono gli Stati membri dell'UE” (European Data Portal, 2018). In questo contesto europeo l'Italia è tra le più attive in questo trend, in particolare si colloca in quarta posizione (Fig. 5.3), con un tasso di maturità complessivo pari al 80% preceduta dall'Irlanda (88%), Spagna (87%) e Francia (83%).

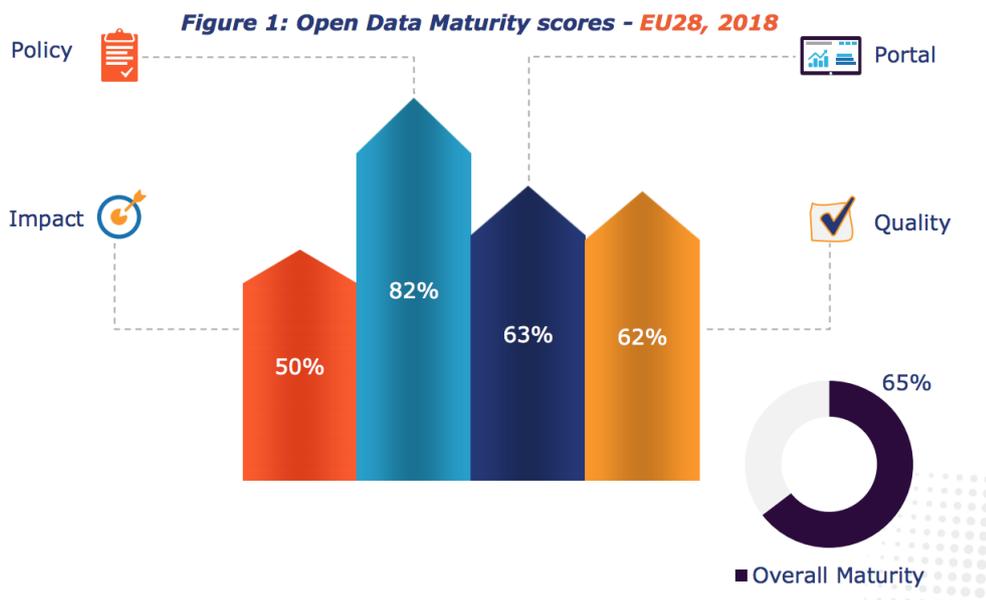


Figura 5.2: Punteggio europeo riguardo la maturità degli Open data
 FONTE: (European Data Portal, 2018)

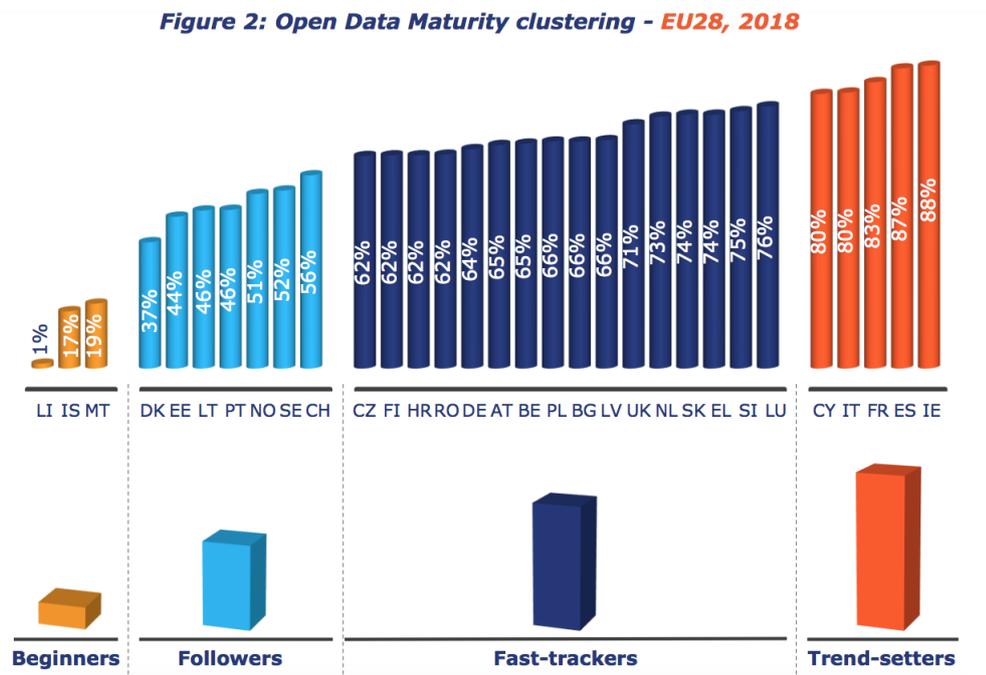


Figura 5.3: Maturità degli Open data - Classificazione paesi europei
 FONTE: (European Data Portal, 2018)

5.3 Progetti e appalti innovativi

All'interno delle linee guida previste dal Piano Triennale per l'informatica nella Pubblica Amministrazione 2019-2021, nasce il portale "appaltinnovativi.gov.it" che si pone l'obiettivo di permettere alle Pubbliche amministrazioni (PA) presenti sul territorio italiano di proporre delle iniziative e progetti innovativi, esplorando nuove possibilità permesse dalla digitalizzazione e da un approccio basato sui dati. Questa piattaforma abilita un percorso d'innovazione attuabile in tre distinte modalità: l'esposizione di un proprio fabbisogno, l'attuazione di consultazioni di mercato (o tramite enti di ricerca) o l'assegnazione di appalti per la realizzazione concreta di un progetto prestabilito. Essa permette quindi l'incontro tra il fabbisogno d'innovazione delle Pubbliche amministrazioni e le aziende che offrono soluzioni all'avanguardia per raggiungere un determinato obiettivo. I progetti sono per la totalità finanziati da fondi regionali o messi a disposizione grazie a bandi o iniziative europee (come Horizon 2020). Tutti i progetti presenti sulla piattaforma, accessibile pubblicamente attraverso il portale web, risultano per lo più nella fase embrionale piuttosto che ultimati come nelle iniziative analizzate precedentemente. Tuttavia, risulta interessante esplorare alcuni di questi progetti per comprendere le prospettive e gli obiettivi della Pubblica amministrazione italiana nel breve-medio periodo.

Per esempio, anche il governo italiano, come dimostrato da molte iniziative internazionali, vuole esaminare i possibili utilizzi dei big data per gestire le situazioni d'emergenza. In particolare, il governo nelle vesti del Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) ha indetto un appalto (per un importo fino a 6 milioni) per esplorare le possibilità d'utilizzo dei dati generati dai cittadini come mezzo per segnalazioni d'emergenza o integrandoli in sistemi automatici. Questo appalto si divide in due bandi distinti: per ricercare nuovi modi per consentire l'allertamento preventivo dell'emergenza (in particolare riguardo i fenomeni franosi) e un supporto operativo per aumentare l'efficienza nelle fasi di soccorso. Un altro progetto interessante è stato proposto dalla Regione Valle d'Aosta che intende trovare metodi innovativi per stabilire la quantità di rifiuti conferiti al servizio pubblico da parte dei singoli cittadini (o imprese) in adempimento alla norma contenuta nel Decreto 20/04/2017 del Ministro dell'Ambiente. Quest'ultima stabilisce come obiettivo finale quello di attuare un modello di tariffazione (per i rifiuti)

commisurato al servizio reso e che consenta una copertura efficiente dei costi. In particolare, la regione ha stanziato 1 milione e mezzo per la realizzazione di un sistema informativo basato sui dati che integri: la georeferenziazione dei punti di conferimento stradale (cassonetti, contenitori interrati, ecc.), l'identificazione delle singole utenze e la determinazione della quantità/volume per ogni conferimento relativo ad un'utenza.

5.4 L'intelligenza artificiale al servizio del cittadino (IA-GOV)

Poc'anzi abbiamo visto come l'utilizzo degli open data e la digitalizzazione della burocrazia permettano di utilizzare un approccio basato sui dati che si rivela utile per migliorare l'efficienza dell'apparato statale e della vita dei suoi cittadini. Tuttavia, una delle sfide future risulta quella di attuare un'integrazione etica e sostenibile di nuove tecnologie all'interno delle amministrazioni statali. In particolare, diventa cruciale le modalità di integrazione di tecnologie basate sull'intelligenza artificiale, che sostanzialmente cercano di sostituire e replicare una condotta umana. Risulta quindi importante formulare dei principi etici da rispettare se ci si vuole allontanare dal riferimento negativo del colonialismo dei dati. Oltre ciò, non è da sottovalutare che un utilizzo improprio e malevole (Fig. 5.4) di questa tecnologia avanzata può recare danni all'intera società, come per esempio: utilizzo di tecniche Deepfake per fini diffamatori o di propaganda politica, attacchi terroristici sfruttando robot autonomi oppure sfruttare l'apprendimento automatico per attacchi informatici.

Proprio per esplorare l'uso etico dell'intelligenza artificiale ho analizzato il lavoro della task force (AI-GOV) che è stata formata all'interno di AgID che si focalizza proprio su questa specifica tematica. Soprattutto dato dal fatto che negli ultimi anni si è assistito alla nascita, con sviluppi esponenziali, del campo legato all'Intelligenza artificiale (AI) che è stata permessa soprattutto grazie a fattori come: l'incremento della disponibilità di fonti di dati, il miglioramento delle tecniche di apprendimento automatico e da una disponibilità di elevata capacità computazionale a basso costo (AgID, 2018b). Questa tecnologia può essere utilizzata per migliorare l'efficienza nelle Pubbliche amministra-

zioni per esempio nel sistema sanitario, scolastico e giudiziario oppure per migliorare i processi d'interazione con i cittadini che potranno essere resi più semplici, efficaci e veloci. Inoltre essa può essere utilizzata per aumentare la sicurezza nazionale; come nel caso del progetto europeo "DANTE" integrato all'interno dell'Arma dei Carabinieri. Insieme ad altri 18 partner europei, l'Italia ha collaborato nella realizzazione di un sistema automatico di rilevazione e analisi di contenuti terroristici online. Il sistema si basa su avanzate tecniche di intelligenza artificiale applicata al web mining, che permettono di individuare e analizzare contenuti eterogenei (anche in diverse lingue), presenti online sia in chiaro che nel dark web. L'obiettivo di questo strumento è quello di tracciare con più completezza ed efficacia le operazioni dei network terroristici (propaganda, addestramenti e finanziamenti) che altrimenti potrebbero non emergere attraverso l'utilizzo di metodi tradizionali.

Nello specifico la task force AI-GOV si pone come obiettivo quello di analizzare come soluzioni basate sull'intelligenza artificiale possano incidere sull'evoluzione dei servizi pubblici e migliorare il rapporto tra Pubblica amministrazione e cittadino. Il gruppo è formato da circa trenta professionisti con formazione multidisciplinare che svolgono principalmente un ruolo di coordinamento ed elaborazione di contenuti anche attraverso la formazione di sottogruppi tematici.

Il principale risultato conseguito da questo gruppo è stato quello di realizzare e presentare nel marzo 2018 il documento (AgID, 2018b) chiamato "Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino", che raccoglie suggerimenti e raccomandazioni per la Pubblica amministrazione italiana riguardo l'adozione dell'intelligenza artificiale, per cercare di seguire questo trend che ormai è pratica comune a livello internazionale. Con il termine "libro bianco" si intende un rapporto ufficiale che contiene delle proposte d'azione comunitaria in uno specifico settore o relative ad un determinato argomento. Oltre che enfatizzare il ruolo dell'intelligenza artificiale, questo documento esplora le sfide che è necessario affrontare per integrare questa tecnologia all'interno delle Pubbliche amministrazioni, o più in generale nella società. Sfide che si basano su un principio antropocentrico, cioè il fatto che "l'intelligenza artificiale debba essere sempre messa al servizio delle persone e non viceversa" (AgID, 2018b). Ho cercato di riassumere e sintetizzare le nove sfide identificate dal gruppo di lavoro (AgID, 2018b):

1. **Etica:** un uso scorretto della tecnologia può contribuire a sviluppare società ingiuste che accrescono le già presenti disuguaglianze, soprattutto se una tecnologia come l'intelligenza artificiale è utilizzata all'interno di processi decisionali legati a questioni delicate come quelle sociali, sanitarie o giudiziarie. I principali problemi che vengono identificati sono quelli relativi alla qualità e neutralità dei dati, l'individuazione di eventuali responsabilità (giuridiche o morali) e la trasparenza nell'utilizzo di questi strumenti.
2. **Tecnologica:** allo stato attuale l'intelligenza artificiale non è ancora in grado di riprodurre il funzionamento complesso di una mente umana, è per questo motivo che l'evoluzione di essa deve passare attraverso l'enfaticizzazione dell'interazione uomo-macchina. Quest'ultima permessa attraverso una modellazione che risponda meglio alle esigenze dell'utente, ai suoi interessi e soprattutto in relazione al contesto reale.
3. **Sviluppo di competenze:** negli ultimi decenni il mondo del lavoro è stato attraversato da una profonda trasformazione e vi è stata e vi sarà una forte evoluzione delle professioni che dovrà per forza essere affrontata attraverso il sistema scolastico e universitario. L'evoluzione di questi ultimi è abilitata soprattutto da una continua formazione degli insegnanti e combattendo l'analfabetismo digitale. Inoltre, un'ulteriore sfida per il settore pubblico sarà quello di attirare (e trattenere) figure professionali altamente qualificate pur non potendo garantire livelli salariali del settore privati.
4. **Ruolo dei dati:** essi ricoprono un ruolo fondamentale nell'utilizzo di tecniche come l'intelligenza artificiale o l'apprendimento automatico, per questa ragione i dati dovrebbero essere di buona qualità e liberi da pregiudizi o errori. Inoltre, l'incremento esponenziale di sorgenti di dati eterogenei e non strutturati ha fatto sì che aumentassero i problemi d'integrazione fra questi dati che potranno risultare frammentati o distribuiti irregolarmente nello spazio e nel tempo. Infine, tecniche di apprendimento automatico come quella di tipo supervisionato (supervised learning), dove viene delegato ad una persona il compito di istruire il sistema, possono introdurre delle distorsioni o pregiudizi nel funzionamento del stesso.

5. **Contesto legale:** è necessario trovare un bilanciamento tra gli interessi della collettività e quelli individuali, come per esempio tra la trasparenza amministrativa e il diritto alla privacy. Inoltre, l'esigenza di avere a disposizione enormi collezioni di dati ha generato un notevole interesse economico riguardo la raccolta di dati anche sensibili, che dovrebbe essere bilanciata da una maggiore tutela giuridica.
6. **Accompagnare la trasformazione:** l'utilizzo di tecniche di intelligenza artificiale rappresenta non solo un'innovazione tecnologica ma anche sociale. Tuttavia, le persone dovrebbero essere affiancate da queste tecnologie e non sostituite; è per questa ragione che lo Stato, le imprese e i cittadini dovrebbero affrontare questa trasformazione in modo coeso. Inoltre, ci dovranno essere maggiori sinergie tra il settore pubblico e quello privato con lo scopo di implementare soluzioni innovative.
7. **Prevenire le disuguaglianze:** l'utilizzo dell'intelligenza artificiale può ridurre il divario sociale, economico o culturale esistente, ma al tempo stesso può accrescere ulteriormente le disuguaglianze presenti. L'aumento di quest'ultime causate per esempio dal funzionamento del sistema stesso, con dati o algoritmi affetti da distorsioni oppure attraverso una limitazione d'accesso dei benefici permessi nell'utilizzo di questa tecnologia. Invece, una riduzione di un divario già presente può essere abilitata per esempio dal supporto di soggetti con disabilità attraverso sistemi intelligenti (assistenti vocali, interpretazione del comportamento, ecc.) che permette loro un'interazione alternativa.
8. **Misurare l'impatto:** l'utilizzo di tecniche d'intelligenza artificiale e di apprendimento automatico permette di misurare l'impatto di politiche socioeconomiche utilizzando un approccio multidisciplinare, attraverso nuovi indicatori e strumenti che permettono di misurare nuove componenti come quelle sociali o psicologiche oltre che quelle tradizionali (economiche, tecnologiche, ecc.). Inoltre, delle politiche socioeconomiche più efficaci permettono ai cittadini di usufruire di un servizio migliore o addirittura di migliorare la qualità della propria vita.
9. **La centralità dell'uomo:** l'inclusività dovrebbe essere messa al centro della narrazione tecnologica, esplicitando il significato e il ruolo dell'intelligenza artificiale

all'interno della nostra società. È necessario rendere tutti consapevoli dei vantaggi e problemi che può derivare nell'utilizzo di questi strumenti.

Oltre ciò, come descritto attraverso il loro sito ufficiale (ia.italia.it), attualmente il gruppo di lavoro si sta occupando di un percorso di sperimentazione legato ad un'implementazione più concreta all'interno delle amministrazioni pubbliche. Questo percorso si basa sul progetto di co-progettazione chiamato "Usa l'intelligenza" che ha l'obiettivo di implementare soluzioni concrete di intelligenza artificiale per migliorare i processi interni alle amministrazioni o i servizi erogati ai cittadini coinvolgendo le imprese che offrono soluzioni innovative basate su questa tecnologia. La sperimentazione coinvolge diverse amministrazioni italiane che hanno mostrato interesse nell'iniziativa (Comune di Milano, Regione Emilia-Romagna, Corpo Nazionale dei Vigili del Fuoco, ecc.) e riguarderà futuri progetti legati ad una vasta gamma di settori, per esempio ambientale (previsioni dei rischi, studio degli eventi atmosferici, ecc.), sanitario (analisi dati relativi a malattie croniche, integrazione di servizi digitali, ecc.), trasporti (analisi flussi di traffico, ecc.) o legati a servizi erogati ai cittadini (piattaforme, chatbot, ecc.).

QUALI POSSONO ESSERE I POTENZIALI DANNI ASSOCIATI A SISTEMI IA		
DIGITALI	FISICI	POLITICI
✓ Phishing automatizzato, creazione di email false, siti web e link per sottrarre informazioni.	✓ Terrorismo automatizzato che utilizza droni o veicoli autonomi come armi.	✓ Propaganda, attraverso immagini e video falsi facilmente generabili.
✓ Hacking massivo, attraverso la scoperta automatizzata delle vulnerabilità dei sistemi.	✓ Sciame di Robot che cercano di raggiungere lo stesso obiettivo.	✓ Rimozione automatica del dissenso, grazie alla possibilità di trovare, analizzare e rimuovere automaticamente testi ed immagini.
✓ Ingannare sistemi IA sfruttando a proprio vantaggio i difetti nell'interpretazione della realtà da parte dell'IA.	✓ Attacchi remoti, resi possibili dal fatto che i robot autonomi possono essere controllati da qualunque distanza.	✓ Persuasione personalizzata, con l'utilizzo di informazioni pubbliche per influenzare l'opinione di qualcuno.

Fonte: AI experts list the real dangers of artificial intelligence, Dave Gershgorin Feb. 2018

Figura 5.4: Potenziali danni causati da sistemi AI

Conclusioni

Dopo aver analizzato tutti questi progetti realizzati all'interno di iniziative internazionali e non, posso affermare che i big data possono essere usati per il bene sociale. Tuttavia, il confine tra il colonialismo dei dati e un utilizzo etico non è sempre così netto. Inoltre, dopo aver esaminato queste iniziative sono in grado di formulare alcune caratteristiche ed elementi comuni che possono abilitare un utilizzo etico e per il bene sociale di questi dati, ossia:

1. **La tutela della privacy come principio fondamentale e indispensabile per tutte le fasi che coinvolgono i dati:** soprattutto utilizzando tecniche e metodologie adeguate come l'aggregazione e anonimizzazione. In quanto, l'obiettivo centrale di un utilizzo per il bene sociale è quello di garantire benefici etici per tutta la popolazione e non identificare specifici individui.
2. **L'utilizzo dei dati deve essere il più trasparente possibile,** in particolare rendere consapevoli i cittadini dei benefici e rischi nell'utilizzare un approccio basato sui dati. Soprattutto non discostarsi da quello dichiarato nell'informativa sulla privacy; quest'ultima tuttavia dovrà essere resa più facilmente comprensibile agli utenti.
3. **Progetti caratterizzati da un modello di business etico e sostenibile:** la maggior parte dei dati sono posseduti da aziende commerciali, che tuttavia molto raramente intraprenderanno progetti pro bono per il bene sociale. Infatti, la maggior parte dei progetti delle organizzazioni con obiettivi sociali utilizzano soluzioni ad-hoc attraverso l'impiego di dati accessibili pubblicamente (open data) o donati per scopi scientifici/statistici.

Quest'ultimo punto è la motivazione principale per cui non vi è un utilizzo strutturato e su larga scala di progetti con obiettivi sociali e/o umanitari. Poiché la maggior parte dei dati sono posseduti da aziende commerciali, che cederanno i loro dati solo a fronte di un corrispettivo economico. Cioè proprio quell'estrazione di valore che caratterizza il fenomeno del Colonialismo dei dati. Proprio la riconsiderazione di questa metodologia da parte delle organizzazioni umanitarie porta quest'ultime a dover trovare dei modelli di business alternativi. In questo contesto l'associazione GSMA (2018) ha elaborato dei possibili modelli di business sostenibili per progetti sociali legati ad un approccio basato sui dati:

- (a) Filantropico (ottenere i dati pro bono);
- (b) Finanziamento attraverso donazioni;
- (c) Generare profitto da altri progetti collaterali basati sui dati (rispettando comunque le finalità etiche);
- (d) Sfruttare piattaforme gratuite o trial-based (utilizzare open data o accessibili parzialmente in modo gratuito);
- (e) Realizzare un prodotto commerciale (generare profitti etici attraverso la creazione di servizi/prodotti a pagamento).

Per quanto riguarda la centralità della Privacy e la trasparenza nell'utilizzo dei dati risultano importanti partire dai principi base elaborati rispettivamente da: *Global Pulse Privacy Advisory* (CAP. 2: "Il concetto di Privacy") e dal gruppo *AI-GOV* all'interno dell'Agenzia per L'Italia Digitale (CAP. 8: "AgID – Agenzia per l'Italia Digitale").

Infine, per quanto riguarda la concessione dei dati dei cittadini verso lo Stato e le Pubbliche amministrazioni ci sono elementi che evidenziano dei benefici concreti per la comunità e i singoli cittadini. Un approccio basato sui dati permette ad un governo di migliorare i servizi erogati e gestire con più efficacia situazioni d'emergenza. Inoltre, il trend della datafication permette allo Stato di generare ed esporre gli open data che possono essere utilizzati come motore per l'innovazione e crescita economica del paese. Tuttavia, la raccolta di molti aspetti economici e burocratici dei cittadini da parte di

un governo nazionale non garantisce un distacco netto dal concetto negativo del colonialismo dei dati. Anzi, la posizione nettamente superiore dello Stato nei confronti dei singoli cittadini potenzialmente può contenere degli elementi di colonialismo dei dati. La sostanziale discriminante per un utilizzo etico sta quindi nella cultura e sensibilità del governo di un determinato paese.

Il trend della datafication ha permesso una visione alternativa dei fenomeni socioeconomici del nostro tempo, tuttavia il tema del colonialismo dei dati è più che mai centrale nello sviluppo della nostra comunità. Bisognerebbe ripensare al ruolo che stiamo dando alla tecnologia all'interno della nostra società, anche mettendo in discussione la pretesa di una visione totalitaria che spesso perde di vista le persone dietro ai freddi numeri.

Bibliografia

AgID (2018a): Indagine sul grado di maturità degli open data e sullo stato di attuazione della direttiva PSI – 2018.

AgID (2018b): L'intelligenza artificiale al servizio del cittadino. AI-Gov. Agenzia per l'Italia Digitale. Disponibile su: <https://ia.italia.it/assets/librobianco.pdf>.

AgID (2019): Piano Triennale per l'Informatica nella Pubblica Amministrazione 2019/21.

Altemeyer B. (2019): Making the business case for AI in HR: two case studies. *Strategic HR Review* Vol. 18 No. 2. pp. 66-70.

Arvidsson A. (2016): Facebook and Finance: On the Social Logic of the Derivative. *Theory Culture & Society* 33 (6): 3–23.

Blumenstock J. (2018): Don't forget people in the use of big data for development. *Nature* Vol. 561, 170-172.

Cannataci A. Joseph (2016): Report of the Special Rapporteur on the right to privacy. UN Human Right Council. A/HRC/31/64.

Chen Jolie Yujie. (2017): Thrown under the Bus and Outrunning It! The Logic of Didi and Taxi Drivers' Labour and Activism in the on-Demand Economy. *New Media & Society* 20 (8): 2691–2711.

Couldry Nick e Ulises A. Mejias. (2019): Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media* 20 (4): 336–49.

Cukier e Mayer-Schöenberger. (2013): Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà. Garzanti.

Dagnino E. (2017a): People Analytics: lavoro e tutele al tempo del management tramite big data. Labour & Law Issues Vol. 3, No. 1.

Dagnino E. (2017b): Big data e lavoro: le sfide della workforce analytics. La nuova grande trasformazione del lavoro. Pag. 136. Adapt University Press.

Degryse C. (2016): Digitalisation of the economy and its impact on the labour markets. ETUI. Working Paper 2016.02.

Elias Daniel, Nadler Friedrich, Stehno Johannes, Krösche Jens, Lindorfer Manuel (2016): SOMOBIL – Improving public transport planning through mobile phone data analysis”. Transportation Research Procedia 14 (4478 – 4485) .

European Data Portal (2018): Open Data Maturity in Europe - Report 2018. European Union.

GSMA (2014): Mobile Privacy: Consumer research insights and considerations for policymakers. February 2014

GSMA (2018): Scaling Big Data for Social Good: The need for sustainable business models. Big Data for Social Good.

GSMA BD4SG (2018): Building communities resilient to climatic extremes. TEF-FAO Case Study. November 2018.

GSMA BD4SG (2019a): Utilising Real-Time Mobile Analytics to Aid in the Rescue of Missing Persons. Megafon Case Study. September 2019.

GSMA BD4SG (2019b): Responding quickly and effectively to disasters in Japan. January 2019.

IOM (2017): Mediterranean migrant arrivals top 363,348 in 2016; deaths at sea: 5,079. IOM Press Release, 6 January.

Kuhlman C., Ramamurthy K. N., Sattigeri P., Lozano A. C., Cao L., Reddy C., Mojilovic' A., Varshney K. R. (2017): How to foster innovation: A data-driven approach to measuring economic competitiveness. IBM j. Res. & dev. Vol. 61 no. 6 paper 11.

La Rue Frank (2013): Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. UN Human Right Council. A/HRC/23/40.

Matthew Crawford (2017): Smart cities and technologies: connected or disconnected citizens?. Field Actions Science Reports [Online] Special Issue 16.

Milan Stefania e Treré Emiliano. (2019): Big Data from the South(s): Beyond Data Universalism. Television & New Media 20 (4): 319–35.

O' Neil Cathy. (2017): Armi di distruzione matematica. Bompiani.

Philippe Sajahau (2016): IBM – Building sustainable cities through. partnerships and integrated approaches. Field Actions Science Reports [Online], Special Issue 16

Ricaurte Paola. (2019): Data Epistemologies, The Coloniality of Power, and Resistance. Television & New Media 20 (4): 350–65.

Rota A. (2017): Rapporto di lavoro e big data analytics: profili critici e risposte possibili. Labour & Law Issues vol. 3 no. 1.

Shamsi A. Jawwad, Khojaye Ali Muhammad (2018): Understanding Privacy Violations in Big Data Systems. IEEE Computer Society IT Professional May/June 2018.

Sturgess P., Sparrey R, DFID (2016): What is Resilience? Evidence on Demand, UK 65 pp.

UN Global Pulse (2017a): Using Big Data to Study Rescue Patterns in the Mediterranean. Project Series no. 29.

UN Global Pulse (2017b): The state of mobile data for social good report. June 2017.

UN Global Pulse (2018): UN Global Pulse Annual report 2018.

United Nations Development Group. (2017): Data privacy, ethics and protection - Guidance note on Big data for achievement of the 2030 agenda.

UNODC (2012): Monitoring the Impact of the Economic Crisis on Crime. In: Rapid Impact and Vulnerability Analysis Fund (RIVAF). pp. 36-41. UN Global Pulse.

Visentin Martina (2019): Alternanza Scuola-Lavoro e Pubblica amministrazione, prime riflessioni su un'indagine Welfare digitale. *Sociologia e Politiche Sociali* vol.22 1/2019 pp. 75-89.

Zuboff Shoshana. (2017): The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Public Affairs.